Populations of *Escherichia coli* in clinical samples of Urinary Tract Infections and Bacteraemia

Fahad Abdullatif Alhashash

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

July 2015

Abstract

Extraintestinal pathogenic *E. coli* (ExPEC) strains are the main etiologic agent of urinary tract infections (UTIs). ExPEC strains are also reported to be the most common cause of bacteraemia in the world, which often originate from UTI. The population structure of UTI *E. coli* strains is well described in the literature with increased prevalence of multidrug resistance driven by extended spectrum β lactamases (ESBLs). ESBL carriage and multidrug resistance of bacteraemia *E. coli* is on the increase yet little information is available about their population structure. With the aim to define the bacteraemia population structure, *E. coli* isolated from urine samples and blood cultures were collected from the Nottingham University Hospital NHS trust over a five month period. Isolates were tested for antimicrobial resistance, ESBL and virulence associated gene (VAG) carriage, and were typed by MLST. Significantly higher ESBL driven multidrug resistant strains were observed in the bacteraemia *E. coli* compared to the UTI isolates with no significant difference in the carriage of VAGs. Our data shows a reduction in population diversity within the bacteraemia isolates compared to the concomitant urine sample population resulting in a small number of dominant sequence types (STs) (ST131, ST73, ST95) which is associated with ESBL conferred multi drug resistance and not specific virulence genes. This suggests that the increased prevalence of ESBL carriage in ExPEC isolates is leading to a selective advantage in a small number of dominant lineages causing bacteraemia in patients. Comparative genome analysis of selected isolates belonging to the dominant ST (ST73) from bacteraemia and UTI was performed to investigate the presence of bacteraemia specific loci that may explain the loss of diversity in bacteraemia. No genomic regions were identified specific for the bacteraemia ST73 isolates other than ESBL carriage. Plasmid profiling of the ESBL positive isolates of this ST73 group from bacteraemia and UTI identified diverse types of plasmids spread between the strains. No specific genomic loci were identified specific for ESBL positive ST73 isolates from bacteraemia and UTI. This concludes

that random acquisition of ESBL plasmids by any ST73 *E. coli* may select for its progression to bacteraemia which is serious and debilitating. Our study provided a comprehensive snapshot of the *E.coli* population structure from contemporaneous clinical cases of UTI and bacteraemia. The large increase in multi-drug resistance in bacteraemia ExPEC populations compared to co-circulating UTI populations is of clinical concern and represents a challenge in control and treatment of serious extra-intestinal *E. coli* infections. This provides an important clinical insight into how common *E. coli* STs could adapt to become dominant bacteraemia agents.

Declaration

I hereby declare that the work presented herein is the result of my original research work, except where references have been made to acknowledge the literature. This work is an intellectual property of the author. You may copy up to 5% of this work for private study, or personal, non-commercial research. Any re-use of the information contained within this document should be fully referenced, quoting the author, title, university, degree level and pagination. Queries or requests of any other use, or if a more substantial copy is required, should be directed in the first instance to the owner(s) of the intellectual property rights. Experiments were performed in the Pathogen Research Group at Nottingham Trent University.

Fahad Alhashash

Acknowledgment

I would like to thank Allah Almighty for giving me the strength and patience to carry out and fulfil all the work needed in this PhD project.

I would like to express my deepest gratitude and thanks to my director of studies Dr Alan McNally, I would like to thank you for your endless support and advice throughout this project. You are my mentor and friend, your guidance and encouragement made me the better researcher I am today. Many thanks for members of the Pathogen Research Group at Nottingham Trent University for their support and endless help. I would like to Dr Georgina Manning for her advice especially in the lab meetings. Many thanks for my second supervisor prof. Stephen Forsythe for his support and Dr Michael Loughlin for their support. I would like to thank our collaborators from the microbiology lab at Nottingham University Hospital NHS Trust Dr Vivienne Weston and Dr Mathew Diggle for their efforts and support in this project and the published work on the population of bacteraemia *E. coli*. Many special thanks for my dearest friends and colleagues Dhahi Alenezi, Abdulaziz Alqasim, Alya Redwan and Miquette Hall for their continuous support, encouragement and help which made this project and enjoyable journey.

Special thanks for my Mother, Father brothers and sister, your prayers sustained me this far. At the end I would like to express my deepest appreciation for my beloved wife Fatma for all the sacrifices you made on my behalf, who spent sleepless nights and supported me when I most needed it. Words cannot express how grateful I am to you, I dedicate this work to you and my lovely sons Abdulrahman, Yousef and Abdullatif, you are my real pride in life.

Publications

1. Research articles

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A. (2013). Multidrug-Resistant *Escherichia coli* Bacteremia. *Emerg. Infect. Dis*. 19 (10): 1699-1701

2. Poster presentation

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A. (2012). Society for General Microbiology Spring Conference March 2012. Dublin Convention Centre.

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A. (2012). SST School Research Conference May 2012. Nottingham Trent University.

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A. (2011). Society for General Microbiology Autumn Conference September 2011. University of York.

Table of Contents

List of tables

List of Figures

Abbreviations

APEC	avian pathogenic *Escherichia coli*
CC	clonal complex
CDS	coding sequences
CFU/ml	Colony-forming unit/ml
CLB	cell lysis buffer
CSB	cell suspension buffer
bp	base pair
DAEC	diffusely adherent *E. coli* (DAEC)
dH_2O	distilled water
DNA	deoxyribonucleic acid
DNTPs	Deoxyribonucleotide triphosphate
EAEC	enteroaggregative *E. coli*
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetic acid, sodium hydroxide
EPEC	enteropathogenic *E. coli*
EHEC	enterohemorrhagic *E. coli*
EIEC	enteroinvasive *E. coli*
ERIC PCR	enterobacterial repetitive intergenic consensus PCR
ESBL	extended spectrum beta lactamase

ETEC	enterotoxigenic *E. coli*
ExPEC	extra intestinal pathogenic *E. coli*
g	gram
gb	gigabytes
GBK file	GENBANK File
G + C	Guanine and cytosine
HCL	hydrochloric acid
HUS	haemolytic uremic syndrome
IBCs	intracellular bacterial communities
kb	kilo base pairs (1,000bp)
LPS	lipopolysaccharide
Mb	mega base pairs (1,000,000bp)
MDR	multi drug resistance
mg	milligram
ml	millilitre
MLEE	multilocus enzyme electrophoresis
MLST	multilocus sequence typing
MRSA	methicillin resistant *Staphylococcus aureus*
MST	minimum spanning tree
NaOH	sodium hydroxide

NCBI	National Centre for biotechnology Information
OD	optical density
ORFs	open reading frames
PAIs	pathogenicity islands
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
Sarcosyl	N-Lauroyl-Sarcosine, Sodium salt
SAT	secreted autotransporter toxin
SDS	sodium dodecyl sulphate
STEC	shiga toxin producing *E. coli*
SNP	single nucleotide polymorphism
ST	sequence type
TBE	Tris base, boric acid and EDTA
TE	Tris base EDTA
TraDIS	transposon directed insertion site sequencing
UPEC	uropathogenic *E. coli*
UTI	urinary tract infection
µl	microliter
VAGs	virulence associated genes
VFs	virulence factors

Chapter one

Introduction

1.0 Introduction

1.1 *Escherichia coli*

Escherichia coli is a Gram negative, rod shaped, non-sporulating, facultative anaerobic bacterium belonging to the *Enterobacteriaceae* family. It is a natural commensal inhabitant of the intestines of humans, animals, lizards and birds (Tenaillon et al., 2010; Odonkor and Ampofo, 2013). *E. coli* is shed in the faeces of these natural hosts into the environment, where it can survive in soil, water or on food for several days. The successful survival of *E. coli* in these secondary environments is dependent upon several factors, such as the availability of nutrients and water, temperature, and acidity (van Elsas et al., 2011). Since *E. coli* is primarily an intestinal inhabitant survival outside the host is limited and the presence of *E. coli* in food or water is an indicator of faecal contamination or poor hygienic practices (Odonkor and Ampofo, 2013).

E. coli is one of the first bacteria that colonize the human intestine, with initial colonization starting from early infancy. This occurs due to the transmission of *E. coli* from the mother to the new-born (Mandar and Mikelsaar, 1996; Watt et al., 2003), other studies have also reported a role for nursing staff in this early transmission (Bettelheim and Lennox-King, 1976). In fact, improved hygienic practices in hospitals not only delays the original transmission of *E. coli* to the infant, it also limits this early colonization (Nowrouzian et al., 2003). This commensal colonization occurs in the mucus of the epithelium of the large intestine, mainly the caecum and colon, where *E. coli* is shed in the lumen and excreted with the faeces (Tenaillon et al., 2010).

Unlike the intestinal commensal *E. coli*, intestinal pathogenic *E. coli* strains have acquired virulence factors enabling them to cause many serious diseases when present in the intestines

(Kaper et al., 2004). Classical examples of these strains are the enterohemorrhagic *E. coli* (EHEC), such as *E. coli* O157:H7, which have recently made the headlines for disease outbreaks of hemorrhagic colitis worldwide (Ateba and Mbewe, 2013). Another group of pathogenic *E. coli* strains can asymptomatically colonize the intestine. However, these *E. coli* strains have the ability to cause diverse and serious diseases outside of the intestinal tract. Hence, they are named extraintestinal pathogenic *E. coli* (ExPEC) strains. One subset of these strains is called uropathogenic *E. coli* (UPEC), which are the main etiologic agent of urinary tract infections (UTIs) (Johnson and Russo, 2002; Yamamoto, 2007). ExPEC strains are also reported as a common cause of bacteraemia and sepsis in the community (Johnson and Russo, 2002). Other ExPEC strains are responsible for wound infections, surgical infections, neonatal meningitis and neonatal sepsis (Russo and Johnson, 2003; Wiles et al., 2008; Ron, 2010).

1.2 Intestinal pathogenic *E. coli*

Diarrhoea related mortality occurs worldwide, causing around 800,000 deaths every year (Liu et al., 2012). Children in developing countries of south Asia and sub Saharan Africa are the most affected, where it is estimated that one in ten children aged five and under die from diarrhoea (Liu et al., 2012). Intestinal pathogenic *E. coli* strains are major aetiological agents. Based on the pathogenesis and symptoms of intestinal pathogenic *E. coli* six major pathotypes have been described in the literature: Shiga toxin producing *E. coli* (STEC) which includes the subtype enterohaemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC) and the diffusely adherent *E. coli* (DAEC) (Croxen et al., 2013). A comprehensive study was conducted to investigate the incidence and causative agents of diarrhoeal disease of children in sub Saharan Africa and South Asia, named the Global Enteric Multicenter Study (GEMS). This study reported that ETEC is one of the main four aetiological agents of severe

to moderate diarrhoea in these children. Furthermore, there was an increased risk of death due to diarrhoea in infants up to 11 months of age, which was highly associated with ETEC and EPEC (Kotloff et al., 2013).

One important intestinal pathogenic *E. coli* subgroup is the serotype O157:H7 (O refers to the somatic lipopolysaccharide (LPS) antigens and H refers to the flagellar antigens), which has been reported to cause outbreaks of human diarrhoeal disease worldwide (Ateba and Mbewe, 2013). Consumption of contaminated undercooked ground beef is the main cause of infection, but other food products have also been identified as sources of human infection (Ateba and Mbewe, 2013). The first two reported outbreaks were in 1982, associated with undercooked hamburgers (Riley et al., 1983). Since then numerous outbreaks have occurred resulting in an estimated 61 deaths in the United States according to the Centers for Disease Control and Prevention (CDC) (Santos Mendonça et al., 2012). The clinical manifestations resulting from *E. coli* O157:H7 infection can vary, from watery to severe bloody diarrhoea, or haemorrhagic colitis and life threatening haemolytic ureamic syndrome (HUS). As such these *E. coli* strains belong to the EHEC subgroup of *E. coli*. EHEC are considered a subset from STEC, because the disease is caused by the production of toxins called Shiga like toxins (referring to *Shigella dysenteriae*) that is responsible for damaging the intestinal lining and other organs, such as the kidney (Armstrong et al., 1996; Waswa et al., 2007; Smith et al., 2013).

Another important intestinal pathogenic strain is EHEC O104:H4, which was responsible for the German gastroenteritis and HUS outbreak in 2011. This outbreak totalled 3816 cases in Germany with 22% suffering from HUS resulted in 54 deaths. Contaminated sprouts were identified as the vehicle of transmission (Frank et al., 2011; Buchholz et al., 2011). *E. coli* O104:H4 is an STEC pathotype producing Shiga toxin, but virulence factor profiling and whole genome sequencing revealed that the chromosomal backbone and some of its plasmids were related to the EAEC pathotype. These discoveries led to the conclusion that *E. coli* O104:H4

is a hybrid strain (Bielaszewska et al., 2011; Richter et al., 2014). This illustrates the plasticity and diversity of the *E. coli* genome and its continued evolution makes it very challenging to assign specific genes to a certain pathotype. New technologies such as whole genome sequencing are needed to produce more reliable and higher resolution typing of these pathogenic strains during outbreaks to discover the population dynamics of pathogenic *E. coli*. This would also enable the relationships of the *E. coli* strains to the host and environment to be studied, which would help improve our understanding of *E. coli* evolution (Croxen et al., 2013).

1.3 Extraintestinal pathogenic *E. coli* (ExPEC)

The ExPEC group of *E. coli* strains are able to stably colonize the intestinal tract without causing any enteric diseases. Based on multilocus enzyme electrophoresis (MLEE), which separates *E. coli* into four main phylogenetic groups (groups: A, B1, B2 and D)(Johnson and Russo, 2002), ExPEC belong to different phylogenetic groups than the commensal and enteric *E. coli* isolates. The ExPEC isolates belong to mainly group B2 and D, whilst the pathogenic intestinal *E. coli* are found in A, B1 or D, and the commensal *E. coli* belong to groups A or B1 (Johnson and Russo, 2002). The acquisition of a combination of specific virulence factors (VFs), enables ExPEC to cause disease when present in organs or sites outside the intestines (Table 1.1). The most common disease caused by ExPEC are UTIs, however, ExPEC can also cause sepsis or bacteraemia, neonatal meningitis, pneumonia, surgical site infections, soft tissue infections and osteomyelitis (Russo and Johnson, 2003; Smith et al., 2007).

Table 1. 1: General classification of pathogenic types of *E. coli* in humans

E. coli strains	Pathogenic type			MLEE phylogenetic groups (A, B1, B2 and D)
	Asymptomatic intestinal colonization	Diarrhea	Extraintestinal infection	
Commensal	+++	----	+	A or B1
Intestinal pathogen	----	+++	----	A, B1 or D
ExPEC	+	----	+++	B2 or D

The level of pathogenicity is expressed as absent (----) or maximum (+++), adapted from (Johnson and Russo, 2002).

ExPEC has been reported to constitute 20% of the faecal bacterial count of healthy individuals (Johnson and Russo, 2002), yet the exact origin of ExPEC colonization of the human intestine is unclear. Food was suggested as a source of ExPEC acquisition in several studies (Manges and Johnson, 2012). In a large sample study, 964 ExPEC isolates from UTI infections, healthy humans, broiler chicken, broiler chicken meat, pigs and pork were investigated for their phylogroups and acquisition of ExPEC related virulence associated genes (VAGs) (Jakobsen et al., 2010a; Jakobsen et al., 2010b). A total of 158 *E. coli* isolates from all sources (human, animals and animal products) were found to belong to phylogenetic group D with similar VAGs carried. The animal and animal-product isolates were also found to be able to cause UTI in a mouse infection model, suggesting animals and animal products may be possible routes of transmission (Jakobsen et al., 2010b). A zoonotic origin of ExPEC has also been documented in several studies. *E. coli* strains isolated from chickens and turkeys with colibacillosis are called avian pathogenic *E. coli* (APEC) (Jakobsen et al., 2010a). APEC has reported similarity with ExPEC, suggesting the zoonotic potential of ExPEC (Jakobsen et al., 2010a). Studies have shown that APEC and ExPEC isolates can be found in the same MLEE phylogroup (B2) and that there is no significant difference in VAG carriage (Johnson et al., 2008). Furthermore,

human ExPEC were found to be able to cause disease in chickens, indicating that there is no host specificity in ExPEC (Moulin-Schouleur et al., 2007). A study conducted by Johnson and Clabots in 2006 surveyed members of a household (5 humans and a dog), in which a woman had suffered from a UTI caused by ExPEC, for the presence of ExPEC clones using random amplified polymorphic DNA (RAPD). During the study, the first woman's ExPEC clone from acute cystitis was later extensively shared between all the members of the house, including the dog, suggesting host to host transmission and zoonosis (Johnson and Clabots, 2006). This conclusion was further investigated by Johnson et al. in 2008 with a larger study consisting of unrelated ExPEC isolates from humans (n = 55), dogs (n = 59) and cats (n = 16). 99% of the ExPEC belonged to the B2 phylogenetic group, but more importantly comparisons by macro restriction pulsed field gel electrophoresis (PFGE) revealed identical clones of ExPEC were shared between humans and dogs (Johnson et al., 2008), which further supported the zoonotic transmission hypothesis.

1.4 Pathogenesis of ExPEC in UTI and bacteraemia

ExPEC strains are able to cause disease due to the acquisition of VAGs that encode for virulence factors (VFs), such as fimbrial adhesins, toxins, iron acquisition proteins, secreted proteins and capsule formation. These VFs have essential roles in causing disease by enabling the ExPEC to colonize and invade mucosal cells, acquire essential nutrients and disrupt the host immune system (Johnson and Russo, 2002; Skjot-Rasmussen et al., 2012). The essential role that these VFs play in the establishment of ExPEC infection has been well established. For example, Type 1 fimbriae, which are encoded by the *fim* operon, is a well characterized VF involved in the colonization of the bladder during UTIs (Wullt et al., 2002). Ascending UTIs and urosepsis are reported to require functional P pili encoded by *pap* genes (Wright et al.,

2007). Acquisition of VFs associated with the pathogenicity island (PAI) is documented to increase ExPEC capability to cause systemic infections (Houdouin et al., 2006).

Typically UTIs are initiated by the colonization of the intestinal flora by an ExPEC strain, followed by a periurethral or vaginal colonization period before entering the urethra and invading the bladder (Stamey and Timothy, 1976; Brumfitt et al., 1987). The hypothesis of rectal flora origins of ExPEC and periurethral colonization prior to UTI is supported by the presence of the same ExPEC isolates from both fecal and urine samples from an individual with a UTI (Moreno et al., 2008). After entering the urethra, ExPEC strains attach to the mannose structures of uroplakin receptors on the surface of bladder epithelial cells by type 1 fimbriae (Sokurenko et al., 1998; Kau et al., 2005). This triggers apoptosis and exfoliation, with the bacteria moving to invade other cells and cause cystitis (Kaper et al., 2004). Type 1 fimbriae also trigger the internalization of some of the attached bacteria into the bladder cell by binding to α3 and β1 integrin receptors (Eto et al., 2007). This causes the formation of intracellular bacterial communities (IBCs) with biofilm like properties to form within the infected bladder cells. This protects the bacteria from the host immune response and may act as a source of recurrent infections (Anderson et al., 2004). VFs encoding iron acquisition proteins are essential for the survival of ExPEC in the urinary tract during infection (Henderson et al., 2009). After the invasion of the bladder cells, released ExPEC may ascend to the kidney via the ureters. Attachment to the kidney epithelial cells is mediated by binding of P fimbriae to digalactoside receptors (Virkola, 1987; Kaper et al., 2004). Once the ExPEC has colonized the kidney other VFs, such as haemolysin and the secreted autotransporter toxin (SAT), are involved in damaging the renal epithelium, which results in pyelonephritis (Guyer et al., 2002; Dhakal and Mulvey, 2012). At this stage, some bacteria can penetrate the endothelial cells of the proximal tubules and gain access to the bloodstream resulting in bacteraemia (Kaper et al.,

2004). This process of ascending infection from UTI to bacteraemia is termed urosepsis (Wagenlehner et al., 2007). Figure 1.1 displays the different steps of ExPEC pathogenesis.

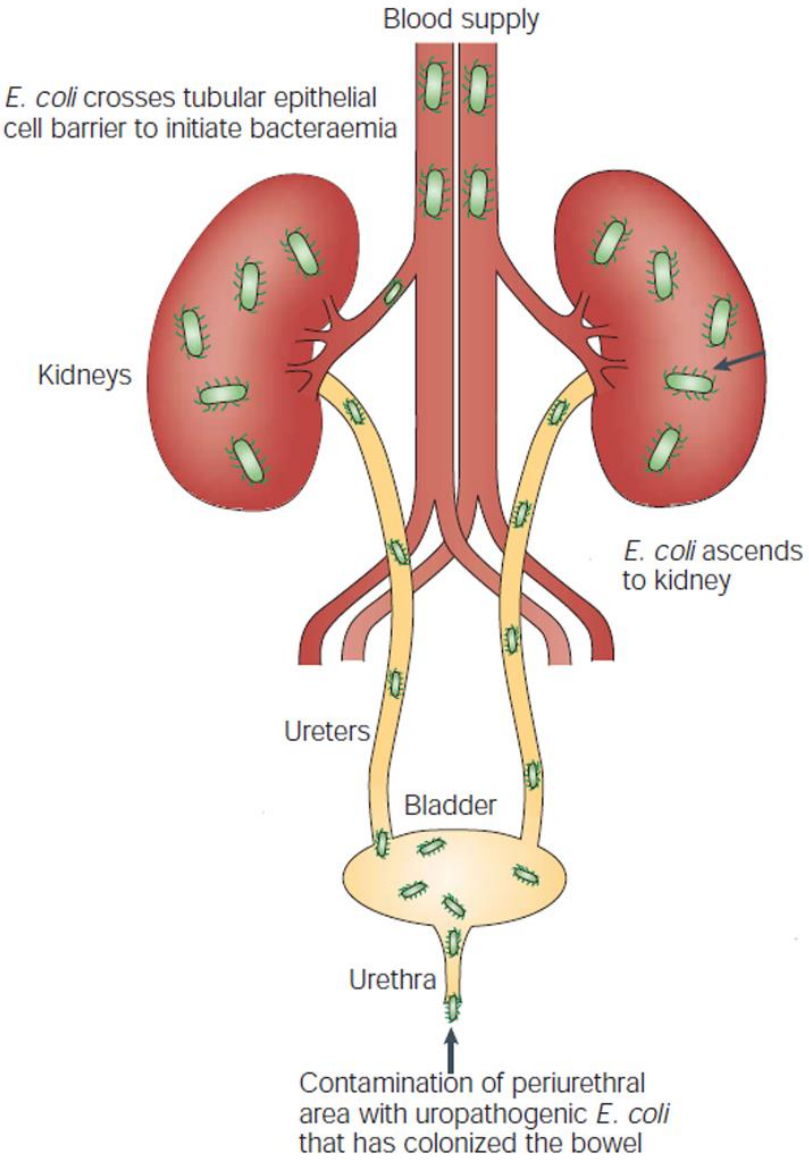


Figure 1. 1: Pathogenesis UTI caused by ExPEC. The figure displays the different stages of UTI caused by ExPEC starting by periurethral colonization, followed by invasion of the bladder and kidney epithelium and in some cases crossing the tubular epithelium resulting in bacteraemia (adapted from Kaper et al., 2004).

1.5 Main virulence factors associated with ExPEC

ExPEC possess a variety of virulence factors (VFs) encoded by virulence associated genes (VAGs) which enables them to effectively colonize and invade host cells and evade host defences (Johnson and Stell, 2000). Most VFs are encoded by mobile genetic elements called pathogenicity islands (PAIs) (Hacker and Kaper, 2000). These PAIs have a G-C content different from that of the remainder of the chromosome and contain regions of mobile DNA sequences inserted adjacent to tRNA genes. Horizontal transfer between PAIs has contributed to their evolution and ExPEC may have several PAIs in their genome (Oelschlaeger et al., 2002). PAIs may encode numerous VFs such as adhesins, iron uptake systems, toxins, capsules and secretion mechanisms (Hacker and Kaper, 2000).

1.5.1 Adhesins:

Adhesins are adhesive molecules that contribute to virulence by triggering host and bacterial cell signalling, promote delivery of bacterial products and promote invasion of host cells (Mulvey, 2002). Type 1 fimbriae are well characterized in animal infection models of UTIs. Encoded by *fim* genes, type 1 fimbriae promote attachment to host mucosal epithelial cells, survival and biofilm formation (Connell et al., 1996; Martinez et al., 2000; Anderson et al., 2003). Attachment is mediated by the FimH subunit that binds to the manosylated glycoprotein uroplakin of the urothelium (Thumbikat et al., 2009).

P fimbriae have been shown to have an essential role in human ascending UTIs and pyelonephritis (Väisänen et al., 1981). They promote the production of cytokines by attaching to the tissue matrix and mucosal surface (Leffler and Eden, 1980; Hedlund et al., 1999). P fimbriae are encoded by *pap* genes (A-K) and PapG recognizes the kidney epithelium glycosphingolipids (Wullt et al., 2000; Kaper et al., 2004). Type 1 fimbriae and P fimbriae

have been shown to play a synergistic role where P fimbriae enhances colonization of the tubular epithelium and type 1 fimbriae promotes colonization in the centre of the tubule resulting in complete obstruction of the nephron and pyelonephritis (Melican et al., 2011).

S fimbriae and F1C fimbriae are involved in binding to human epithelial cell lines of the lower urinary tract and kidney (Mulvey, 2002; Bien et al., 2012). Fimbrial Dr and Afa adhesins are described to have a role in recurring cystitis and gestational pyelonephritis (Servin, 2005). Dr fimbrial adhesin binds to decay accelerating factor (DAF) and type IV collagen (Nowicki et al., 2001) and displays a high tropism to the basement membrane of mouse renal interstitium (Goluszko et al., 1997). Afa adhesin shows a unique tropism to renal tissue and possibly plays a role in recurrent and chronic infection (Bouguénec, 2005).

1.5.2 Capsule and lipopolysaccharide (LPS)

The capsule and LPS are two of the surface virulence factors important for the survival of the bacterium. The capsule consists mainly of a polysaccharide structure covering the bacteria and protects it from complement mediated killing (Bien et al., 2012). K1 and K5 capsules have been described to help avoid the humoral immune response of the host by the molecular mimicking of tissue components and protect against killing by human serum (Buckles et al., 2009; Bien et al., 2012). LPS is known to cause acute renal failure due to systemic response to cytokine production by the host immune response (Backhed et al., 2001), rather than expression of functional LPS receptors (TLR4) in the kidney in an animal model (Cunningham et al., 2004).

1.5.3 Flagella

Flagellated ExPEC causes 70 to 90% of all UTIs and pathogenesis is mediated by interactions between the flagella and the surface of host epithelial cells. It has been shown that flagellin acts as an invasin to invade the renal collecting duct in pyelonephritis (Pichon et al., 2009). It has also been shown that motility provided by the flagella is important for ascending UTIs and antibodies specific to the flagella have prevented the spread of infection to the kidneys (Schwan, 2008).

1.5.4 Toxins

One of the most important secreted toxins by ExPEC during upper UTIs and pyelonephritis is the lipoprotein α-haemolysin (HlyA) (Bien et al., 2012). HlyA belongs to the RTX (repeats in toxin) toxin family that is present among many Gram-negative pathogens (Bhakdi, 1989). HlyA is a pore forming protein that has two different activities depending on its concentration (Laestadius et al., 2002). At high concentration α-haemolysin lyses erythrocytes, immune cells and mucosal epithelium that enables the bacteria to cross the mucosa, acquire host nutrients and iron reserves and damage the host defence system (Bien et al., 2012). At low concentration, α-haemolysin induces apoptosis of host immune cells such as lymphocytes, neutrophils, and renal cells which results in the exfoliation of host epithelial cells (Chen et al., 2006; Smith et al., 2006). About 50% of complicated renal infections are linked to HlyA production and may lead to permanent renal scarring (Jakobsson et al., 1994).

Cytotoxic necrotising factor 1 (CNF1) is reported to induce formation of actin stress fibres and Rho GTPase dependant membrane ruffle formation resulting in cell invasion. It is reported in one third of pyelonephritis isolates showing invasion of the kidneys (Landraud et al., 2000). It is also reported to cause apoptosis and exfoliation of epithelial cells in the bladder (Mills et al.,

2000). Secreted autotransporter toxin (SAT) is toxic to bladder and kidney cell lines indicating a role in pyelonephritis (Guyer et al., 2002).

1.5.5 Iron acquisition systems

ExPEC have different types of iron acquiring mechanisms to be able to survive and compete with host cells. They use iron chelators called siderophores, such as yersiniabactin (*fyuA*) and aerobactin (*iutA*) (Johnson et al. 2007). Yersiniabactin is also described to have a role in biofilm formation and ascending UTIs in a mouse infection model (Hancock et. Al., 2008; Spurbeck et. Al., 2012). Studies have shown an essential role of aerobactin in ExPEC iron acquisition *in vivo* (Garcia et. Al., 2011).

1.6 Incidence of UTI due to ExPEC

In clinical practice, bacterial infectious diseases are the most common causes of UTI (Foxman, 2003). In USA, the annual incidence of UTIs is estimated to be 12% in women and 3% in men (Foxman, 2003). *E. coli* is the predominant causative agent of UTIs (Yamamoto, 2007). The *E. coli* strains responsible for causing UTIs are described as a pathotype of ExPEC and termed as UPEC. UPEC are characterised by the acquisition of a set of VFs (Johnson and Russo, 2005). UPEC accounts for more than 68% of community acquired UTI and more than 35% of nosocomial UTI infections (Bouza et al., 2001; Kashef et al., 2010). The complications associated with UTIs are cystitis, pyelonephritis and bacteraemia (Warren et al., 1999). UPEC is reported to account for up to 90% of acute cystitis and quarter of a million cases of pyelonephritis in the USA per annum (Foxman, 2003). Worldwide UPEC is estimated to be responsible for 130 to 175 million cases of uncomplicated cystitis and 5.4 million cases of pyelonephritis every year (Russo and Johnson, 2003). In a UK study conducted by Woodford

et al. in 2004, *E. coli* causing UTI were collected from 42 UK centres, 24% of which were community acquired. Nosocomial *E. coli* UTI which are the most rapidly spreading cause of community acquired UTIs have a significant economic impact in the UK (Plowman et al 2001; Woodford et al., 2004).

1.7 Increasing incidence of bacteraemia due to ExPEC

The incidence of blood stream infection (bacteraemia) is becoming more common and is associated with increased levels of mortality (Wagenlehner et al., 2007; Jensen et al., 2010). In the USA, around 250,000 cases of bacteraemia are recorded annually and considered one of the leading causes of death with a mortality rate of 18% (Marschall et al., 2008). Bacteraemia due to *E. coli* infection, which can be community or hospital acquired, is increasingly reported worldwide (Ron, 2010). *E. coli* bacteraemia infections account for 17 to 37% of bacteraemia cases globally (Russo and Johnson, 2003), and is associated with a high mortality rate (Laupland et al., 2008). In the UK, the Health Protection Agency reported that *E. coli* bacteraemia infections between 2008 and 2010 increased by 33%, overtaking bacteraemia caused by methicillin resistant *Staphylococcus aureus* (MRSA) (Wilson et al., 2011). In response, the Department of Health extended mandatory reporting of *E.coli* bacteraemia infections in 2011, with the aim of reducing the rate of infections through increased awareness of circulating bacteraemia causing *E. coli* (Underwood et al., 2011). Analysis of the Guy's and St Thomas' NHS Foundation Trust during the year 2010 revealed that *E.coli* was isolated from 29% of total bacteraemia infections, of which the majority (63%) were community acquired (Underwood et al., 2011). According to the latest quarterly epidemiological commentary released by the Public Health England on December 2014, *E. coli* bacteraemia is still on the increase since 2011 with a rate of 70 persons per 100 000 of population affected (Table 1.1) (Public Health England, 2015b).

Table 1.1: Table of quarterly epidemiological commentary of *E. coli bacteraemia* by Public Health England.

Year and quarter (Q)		Total *E.coli* bacteraemia reports	Rate (per 100,000 population)
2011	**Q3**	8275	61.82
	Q4	8098	60.50
2012	**Q1**	7698	57.88
	Q2	8074	60.71
	Q3	8676	64.52
	Q4	7957	59.18
2013	**Q1**	7602	57.24
	Q2	8193	61.02
	Q3	9079	66.87
	Q4	8623	63.51
2014	**Q1**	8380	63.09
	Q2	8886	66.17
	Q3	9473	69.77

The table describes the Quarterly counts and rates of *E. coli* bacteraemia from the mandatory reports of *E. coli* bacteraemia by Public Health England (Public Health England, 2015b).

Studies have reported that *E. coli* is also the most frequent organism isolated from septicaemia, which has developed following UTI (McBean and Rajamani, 2001). A study in the North West of England reported that *E. coli* accounts for 74.9% of the Gram negative bacteraemia samples of UTI origin (Al-Hasan et al., 2010). The Guy's and St Thomas' NHS Foundation Trust reported that 49% of the *E. coli* bacteraemia were of UTI origin (Underwood et al., 2011). The association of *E. coli* bacteraemia with a UTI origin is still significant in the UK. The last annual epidemiological commentary released by Public Health England in 2014 reported that 50.4% of community acquired *E. coli* bacteraemia and 39.6% of the hospital acquired *E. coli* bacteraemia were of UTI origin (Figure 1.2) (Public Health England, 2015a).

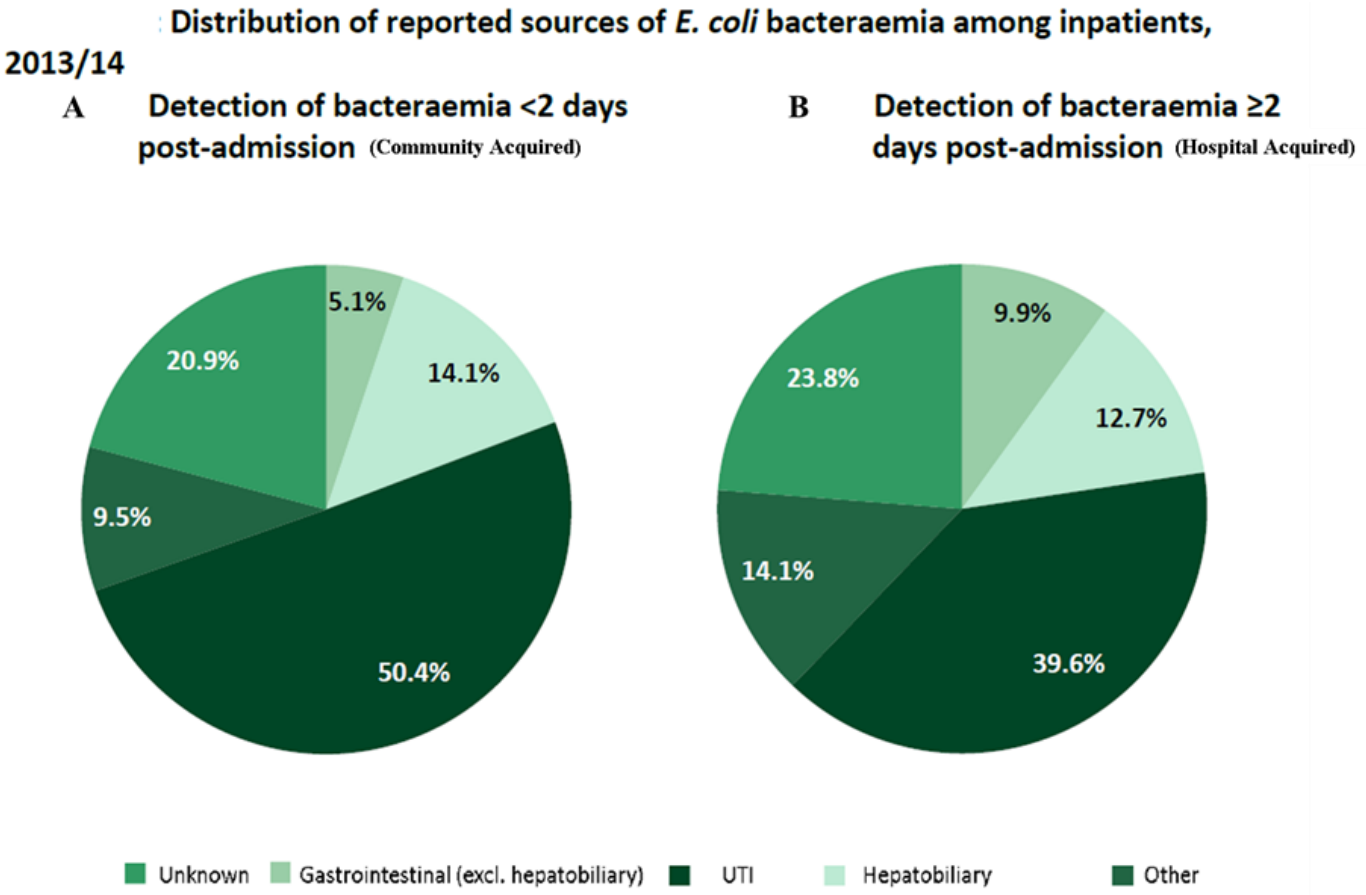


Figure 1. 2: Distribution of sources of *E. coli* bacteraemia in the UK in 2014. The figure obtained from the annual epidemiological commentary released by Public Health England on July 2014 (Public Health England, 2015a). A: UTI (dark green) accounts for 50.4% of community acquired bacteraemia (cultures obtained within 2 days of admission). B: UTI accounts for 39.6% of hospital acquired bacteraemia (cultures obtained in after 2 days of admission).

1.8 Extended spectrum β-lactamases (ESBLs)

ESBLs are defined as class A or D β-lactamases, which are able to inactivate oxyimino cephalosporins by hydrolytic activity with an active site serine, and inhibited generally by β-lactamase inhibitors such as sulbactam, tazobactam or clavulanic acid (Bush et al., 1995). Most ESBLs are encoded on large plasmids that can be transferred between strains, or even between bacterial species (Jacoby and Medeiros, 1991). ESBLs have evolved from amino acid substitutions in plasmid enzymes Temoniera (TEM, encoded by *bla*-$_{TEM}$), sulfhydryl variable (SHV, encoded by *bla*-$_{SHV}$) and oxacillinase (OXA, encoded by *bla*-$_{OXA}$). A novel, non TEM, non SHV, non OXA lineage has also developed, first discovered in Germany named CTX-M (CefoTaXime active in Munich, encoded by *bla*-$_{CTX\text{-}M}$) (Knothe et al., 1983; Kliebe et al., 1985; Sirot et al., 1987; Sougakoff et al., 1988). The intrinsic mutations in these enzymes caused major amino acid substitutions and alterations in the structure of the active site which increased their hydrolytic activities against third generation cephalosporins. Thus in addition to resistance against penicillins, first and second generation cephalosporins, resistance to oxyimino cephalosporins (ceftazidime, cefotaxime, ceftriaxone) and monobactams such as aztreonam is also observed. The Cephamycins and carbapenems (meropenem, imipenem, ertapenem) on the other hand are generally unaffected (Stürenburg and Mack, 2003).

1.9 Increasing incidence of ESBL producing *E. coli*

The incidence of CTX-M has rapidly surpassed that of the classical TEM and SHV ESBLs worldwide (Livermore et al., 2007). CTX-M ESBL producing members of the *Enterobacteriaceae* family are to cause hospital and community acquired infections (Pitout et al., 2005). However, since 2000 CTX-M producing *E. coli* has emerged as a leading cause of community acquired UTI globally; this has been termed the CTX-M pandemic (Cantón and

Coque, 2006; Rossolini et al., 2008). CTX-M-15 is the CTX-M ESBL most widely distributed worldwide, since the first report in 2001(Karim et al., 2001). Since then, CTX-M-15 producing multidrug resistant (MDR) *E. coli* is emerging as an important cause of community, or hospital acquired infections worldwide (Woodford et al., 2004; Coque et al., 2008; Nicolas-Chanoine et al., 2008).

An established *E. coli* multilocus sequence typing (MLST) scheme is currently used worldwide to study the population dynamics of pathogenic ExPEC strains. Several MLST sequence types (STs) such as ST95, ST73, ST131, ST69, ST127 represent well characterized, commonly isolated human ExPEC strains (Lau et al., 2008b, Johnson et al., 2008b, Johnson et al., 2008a, Mora et al., 2009). The most well recognised ST is *E. coli* ST131, which is increasingly reported in cases of UTIs worlwide (Nicolas-Chanoine et al., 2008). Recently, *E. coli* ST131 has caused increasing concern in the health care community, as it is resistant to most commonly used antibiotics and associated with CTX-M-15 carriage (Lau et al., 2008a; Nicolas-Chanoine et al., 2008; Pitout and Laupland 2008). *E. coli* ST131 is a major contributor to the CTX-M pandemic (Totsika et al., 2011). The *bla*$_{\text{-CTX-M}}$ gene is carried on IncFII plasmids in *E. coli* ST131 (Coque et al., 2008), however other types of plasmids such as FII, FIA and FIB were also associated with ST131 and ExPEC ESBL carriage (Rogers et al., 2011). *E. coli* ST131 is also frequently resistant to fluoroquinolones (Johnson et al., 2009; Johnson et al., 2010). In addition to mutations in chromosomal *gyrA* and *parC* gene, fluoroquinolone resistance can be mediated by plasmid encoded quinolone resistance genes *qnr (A, B and S)* and *aac(6')-Ib-cr* (aminoglycosides/fluoroquinolone acetyltransferase) (Baudry et al., 2009). Many studies in the UK have reported a high prevalence of CTX-M-15 producing ST131 *E. coli* in hospital and community onset UTIs (Gibreel et al., 2011; Croxall et al., 2011). In addition to ST131, other sequence types were reported to have a role in the worldwide distribution of CTX-M, such as: ST38, ST405 and ST648 (Peirano and Pitout, 2010).

Reports of multidrug resistant ESBL producing *E. coli* are on the increase worldwide. A 4 year study in Spain investigating incidence of bacteraemia reported that 70% of *E. coli* bacteraemia isolates produced a CTX-M type ESBL, of which 46% were of urinary origin (Rodriguez-Bano et al., 2006). In China, 24.2% of bacteraemia *E. coli* strains were ESBL positive (To et al., 2013). In Turkey, ESBL producing *E. coli* accounted for 38.9% of *E. coli* bacteraemia cases between 2007 and 2011 (Kaya et al., 2013). The increasing prevalence of ESBL producing bacteraemia *E. coli* raises concerns and challenges in the medical community (Peirano and Pitout 2010). A study was performed to investigate and compare ESBL producing to non-ESBL producing bacteraemia *E. coli* isolates between June 2003 to November 2005 in the UK (Melzer and Petersen, 2007). ESBL producing *E. coli* were reported in 6.6% of community and in 26% of hospital acquired bacteraemia infections where the most common source was UTI's, which accounted for 67% of the cases (Melzer and Petersen, 2007). This study revealed that a significantly higher percentage of patients with ESBL producing *E. coli* bacteraemia (60.8%) died compared to the percentage of deaths in patients resulting from a non-ESBL producing *E. coli* bacteraemia (23.7%) (Melzer and Petersen, 2007). In a more recent investigation of *E. coli* bacteraemia in the UK by Wilson et al. (2011), a significant proportion of ESBL producing *E. coli* isolates were identified as the infecting agents. ESBL producing *E. coli* isolates were reported in 16% of *E. coli* bacteraemia cases, 55% of which were of community onset and 35% were of urinary origins (Wilson et al., 2011).

1.10 Population structure of ExPEC

MLST is a molecular typing method based on the DNA sequencing of approximately 500 bp from seven unlinked housekeeping genes, which are used to identify bacterial strains (Maiden et al., 1998). Due to the reproducibility and comparability between different laboratories, MLST is considered the gold standard for ExPEC genotyping (Tartof et al., 2005). The

'Achtman' scheme is the established MLST typing scheme for *E. coli* (Wirth et al., 2006), and the database is available via a publically accessible website (www.mlst.warwick.ac.uk/mlst/dbs/Ecoli). In the MLST database, strains that share the same seven housekeeping gene alleles are identified as a unique ST (Wirth et al., 2006). A minimum spanning tree (MST) of the structure of the ExPEC population using the results from the MLST scheme as of January 2011 was produced by Köhler and Dobrindt (2011) in Figure 1.3. The MST uses eBURST algorithms to identify strains which differ in only one of the seven housekeeping gene alleles (single locus variants) into groups called a clonal complex (CCs) (Feil et al., 2004). The MST contained 3454 entries consisting of 32% ExPEC isolates (1117 entries), 27% intestinal *E. coli* isolates (922 entries), 16% non-pathogenic isolates (535 enteries) and the rest are isolates that do not belong to any of these groups. Of the total of 1859 STs in the MLST database (as of January 2011), 114 CCs were determined consisting of 1197 STs (containing 2610 entries). The clinical source of the isolates was mapped on the MST and the vast majority of the ExPEC STs and CCs also included other pathotypes and non-pathogenic strains (Figure 1.3).

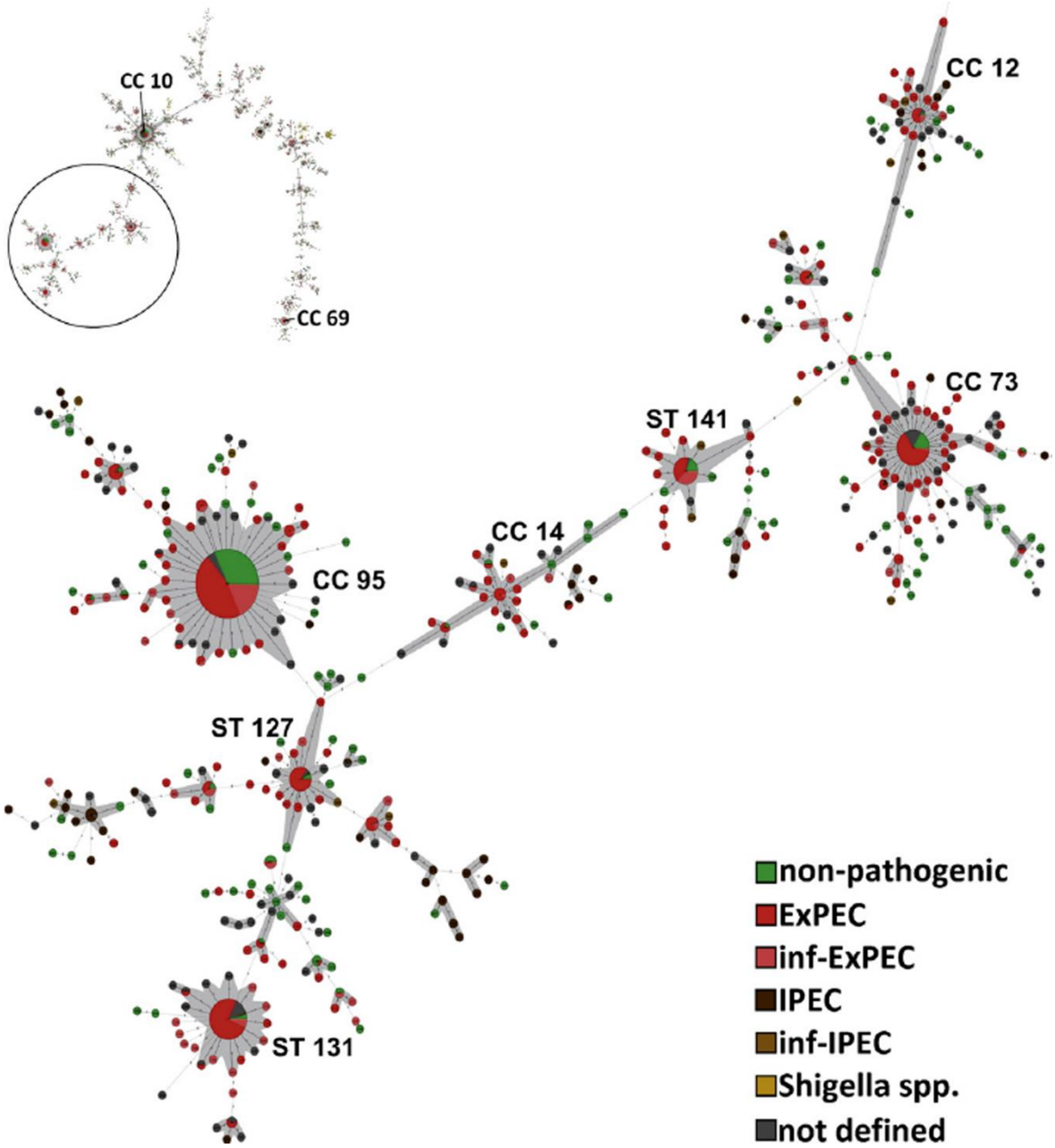


Figure 1. 3: Minimum spanning tree presenting the population structure of ExPEC isolates. The Figure was produced by Köhler and Dobrindt, (2011) with reference to Achtman MLST database (www.mlst.ucc.i.e./mlst/dbs/Ecoli) as of January 2011 (Köhler and Dobrindt, 2011). In the Figure (inf) refers to clinical isolates from different infections and (IPEC) refers to intestinal pathogenic *E. coli*. The main STs and CCs of ExPEC determined by the circle on top are enlarged with pie charts mapping the clinical source. The size of the pie chart reflects the number of isolates included in the STs or CCs.

CC10 is located in the middle of the MST. It represents the largest CC in the database including many intestinal pathogenic *E. coli*, non-pathogenic *E. coli*, non-defined isolates, and only a small proportion of ExPEC isolates. To the right of the CC10 are CCs comprised of predominantly intestinal pathogenic *E. coli* and non-pathogenic isolates and others. At the far right of the tree is CC69, which mainly consists of ExPEC isolates. On the left of CC10 are the STs and CCs that include mainly ExPEC and non-pathogenic isolates. These major ExPEC CCs and STs (95, 73, 131, 127, 17, 14, 144 and 12) are extensively reported in the literature as pathogenic ExPEC STs (Tartof et al., 2005; Wirth et al., 2006; Zdziarski et al., 2008; Peirano and Pitout, 2010; Croxall et al., 2011; Rogers et al., 2011;). These STs and CCs comprise the largest number of isolates in the Achtman MLST database, 47.7% of CC95 are ExPEC isolates, which comprise 11.6% of the total ExPEC in the database, 62.7% of CC73 are ExPEC isolates, which comprise 8.4% of the total ExPEC in the database and 66.3% of CC131 are ExPEC isolates which comprise 7.5% of ExPEC in the database. Almost all the isolates in the database of this ExPEC branch of the MST belongs to the B2 phylogenetic group which confirms the ExPEC pathotype, while ST69 isolates belongs to phylogenetic group D hence they are located to the far right of the MST (Köhler and Dobrindt, 2011). The main ExPEC STs and lineages such as (ST95, ST73, ST131) are well characterized in the literature in terms of the acquisition of VFs, which enables them to be virulent extraintestinal pathogens compared to the other CCs in the *E. coli* population (Manges et al., 2008; Martinez-Medina et al., 2009; Johnson et al., 2010).

1.11 Most common STs of ExPEC

1.11.1 *E. coli* ST131

In the early 2000s PFGE analysis was performed on two *E. coli* isolates that produced CTX-M-15 ESBL, one from Canada and one from the UK (Woodford et al., 2004; Pitout et al., 2007). The PFGE identified more than 80% pulsotype cluster similarities, and were named clone A in the UK and clone 15A in Canada (Woodford et al., 2004; Pitout et al., 2007). With the development of MLST these strains were then identified as belonging to ST131, which is characterized by serotype O25b:H4 and MLEE phylogenetic group B2 (Coque et al., 2008; Nicolas-Chanoine et al., 2008). Other serotypes have also been included in ST131, such as O16:H5 and non typable O antigen (Matsumura et al., 2012; Dahbi et al., 2013; Olesen et al., 2013). In Australia, an ST131 strain serotype O157 was also reported (Platell et al., 2011a). Most of the increased antibiotic resistance in ExPEC infections, especially ESBL producing and fluoroquinolone resistant ExPEC, is attributed to spread of ST131 (Rogers et al., 2011). Resistance is mediated mainly by ESBL type CTX-M-15 encoded by $bla_{\text{CTX-M-15}}$ gene on plasmids while fluoroquinolone resistance is due to mutations in chromosomal genes *gyrA* and *parC* (Cagnacci et al., 2008; Nicolas-Chanoine et al., 2008; Johnson et al., 2013). Other CTX-M types are also associated with ST131, such as CTX-M-2, CTX-M-3 and CTX-M-9 (Suzuki et al., 2009; Mora et al., 2010).

The origins of ST131 are unknown. Some studies reported that ST131 strains were identified as early as 1994 in France (Clermont et al., 2009). ST131 *E. coli* can be isolated from community and hospital infections, but the source of infection is not well characterised. In a study at a general hospital in San Francisco between 2007 and 2010, ST131 *E. coli* infections were identified in 23% of *E. coli* bacteraemia patients, but more than 70% from patients within 48 hours of admission, which indicates community acquired origins (Adams-Sapper et al., 2013). In the UK, ST131 *E. coli* was the most dominant ST associated with ESBL CTX-M-15

(Croxall et al., 2011). Intestinal carriage of ST131 was also reported in livestock, wild and domestic animals (Platell et al., 2011b). Several main VAGs are described in association with ST131 clinical isolates, *iha* (siderophore associated), *fimH* (type I fimbriae), *sat* (toxin), *fyuA/irp2* and *iutA/iucD* (siderophore associated), *kpsM II* (capsule), *usp* (uropathogenic specific protein), *traT* (serum resistance), *ompT* (protease of outer membrane) and *malX* (pathogenicity island) (Johnson et al., 2010; Platell et al 2011a; Matsumura et al., 2012; Dahbi et al., 2013).

1.11.2 *E. coli* ST69

In California between January 1999 and 2000, 255 *E. coli* strains were isolated from 228 clinical cases of community acquired UTIs (Manges et al., 2001). Of these, 55 isolates were found resistant to trimethoprim-sulphamethazole. When analysed by enterobacterial repeat intergenic consensus (ERIC2) PCR fingerprinting, 51% of these resistant isolates were found to have the same banding pattern (Johnson and O'Bryan, 2000) and were designated the name clonal group A (CgA) (Manges et al., 2001). This clonal group was then identified as ST69 complex by MLST (Tartof et al., 2005). ST69 *E. coli* strains belong to the phylogenetic group D and include many serotypes of O antigen variants including: O11, O15, O73, O77, O86, O125ab and O25b (Manges et al., 2008; Blanco et al., 2011; Colomer-Lluch et al., 2013; Skjøt-Rasmussen et al., 2013). ST69 strains are isolated worldwide from UTIs and bacteraemia from both hospital and community acquired cases (Dias et al., 2009; Johnson et al., 2009; Adams-Sapper et al., 2013). Most ST69 have a class I integron containing the arrangement of (*dfrA17 – aadA5*) gene cassette encoding dihydrofolate reductase and aminoglycoside adenyltransferase that makes them multidrug resistant (Solberg et al., 2006; Ajiboye et al., 2009). In a study of *E. coli* bacteraemia isolates collected in San Francisco between 2007 and 2010, ST69 complex was found to be the fourth most prevalent ExPEC ST after ST131, ST95

and ST73 (Adams-Sapper et al., 2013). 83% of all these ST69 *E. coli* cultures were obtained less than 48 hours after admission, which suggests they are of community acquired origin. ST69 *E. coli* isolates may have nonhuman origins as they have been isolated from pork, chicken and beef (Ramchandani et al., 2005; Vincent et al., 2010; Jakobsen et al., 2011). Nonhuman ST69 isolated from chicken meat were found to be as pathogenic as human ST69 isolates and caused UTIs in a mouse infection model, which may suggest a zoonotic origin of ST69 (Jakobsen et al., 2010). A variety of VAGs are found associated with ST69 complex isolates, *afa/dra* (Dr adhesins), K1 and *kpsMT* (capsule), *papA* and *papG* alleles (P fimbriae adhesion) and *sfa/focDE* (S and FIC fimbriae) (Johnson et al., 2002; Johnson et al., 2005).

1.11.3 *E. coli* ST95

E. coli ST95 strains belong to phylogenetic group B2 and include serotypes O1/O2/O18:K1:H7 (Weissman et al., 2006). ST95 *E. coli* K1 capsule serotypes are associated with neonatal meningitis (Glode et al., 1977; Tivendale et al., 2010). ST95 *E. coli* strains also include APEC, which cause colibacillosis in wild and domestic birds (Blanco et al., 1998; Mora et al., 2009). ST95 *E. coli* isolates were the second most prevalent ST complex isolated from bacteraemia *E. coli* infections from a general hospital in San Francisco between 2007 and 2010 (Adams-Sapper et al., 2013). In France, ST95 complex was the most common *E. coli* ST isolated from blood and ascitic fluid between 1997 and 2006 (Bert et al., 2010). ST95 complex isolates are characterised by low multidrug resistance frequency. More than half of the ST95 *E. coli* isolates were sensitive to all antibiotics in a San Francisco study (Adams-Sapper et al., 2013). One of four ST95 *E. coli* isolated from UTIs was found resistant to any antibiotic in Montreal between 2007 and 2009 (Manges et al., 2008). In North-West England, ST95 *E. coli* isolates from UTIs presented the lowest antibiotic resistance score of the nine common *E. coli* ST complexes reported (Gibreel et al., 2011). ST95 *E. coli* complex may have zoonotic potential, as ST95

isolates from avian colibacillosis were able to cause neonatal meningitis in a rat infection model (Tivendale et al., 2010). At the same time, human ST95 *E. coli* isolates from neonatal meningitis were able to cause colisepticemia in poultry (Moulin-Schouleur et al., 2007). Many VAGs are described for ST95 isolates which encode for type I fimbriae (*fimH*), P fimbriae (*papG*), siderophore system (*iucD*), capsule (*kpsM II* and *II-K1*), serum resistance (*traT*), pathogenicity island (*malX*) and uropathogenic specific protein (*usp*) (Manges et al., 2008; Bert et al., 2010; Mora et al., 2013).

1.11.4 *E. coli* ST73

E. coli ST73 isolates belong to phylogenetic group B2 and are only associated with serotype O6:H1 (Johnson et al., 2008; Martinez-Medina et al., 2009). In the UK, ST73 *E. coli* isolates were the most common ST from UTI (Gibreel et al., 2011) and bacteraemia (Horner et al., 2014), and second dominant in UTI (Croxall et al., 2011). It was the third most common *E. coli* ST isolated from bacteraemia in a general hospital in San Francisco and ten hospitals in France (Brisse et al., 2012; Adams-Sapper et al., 2013). In Egypt, ST73 *E. coli* isolates were reported as one of the most common ExPEC STs associated with CTX-M-15 ESBL carriage (Fam et al., 2011). The pyelonephritis reference strain *E. coli* CFT073, which is used in animal models and ExPEC virulence studies, is an ST73 strain isolated from the blood of a woman with pyelonephritis (Johnson et al., 2008). Prototypic VAGs of ExPEC associated with ST73 *E. coli* isolates encode for P fimbriae adhesion (*papC*, *papEF*, *papG*), S and FIC fimbriae adhesin (*sfa/foc*), type I fimbriae (*fimH*), yersiniabactin receptor (*fyuA*), capsule (*kpsM II*) and outer membrane protease (*ompT*) (Martinez-Medina et al., 2009; Bert et al., 2010).

.

1.11.5 *E. coli* ST10

E. coli ST10 strains and the closely related STs in ST10 CC belong to phylogenetic group A, which is associated with colonization of the human intestine. There are numerous serotypes associated with ST10. They are usually associated with low virulence and antibiotic susceptibility (Manges and Johnson, 2012). Few ST10 *E. coli* isolates have been associated with human disease, ESBL carriage and livestock (Oteo et al., 2009; Cortes et al., 2010; Peirano et al., 2012). In the Netherlands, ST10 *E. coli* were isolated from human and poultry blood cultures and were CTX-M-1 producers (Leverstein-van Hall et al., 2011). In Canada, multidrug resistant ST10 *E. coli* strains were isolated from human clinical samples, chickens faeces, retail chicken, pig faeces and pork meat (Bergeron et al., 2012).

1.12 Aims of the project

Bacteraemia infections as a result of multi-drug resistant extra-intestinal pathogenic *E. coli* have overtaken MRSA as the leading cause of health care associated blood stream infections in the world. More concerning is the huge rise in associated antimicrobial resistant infections as a result of the increasing prevalence of ESBL carriage among ExPEC isolates. Despite ESBL ExPEC infections now being a major global health threat, and ESBL carriage in bacteraemia ExPEC isolates being on the increase, there are still few studies on the population genetics of ExPEC bacteraemia isolates, in comparison to the abundance of literature on urinary tract infections, which are often the source of such *E. coli* bacteraemia. Furthermore numerous studies have demonstrated the enormous diversity within ExPEC populations associated with UTIs, but with little information on bacteraemia isolates. We sought to define the ExPEC population causing bacteraemia by collecting clinical isolates from the largest clinical microbiology laboratory service provider in the United Kingdom, the Nottingham University

Hospitals microbiology laboratory. Isolates were collected over a 5 month period, alongside urine sample isolates from the concomitant period to provide a reference point for circulating ExPEC strain population characteristics. The general aims of the study are:

A) Screening and testing the clinical *E. coli* isolates from bacteraemia and UTIs for antibiotic resistance, ESBL and virulence gene carriage.
B) Determining the genetic structure of bacteraemia and UTI *E. coli* populations by MLST.
C) Comparing the two populations of *E.coli* isolates to provide a comprehensive snapshot of the current ExPEC genetic epidemiology in bacteraemia and UTIs in the region.
D) Determining and comparing the gene content of a selected group of isolates from bacteraemia and UTIs to identify unique genetic loci, or regions specific for bacteraemic *E. coli*.

Chapter two

Materials and Methods

2.0 General Materials and Methods

2.1 Collection of isolates

In the period between March 2011 and July 2011, one hundred and forty *E.coli* isolates from clinical cases of bacteraemia and another one hundred and twenty five *E.coli* isolates from clinical cases of UTI were collected from the Clinical Microbiology Department at Nottingham University Hospital. The selection of the cases was performed at random with no additional information except the age of the patients. Isolates were assigned a serial number as they were collected with letter B for bacteraemia clinical isolates and letter U for UTI clinical isolates. On average, 10 *E.coli* UTI isolates and 5 *E. coli* bactereamia isolates were collected per week.

2.1.1 Bacterial cultures maintenance and storage

All the *E.coli* isolates from clinical cases of bacteraemia and UTIs were collected as nutrient agar slopes and were sub-cultured on Cystine-Lactose-Electrolyte Deficient Agar (CLED, Sigma-Aldrich) to confirm purity of the isolates. Pure cultures were assigned a number with letter B for bacteraemia isolates, or U for UTI isolates. Stocks of all the bacterial cultures were maintained in 1ml of Luria Bertani Broth (LB) and 20% glycerol (Fisher Scientific) and stored at -80˚C.

2.1.2 Species Confirmation of Bacterial Cultures

All isolates were confirmed as *E. coli* species by API 20E identification kit (BioMérieux).

2.2 Culture media

2.2.1 Luria-Bertani (LB) Agar

The media was ordered from Sigma-Aldrich. It consists of 5g/L yeast extract, 10g/L sodium chloride, 10g/L tryptone, and 15g/L granulated agar. 40g of LB agar was dissolved in 1L distilled water (dH_2O) and sterilised by autoclaving for 15 minutes at 121°C.

2.2.2 LB broth

The media was ordered from Sigma-Aldrich. It consists of 5g/L yeast extract, 10g/L sodium chloride, 10g/L tryptone, and 15g/L granulated agar. 20g of LB broth was dissolved in 1L (dH_2O) and sterilised by autoclaving for 15 minutes at 121°C.

2.2.3 Cystine Lactose Electrolyte Deficient (CLED) agar

The media was ordered from Oxoid Limited UK. It consists of 0.128g/L L-cystine, 10g/L lactose, 4g/L tryptone, 0.02g/L bromothymol blue and 4g/L peptone. 36.2g was dissolved and sterilised by autoclaving for 15 minutes at 121°C.

2.3 Antibiotic Susceptibility profiles

Antibiotic resistance tests of all the *E. coli* isolates from bacteraemia and UTI samples were performed with reference to the protocol for standardized disc susceptibility testing method provided by The British Society for Antimicrobial Chemotherapy (BSAC) (Andrews and Howe, 2011). A panel of ten antibiotics were selected described in Table 2.1.

Table 2. 1: Antibiotics details used in the standardized disc susceptibility test

Antibiotic Group	Antibiotic Name	Disc concentration (µg)
Aminoglycoside	Gentamicin (GM)	10µg
Third Generation Cephalosporin	Ceftazidime (CAZ)	30µg
Carbapenem	Meropenem (MEM)	10µg
Combination of penicillins and β-lactamase inhibitor	Piperacillin-Tazobactam (PTZ)	75/10µg
Combination of β-lactamase inhibitors	Co-Amoxiclav (Augmentin) (AUG)	20/10µg
Dihydrofolate reductase inhibitors	Trimethoprim (TM)	2.5µg
Second generation fluoroquinolone	Ciprofloxacin (CIP)	1µg
First generation cephalosporin	Cefradine (CRD)	30µg
Miscellaneous	Nitrofurantoin (NI)	200µg
Penicillins	Ampicillin (AP)	10µg

All the *E. coli* isolates were tested for their ability to produce extended spectrum *β* lactamases (ESBL) by using ESBL combination discs provided by MAST GROUP Ltd. Control Strains *E. coli* NCTC 13353, *E. coli* NCTC 13351, *E. coli* NCTC 10418 and *Klebsiella pneumoniae* NCTC 13368 were used as controls. The isolates were sub-cultured into LB agar and incubated overnight at 37ºC. Inoculums were prepared by transferring four identical colonies of each isolate into saline solution to yield a density equal to 0.5 McFarland Standard. Cotton swabs were dipped in the inoculums and used to inoculate Iso-sensitest agar plates within 15 minutes of preparation of the inoculums. A maximum of six antibiotic discs were applied per plate and the cultures were incubated at 37ºC for 18-20 hours. Zones of inhibition of growth around the antibiotic discs for each isolate were then measured and referred to the BSAC tables to determine their antibiotic sensitivity.

2.4 PCR screening of extended-spectrum β-lactamase (ESBL) encoding genes

Bacterial genomic DNA was extracted using GenElute extraction kits (Sigma-Aldrich). PCR detection of ESBL encoding genes *bla*$_{OXA}$, *bla*$_{CTX-M}$, *bla*$_{TEM}$ and *bla*$_{SHV}$ was performed using previously described multiplex primers (Fang et al., 2008). Control cultures were used to confirm the validity of the primers. These control strains were NCTC 13353 *E. coli* (*bla*$_{OXA}$, *bla*$_{CTX-M-15}$, and *bla*$_{TEM}$), NCTC 13351 *E. coli* (*bla*$_{TEM-3}$) and NCTC 13368 *Klebsiella pneumoniae* (*bla*$_{TEM}$ and *bla*$_{SHV-18}$). The primer sequences, the PCR reagents concentrations and the PCR conditions are described in Table 2.2, Table 2.3 and Table 2.4 respectively. PCR products were run on 2% agarose gels at 90 volt for 50 minutes and the ESBL gene carried was determined for each isolate. All the CTX-M positive PCR products of the isolates were sequenced. The CTX-M type of all the isolates in this study was CTX-M-15.

Table 2. 2: Multiplex ESBL PCR primer sequences and their product size

Target Gene	Forward and reverse primer names and sequences	Product size
***bla*$_{SHV}$**	shvF 5'-CTTTATCGGCCCTCACTCAA-3' shvR 5'-AGGTGCTCATCATGGGAAAG-3'	237 bp
***bla*$_{TEM}$**	temF 5'-CGCCGCATACACTATTCTCAGAATGA-3' temR 5'-ACGCTCACCGGCTCCAGATTTAT-3'	445 bp
***bla*$_{CTX-M}$**	ctxmF 5'-ATGTGCAGYACCAGTAARGTKATGGC-3' ctxmR 5'-TGGGTRAARTARGTSACCAGAAYCAGCGG-3'	593 bp
***bla*$_{OXA}$**	oxaF 5'-ACACAATACATATCAACTTCGC-3' oxaR 5'-AGTGTGTTTAGAATGGTGATC-3'	813 bp

Table 2. 3: PCR reagents concentrations and quantities for ESBL multiplex PCR

Reagent	**Final Concentration**	**Quantity/reaction (µl)**
5X PCR reaction buffer	5X	5.0
20mM dNTPs	0.2mM	0.25
25mM $MgCl_2$	2.5 mM	2.5
5 U/µM Taq DNA polymerase	1.25 U/µM	0.25
Sterile dH_2O		11
2.5 pmol Forward Primer	0.2 pmol	2.0
2.5 pmol Reverse Primer	0.2 pmol	2.0
DNA	~3 ng/ µl	2.0
Total		25

Table 2. 4: PCR conditions for ESBL multiplex PCR

Step	**Temperature**	**Time**
Preheat lid	95ºC	1 min
Initial denaturation	94ºC	5 min
30 cycles of steps 1,2,3		
1) Denaturation	94ºC	30 sec
2) Annealing	62ºC	90 sec
3) Elongation	72ºC	60 sec
Final elongation	72ºC	10 min
Final hold	10ºC	∞

2.5 PCR detection of Virulence Associated Genes (VAGs) carriage

E. coli isolates from the clinical samples of bacteraemia and UTIs were screened for the presence of VAGs by multiplex PCR with reference to a previously published protocol (Johnson and Stell, 2000). The protocol describes thirty different VAGs associated with pathogenicity islands, adhesins, toxins and others that are divided in 5 multiplex PCR pools. Primer sequences, PCR reagent concentrations and PCR conditions are described in Table 2.5, Table 2.6 and Table 2.7 respectively. PCR products were run on a 2% agarose gel and VAGs carriage was determined for each isolate.

Table 2. 5: VAGs multiplex PCR primer sequences and PCR product sizes

Gene/function	Primer Name	Primer sequence	size (bp)
adhesins			
papAH	PapA f PapA r	5'-ATGGCAGTGGTGTCTTTTGGTG-3' 5'-CGTCCCACCATACGTGCTCTTC-3'	720
papC	PapC f PapC r	5'-GTGGCAGTATGAGTAATGACCGTTA-3' 5'-ATATCCTTTCTGCAGGGATGCAATA-3'	200
papEF	PapEF f PapEF r	5'-GCAACAGCAACGCTGGTTGCATCAT-3' 5'-AGAGAGAGCCACTCTTATACGGACA-3'	336
papG **II, III** **I(flanking region)**	pGf pGr pG1"r	5'-CTGTAATTACGGAAGTGATTTCTG-3' 5'-ACTATCCGGCTCCGGATAAACCAT-3' 5'-TCCAGAAATAGCTCATGTAACCCG-3'	1070 1190
allele I	AlleleI-f AlleleI-r	5'-TCGTGCTCAGGTCCGGAATTT-3' 5'-TGGCATCCCCCAACATTATCG-3'	461

allele I'	AlleleI'-f AlleleI'-r	5'-CTACTATAGTTCATGCTCAGGTC-3' 5'-CTGACATCCTCCAACATTATCGA-3'	474
allele II	AlleleII-f AlleleII-r	5'-GGGATGAGCGGGCCTTTGAT-3' 5'-CGGGCCCCCAAGTAACTCG-3'	190
allele III	AlleleIII-f AlleleIII-r	5'-GGCCTGCAATGGATTTACCTGG-3' 5'-CCACCAAATGACCATGCCAGAC-3'	258
sfa/focDE	sfa1 sfa2	5'-CTCCGGAGAACTGGGTGCATCTTAC-3' 5'-CGGAGGAGTAATTACAAACCTGGCA-3'	410
sfaS	SfaS f SfaS r	5'-GTGGATACGACGATTACTGTG-3' 5'-CCGCCAGCATTCCCTGTATTC-3'	240
focG	FocG f FocG r	5'-CAGCACAGGCAGTGGATACGA-3' 5'-GAATGTCGCCTGCCCATTGCT-3'	360
afa/draBC	Afa f Afa r	5'-GGCAGAGGGCCGGCAACAGGC-3' 5'-CCCGTAACGCGCCAGCATCTC-3'	559
bmaE	bmaE-f bmaE-r	5'-ATGGCGCTAACTTGCCATGCTG-3' 5'-AGGGGGACATATAGCCCCCTTC-3'	507
gafD	gafD-f gafD-r	5'-TGTTGGACCGTCTCAGGGCTC-3' 5'-CTCCCGGAACTCGCTGTTACT-3'	952
nfaE	nfaE-f nfaE-r	5'-GCTTACTGATTCTGGGATGGA-3' 5'-CGGTGGCCGAGTCATATGCCA-3'	559
fimH	FimH f FimH r	5'-TGCAGAACGGATAAGCCGTGG-3' 5'-GCAGTCACCTGCCCTCCGGTA-3'	508
Toxins			
hlyA	hly f hly r	5'-AACAAGGATAAGCACTGTTCTGGCT-3' 5'-ACCATATAAGCGGTCATTCCCGTCA-3'	1177
cnf1	cnf1 cnf2	5'-AAGATGGAGTTTCCTATGCAGGAG-3' 5'-CATTCAGAGTCCTGCCCTCATTATT-3'	498

cdtB	cdt-a1 cdt-a2 cdt-s1 cdt-s2	5'-AAATCACCAAGAATCATCCAGTTA-3' 5'-AAATCTCCTGCAATCATCCAGTTTA-3' 5'-GAAAGTAAATGGAATATAAATGTCCG-3' 5'-GAAAATAAATGGAACACACATGTCCG-3'	430
Siderophores			
fyuA	FyuA f FyuA r	5'-TGATTAACCCCGCGACGGGAA-3' 5'-CGCAGTAGGCACGATGTTGTA-3'	880
iutA	AerJ f AerJ r	5'-GGCTGGACATCATGGGAACTGG-3' 5'-CGTCGGGAACGGGTAGAATCG-3'	300
Polysaccharide coatings			
***kpsMT* II**	kpsII f kpsII r	5'-GCGCATTTGCTGATACTGTTG-3' 5'-CATCCAGACGATAAGCATGAGCA-3'	272
***kpsMT* III**	KpsIII f KpsIII r	5'-TCCTCTTGCTACTATTCCCCCT-3' 5'-AGGCGTATCCATCCCTCCTAAC-3'	392
***kpsMT* K1**	K1-f	5'-TAGCAAACGTTCTATTGGTGC-3'	153
***kpsMT* K5**	K5-f	5'-CAGTATCAGCAATCGTTCTGTA-3'	159
Miscellaneous			
rfc	rfc-f rfc-r	5'-ATCCATCAGGAGGGGACTGGA-3' 5'-AACCATACCAACCAATGCGAG-3'	788
ibeA	ibe10 f ibe10 r	5'-AGGCAGGTGTGCGCCGCGTAC-3' 5'-TGGTGCTCCGGCAAACCATGC-3'	170
cvaC	ColV-Cf ColV-Cr	5'-CACACACAAACGGGAGCTGTT-3' 5'-CTTCCCGCAGCATAGTTCCAT-3'	680
traT	TraT f TraT r	5'-GGTGTGGTGCGATGAGCACAG-3' 5'-CACGGTTCAGCCATCCCTGAG-3'	290
PAI	RPAi f RPAi r	5'-GGACATCCTGTTACAGCGCGCA-3' 5'-TCGCCACCAATCACAGCCGAAC-3'	930

Table 2. 6: PCR reagents concentrations for each pool of VAGs multiplex PCR.

Multiplex primer Pools	**Reagent**	**Final Concentration**	**Quantity per reaction (µl)**
Pool 1 PAI, *papA, fimH, kps MT* III, *papEF* and *ibeA.*	5X PCR reaction buffer	5X	5.0
	20mM dNTPs	0.8mM	1.0
	25mM $MgCl_2$	4 mM	4.0
	5 U/µM Taq DNA polymerase	2.5 U/µM	0.5
	Sterile dH_2O		6.5
	30 µM Forward Primer	600 pmol	0.5
	30 µM Reverse Primer	600 pmol	0.5
	DNA Template	~3 ng/ µl	2.0
	Total		25
Pool 2 *fyuA, bmaE, sfa/foc, iutA, papG* allele III and K1 (KpsIIr with K1-f)	5X PCR reaction buffer	5X	5.0
	20mM dNTPs	0.8mM	1.0
	25mM $MgCl_2$	4 mM	4.0
	5 U/µM Taq DNA polymerase	2.5 U/µM	0.5
	Sterile dH_2O		6.5
	30 µM each Forward Primer	600 pmol	0.5
	30 µM each Reverse Primer	600 pmol	0.5
	DNA Template	~3 ng/ µl	2.0
	Total		25
Pool 3 *hlyA, rfc, *nfaE,*papG* allele I (internal), **kps MT* II and **papC* * require 10µM concentration	5X PCR reaction buffer	5X	5.0
	20mM dNTPs	0.8mM	1.0
	25mM $MgCl_2$	4 mM	4.0
	5 U/µM Taq DNA polymerase	2.5 U/µM	0.5
	Sterile dH_2O		8.5
	30 µM each Forward Primer	600 pmol	0.5
	30 µM each Reverse Primer	600 pmol	0.5
	*30 µM Forward Primer	300 pmol	0.25
	*30 µM Reverse Primer	300 pmol	0.25
	DNA Template	~3 ng/ µl	2.0
	Total		25

Pool 4 *gafD, cvaC, cdtB, focG, traT* and *papG* allele II	5X PCR reaction buffer	5X	5.0
	20mM dNTPs	0.8mM	1.0
	25mM $MgCl_2$	4 mM	4.0
	5 U/µM Taq DNA polymerase	2.5 U/µM	0.5
	Sterile dH_2O		6.5
	30 µM Forward Primer	600 pmol	0.5
	30 µM Reverse Primer	600 pmol	0.5
	DNA Template	~3 ng/ µl	2.0
	Total		25
Pool 5 *papG* allele I (flanking), *papG* allele II (flanking), *papG* allele III (flanking), **afa/draBC*, **cnf1*,**sfaS* and K5 (KpsIIr with K5-f) * require 10µM concentration	5X PCR reaction buffer	5X	5.0
	20mM dNTPs	0.8mM	1.0
	25mM $MgCl_2$	4 mM	4.0
	5 U/µM Taq DNA polymerase	2.5 U/µM	0.5
	Sterile dH_2O		9
	30 µM each Forward Primer	600 pmol	0.5
	30 µM each Reverse Primer	600 pmol	0.5
	*30 µM Forward Primer	300 pmol	0.25
	*30 µM Reverse Primer	300 pmol	0.25
	DNA Template	~3 ng/ µl	2.0
	Total		25

Table 2. 7: PCR reaction conditions for VAGs multiplex PCR

Step	**Temperature**	**Time**
Preheat lid	95ºC	1 min
Initial denaturation	95ºC	5 min
25 cycles of steps 1,2,3		
1) Denaturation	94ºC	30 sec
2) Annealing	63ºC	30 sec
3) Elongation	68ºC	3 min
Final elongation	72ºC	10 min
Final hold	10ºC	∞

2.6 Multilocus Sequence Typing (MLST)

DNA extractions of overnight LB broth cultures for each of the isolates from clinical samples of bacteraemia and UTI were obtained by using GenElute™ Bacterial Genomic DNA Kits (Sigma-Aldrich) to be used as template. PCR was performed to amplify the target seven housekeeping genes (*adk, icd, fumC, purA, mdh, gyrB and recA*) with reference to the Achtman scheme (mlst.ucc.ie). Primer sequences of each of the target genes, their PCR product size and annealing temperatures are described in Table 2.8. Tables 2.9 and 2.10 describe the PCR reagent concentrations and the PCR conditions respectively. All the PCR products were run on 1% agarose gel for visualization.

Table 2. 8: PCR primer sequences, annealing temperatures (TM) and PCR product size for MLST

Gene	Forward and reverse primer names and sequences	Product size	TM
adk	*adkF* 5'-ATTCTGCTTGGCGCTCCGGG-3' *adkR* 5'-CCGTCAACTTTCGCGTATTT-3'	583 bp	54° C
fumC	*fumCF* 5'-TCACAGGTCGCCAGCGCTTC-3' *fumCR* 5'-GTACGCAGCGAAAAAGATTC-3'	806 bp	54° C
gyrB	*gyrBF* 5'-TCGGCGACACGGATGACGGC-3' *gyrBR* 5'-ATCAGGCCTTCACGCGCATC-3'	911 bp	60° C
icd	*icdF* 5'-ATGGAAAGTAAAGTAGTTGTTCCGGCACA-3' *icdR* 5'-GGACGCAGCAGGATCTGTT-3'	878 bp	54° C
mdh	*mdhF* 5'-ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG-3' *mdhR* 5'- TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT-3'	932 bp	60° C
purA	*purAF* 5'-CGCGCTGATGAAAGAGATGA-3' *purAR* 5'-CATACGGTAAGCCACGCAGA-3'	816 bp	54° C
recA	*recAF* 5'-CGCATTCGCTTTACCCTGACC-3' *recAR* 5'-TCGTCGAAATCTACGGACCGGA-3'	780 bp	58° C

Table 2. 9: PCR reagent concentrations and quantities for MLST

Reagent	Final Concentration	Quantity/reaction (µl)
5X PCR reaction buffer	5X	6.0
10mM dNTPs	0.2mM	0.6
50mM $MgCl_2$	1.67 mM	1.0
5 U/µM Taq DNA polymerase	0.03 U/µM	0.2
Sterile dH_2O		19.2
10 pmol Forward Primer	1.67 pmol	0.5
10 pmol Reverse Primer	1.67 pmol	0.5
DNA	~3 ng/ µl	2.0
Total		30

Table 2. 10: PCR reaction conditions for MLST

Step	**Temperature**	**Time**
Preheat lid	95ºC	1 min
Initial denaturation	95ºC	10 min
30 cycles of steps 1,2,3		
1) Denaturation	95ºC	30 sec
2) Annealing	Primer specific	30 sec
3) Elongation	72ºC	30 sec
Final elongation	72ºC	10 min
Final hold	10ºC	∞

2.6.1 MLST PCR products cleanup

All the MLST PCR products were prepared for sequencing by using Exo-SAP cleanup kit (affymetrix company item: 78201). A volume of 2µl of the Exo-SAP enzyme were added to 5µl of PCR product in 96 wells plate which was sealed and placed in the thermal cycler. The thermal cycler was set to 15 minutes at 37ºC followed by 15 minutes at 80 ºC and 4ºC final hold.

2.6.2 Sequencing of MLST PCR products

All the MLST PCR products were Sanger sequenced by Source Bioscience LifeSciences. The MLST PCR products were sent in 96 wells plates prepared in the required concentrations of 1ng/µl for PCR products and 3.2pmol/µl of primers. The volume required per read was of 15µl of each PCR product.

2.6.3 Analysis of MLST sequence data and determination of the sequence types (ST)

All the raw sequence data received from Source Bioscience LifeSciences as processed using the software CLC Sequence Viewer (version 6.5.3). Forward and reverse sequences were imported for each gene. The reverse sequence was reverse complemented and aligned to its forward sequence. Any nucleotide differences between the two DNA strands were corrected by referring to the original chromatogram file using GENtle software. Bad sequence files were discarded and a new MLST PCR reaction was prepared and sent for sequencing. The consensus sequence resulting from the alignment for each MLST target gene for each isolate was obtained. The consensus sequence for each of the MLST target genes was submitted to the Achtman MLST scheme website (www.mlst.warwick.ac.uk/mlst/mlst/dbs/Ecoli) and their allelic numbers were obtained. The combination of all the allelic numbers of the seven target genes was then submitted to the website for each isolate and their sequence type was determined.

2.6.4 Generation of Minimum Spanning Trees (MSTs)

Phyloviz 1.0 software was used to create minimum spanning trees for the *E.coli* populations from bacteraemia and UTI using non-concatenated sequences of the STs produced by MLST. The software utilizes the goeBURST algorithm to produce a complete minimum spanning tree clustering isolates depending on the level of similarities and differences in their ST allelic profiles (Francisco et al., 2012).

2.7 Serum resistance assay

All twenty-two sequenced ST73 *E.coli* strains were tested for their response to the bactericidal effect of human serum. The procedure was adapted from the published protocol (Podchun et al., 2001). Initial bacterial suspensions were diluted to 2 X 10^6 in saline, then 250µl of from each suspension was added to 750µl of human serum in 24 well plates, mixed well and incubated at 37ºC for 3 hours. Viable counts were determined for each strain on LB agar by Miles and Misra method for both the initial inoculum (T_0) and after three hours incubation in normal human serum (T_3) and compared. Strains were recorded as resistant when they maintained or increased their viable counts compared to their initial counts (T_0). Strains with decreased viable counts at (T_3) to less than 50% of their initial count (T_0) were recorded sensitive. *E. coli* MG1655 was used as serum sensitive control strain as they are cleared completely after incubation in normal human serum. *E. coli* CFT073 was used as a serum resistant control strain (Miajlovic et al., 2014).

2.8 Genome sequencing of ST73 *E. coli* isolates

A selection of ST73 *E.coli* isolates (n= 22) were cultured overnight in LB broth and DNA extraction was performed using GenElute™ Bacteria Genomic DNA Kits (Sigma-Aldrich). The DNA samples were sent for sequencing at Exeter Sequencing Service at the University of Exeter (Ess-wiki.exeter.ac.uk, 2014). The sequencing system Illumina HiSeq 2500 was used and 100bp paired end sequencing was performed (Systems.illumina.com, 2014).

2.9 Sequence assembly using Velvet and PAGIT

All the raw sequence data were received from Exeter Sequencing Service in FASTQ file format. FASTQ files are text files that include the sequencing reads of the nucleotides and their associated quality scores (Cock, et al., 2010). Velvet software was used to assemble the sequences (Zerbino and Birney, 2008). Velvet is a de Bruijn graph assembler used for the production of *de novo* assembled genomes, which means it assembles the genomes without the use of a reference genome of a closely related strain (Swain, et al., 2012). Velvet assembly parameters were k-mer size of 31, expected coverage of 75-fold and coverage cutoff of 8 fold. Assembled genome sequences were used in the post assembly genome improvement toolkit (PAGIT). PAGIT software produces improved quality sequence from *de novo* assembled genomes by utilizing ABACAS, IMAGE, iCORN and RATT and a closely related reference strain (Swain, et al., 2012). In the PAGIT assembly CFT073 *E. coli* genome sequence (NCBI Reference Sequence: NC_004431.1, genome size is 5231428 bps) was used as a reference genome because it belonged to sequence type ST73. Full details of command lines and parameters are provided in the appendix (chapter six).

2.10 Genome annotation

Prokka is a rapid accurate unguided annotation tool compared to other transfer annotation softwares that produce errors as they force the transfer of annotations (Seemann, 2014). Only FASTA files of the assembled genomes are required to start the Prokka script where it utilizes a built in database to annotate the genome and produces output files in GFF and GENBANK (GBK) formats which are compatible annotation formats ready for further analysis (Seemann, 2014). All the selected twenty two ST73 *E. coli* isolate genomes were annotated by using Prokka software. Command lines are provided in the appendix (chapter six).

2.11 High resolution SNP typing and phylogeny

2.11.1 Alignment of sequenced genomes

The raw FASTQ sequence files of all sequenced ST73 *E. coli* isolates were aligned against *E. coli* CFT073 as a reference genome using SMALT software, which is developed and optimized by the Wellcome Trust Sanger Institute (Sanger.ac.uk, 2014). SMALT provides a more accurate alignment of paired end sequencing reads produced by Illumina, Roche 454 or ABI Sanger compared to other alignment programmes such as BWA and BOWTIE (Wellcome Trust Sanger Institute, 2014a). The SMALT alignment process involves first forming an index table of the reference genome consisting of specific length words or hashing with equidistant spaces, then sequence reads are mapped against the hash index (Wellcome Trust Sanger Institute, 2014b). SAMtools was used to extract the required information from SAM files generated by SMALT in the previous step. SAMtools software used to manipulate the data in SAM format files such as indexing, merging, sorting and other post processing requirements (Li et al., 2009; Evolution and Genomics, 2011). This results in VCF files containing all the SNP data for each strain against the reference genome *E. coli* CFT073. In house scripts were used by Alan McNally at NTU to filter and produce high fidelity SNPs. The filtering criteria for obtaining high fidelity SNPs for a phylogenetic tree was removing multiple alleles, removing reads with quality scores of less than thirty, removing SNPs reads with depth of coverage less than eight and with allele frequency of less than 75%. Full parameters and scripts are provided in the appendix (chapter six).

2.11.2 Construction of a phylogenetic tree

The VCF files were used to create a consensus sequence for each strain based on the CFT073 reference. RAxML-HPC was used to produce a maximum-likelihood phylogenetic tree with rapid bootstrapping and the (GTR) or general time reversible model with gamma distribution (Felsenstein, 1981; Rodriguez et al., 1990). The resulting phylogenetic tree was viewed using Figtree software (tree.bio.ed.ac.uk/software, 2014). More details and scripts are provided in the appendix (chapter six).

2.11.3 Measurement of pairwise distance

The FASTA files generated in section 4.2.7.3 were used to compute the number of base differences between the ST73 *E coli* strains using MEGA software (Tamura et al., 2013). A table was generated and exported as spreadsheet file for further analysis.

2.12 Comparative Genome analysis

Whole genome comparisons were performed using the software Gegenees (Gegenees.org, 2014). Gegenees software allows the comparative analysis of hundreds of complete or draft genomes by aligning fragments of all the genomes utilizing a multithreaded BLAST approach. The alignment data can provide a phylogenetic view of the genomes, but more importantly for our study, it can be used to investigate the presence of genomic regions unique to a target group when compared to a background group (Agren et al., 2012). Gegenees software was also used to determine core and pangenome statistics.

2.12.1 Investigation of bacteraemia and ESBL carrier specific loci

Gegenees software was used to investigate the presence or absence of genomic regions or loci specific only to bacteraemic or to ESBL carrier strains of the sequenced ST73 *E. coli*. This was performed by utilizing the Fragmented all-all comparison option (Gegenees.org, 2014). First, all twenty two GENBANK (GBK) files were imported into a new database in Gegenees. Then, in the Fragmented all-all comparison window, nucleotide comparison (BLASTN) was chosen with the highest resolution parameters for the alignment (200/100), which were 200 bps of sliding fragment window size and 100bps of progressive step size respectively. When the alignments and fragments comparison were performed and displayed, the bacteraemia strains were selected as target group against a background group of all the UTI strains in order to investigate for the presence of bacteraemia specific genomic regions or loci. The strain B134 was chosen as reference genome from the target group to refer the unique loci to with annotations. In the other test, all the ESBL carrier strains were selected as a target group from bacteraemia and UTI clinical samples against a background group of ESBL absent strains in order to investigate ESBL carrier specific genomic regions, or loci with B134 strain as a reference from the target group.

2.12.2 Determination of core and pangenome

An overview of the core and pangenome of all the sequenced ST73 *E. coli* strains was created by Gegenees software. All the genomes were fragmented then compared to each other and their BLASTN scores were determined. The pangenome was constructed progressively by analysing the fragments BLASTN scores and then each new unique fragment was deposited in the pangenome (Gegenees.org, 2014). In the New pangenome option, 1% threshold was selected for maximum distinctive selection of unique fragments.

2.13 Plasmid profiling by S1 nuclease Pulsed Field Gel Electrophoresis (S1 PFGE)

The protocol of S1 nuclease treatment of genomic DNA for profiling plasmids was adapted from the published protocol of Barton et al., (1995). S1 PFGE is an established method used to detect, size and isolate plasmid DNA from bacteria (Keyes et al., 2000; Basta et al., 2004; Zong et al., 2010; Brolund et al., 2013). S1 nuclease is an endonuclease enzyme isolated from *Aspergillus oryzae* that cleaves single stranded nucleic acids. It can also degrade double stranded DNA in specific regions where single stranded DNA is exposed by a nick, a mismatch, a gap or a loop (Lehman, 1981). In case of supercoiled plasmids torsional stress occurs in the molecule that becomes sensitive to S1 nuclease activity. Once one strand is cleaved, S1 nuclease cuts the intact strand at the nick opposite to the initial break (Barton et al., 1995). The S1 nuclease is known to cut double stranded genomic DNA, but much less frequently. Subsequently, linearized plasmids are isolated and separated by PFGE as bands in a faint smear of genomic DNA background. The size of the plasmid is accurately determined with reference to DNA markers (Barton et al., 1995). All thirteen ESBL carrier strains (B10, B18, B29, B36, B40, B72, B73, B84, B91, B134, U42, U50 and U76) were selected and three CTX-M negative strains (B14, B102 and U24) were selected (Table 4.1A and Table 4.1B). NCTC13353 *E. coli* was used as a positive control for a plasmid carrier.

2.13.1 Reagents and Buffers

1 M Tris-HCL (pH 8.0)

A final concentration of 1 M Tris-HCL was obtained by dissolving 121.1 g of Tris base (Fisher Scientific) in 700 ml of dH_2O at 56°C with magnetic steering. Concentrated HCL and ultra-pure water was added gradually maintaining pH of 8.0 to complete 1 L. The reagent was autoclaved and stored at room temperature.

0.5 M EDTA (pH 8.0)

A final concentration of 0.5 M ethylenediamine tetra-acetic acid, sodium hydroxide (EDTA) buffer was prepared by dissolving 186.1 g of EDTA (Sigma Aldrich) in 700 ml of dH_2O with the addition of Sodium hydroxide (NaOH) to adjust the pH to 8.0, the solution was then made up to 1 L by adding de-ionized water. The reagent was autoclaved and stored at room temperature.

10 X TBE Buffer

A final concentration of 10 X Tris base, boric acid and EDTA (TBE) buffer was obtained by dissolving 108 g Tris base and 55 g boric acid (Fisher Scientific) in 700 ml of dH_2O. Then, 80 ml of 0.5 M EDTA (pH 8.0) was added and the mixture was completed to 1 L by adding dH_2O.

Tris:EDTA (TE) buffer

The final concentration of TE buffer was 10 mM Tris-HCL:1 mM EDTA, pH 8.0. It was prepared by adding 10 ml of 1 M Tris-HCL (pH 8.0) to 2 ml of 0.5 M EDTA (pH 8.0).

Cell Suspension Buffer (CSB)

The final concentration of CSB was 100 mM Tris-HCl:100 mM EDTA, pH 8.0. It was prepared by adding 10 ml of 1 M Tris-HCl (pH 8.0) to 20 ml of 0.5 M EDTA (pH 8.0). Sterile pure water was used to complete the buffer to 1 L.

Cell Lysis Buffer (CLB)

CLB was prepared by mixing 25 ml of 1 M Tris-HCl (pH 8.0) with 50 ml of 0.5 M EDTA (pH 8.0). Then 50 ml of 10% Sarcosyl (N-Lauroyl-Sarcosine Sodium salt, Sigma Aldrich) was added and the mixture was diluted by sterile pure water to 500 ml. Just before use, 25µl Proteinase K (20 mg/ml, Sigma Aldrich) was added to every 5 ml of CLB. The final concentration of CLB was 50 mM Tris-HCl:50 mM EDTA, pH 8.0, 1% Sarcosyl and 0.1 mg/ml Proteinase K.

DNA Markers and other products

Other reagents and DNA markers were used in the S1 PFGE as follows: MidRange I PFGE Marker (New England Biolabs), Low Range PFGE Marker (New England Biolabs), S1 Nuclease (Promega), SeaKem® Gold Agarose (Biozym Diagnostics), N-Lauroylsarcosine sodium salt (Sigma Aldrich), Lysozyme from chicken egg white (Sigma Aldrich), Thiourea (Sigma Aldrich) and Proteinase K (Sigma Aldrich).

2.13.2 S1 PFGE protocol

Bacterial cultures

E. coli cultures were incubated for overnight at 37°C on LB agar.

Preparation of the agar plugs with the bacteria

0.1g of SeaKem® Gold Agarose (Biozym Diagnostics) was dissolved in 9.5 ml TE and 500 µl of 20% sodium dodecyl sulphate (SDS) and kept in a water bath at 55℃. Working on ice, two or three loops of bacterial colonies were suspended in 1.6ml of Cell Suspension Buffer (CSB) and Centrifuged at 7000rpm for 4 min. The supernatant was discarded and the pellet was washed again with another 1.6ml CSB. Centrifugation was repeated and the cells were resuspended in 1ml of CSB. The optical density (OD) of the bacteria suspension was adjusted to 1.4 at wave length of 600 nm. In a 1.5 ml Eppendorf, 400 µl of the bacterial suspension was transferred and incubated at 37℃ for 10 min in water bath. After the incubation, 25µl of proteinase K (20mg/ml) was added and mixed gently by tapping them 3 to 4 times. Then, 400µl of agarose was added and gently mixed using non sharp tips. Finally, the mixture was added into the plug moulds carefully to avoid the formation of bubbles and the moulds were incubated at 4℃ for 10 min.

Bacterial Cell Lysis

Cell lysis buffer (CLB) was prepared for bacterial cell lysis. For every 5ml of CLB, 25µl of proteinase K (20mg/ml) and 0.5mg lysozyme (Sigma Aldrich) were added and mixed thoroughly. Then, each of the prepared plugs earlier were transferred into 2 ml of the prepared CLB ensuring that they were completely immersed in the buffer and incubated in a water bath at 55℃ for 2 hours.

Washing the plugs after the bacterial cell lysis

Ultra-pure water and TE buffer were equilibrated in a water bath at 55℃ before use. The plugs were transferred to 2 ml Eppendorf tubes filled with ultra-pure water and washed by pipetting. Then the plugs were transferred to a 15 ml Falcon tube filled with ultra-pure water and incubated at 50℃ in a water bath with shaking for 15 min. The washing was repeated once with ultra-pure water and repeated all again three times with TE buffer. After discarding the liquid of the final wash, 10 ml of TE buffer (stored at room temperature) was added to the plugs which were then stored at 4℃.

Digestion with S1 nuclease

The plugs were transferred to a new Eppendorf tube containing 200 µl of 1X S1 nuclease buffer (10 X, Promega) and incubated for 20 min at room temperature. Then, the buffer was discarded and 100 µl of fresh buffer with 0.2µl of nuclease S1 (89 U/µl, Promega) was added, and the tube was incubated for 45 min at 37°C. The reaction was stopped by adding 10 µl of 0.5 M EDTA (pH 8) and left for 30min at room temperature. Finally, 1ml of 1 X TE buffer was added to the Eppendorf with the plugs and left at 4°C for 30 min.

Electrophoresis

A concentration of 200 uM Thiourea (Sigma Aldrich) in 0.5 X TBE buffer was prepared by dissolving 15.22mg for 1L of 0.5 X TBE, a total of 2500 ml of 0.5 X TBE was used for a single electrophoresis run. The agar was prepared by dissolving 1g of SeaKem® Gold Agarose (Biozym Diagnostics) in 100ml mixture of TBE and Thiourea by incubation in a water bath at 55℃. The plugs and DNA Markers were placed on the comb and dried with a filter paper then the comb was fitted in the electrophoresis tray. The gel was then poured and kept for 30 min at

room temperature to solidify. The comb was removed and the wells were filled with liquid gel. Finally, the gel was placed in the tank and electrophoresis was run on the conditions: Block 1, 6V/cm, initial switch 5s, final switch 40s, run time 12h; Block 2, 6V/cm, initial switch 3s, final switch 8s, run time 8h. The gel was stained with Ethidium Bromide (1µg/ml) for 30 min and photographed.

2.14 Plasmid typing by Plasmid Finder database

Plasmid finder is a web based tool developed to identify the presence of plasmids and determine their replicon types from whole or partial, raw or sequenced genome and plasmid sequences of bacteria belonging to the *Enterobacteriaceae* family (Carattoli et al., 2014). The database was constructed from 559 sequenced plasmids collected from the database of the National Centre for Biotechnology (NCBI) to produce a plasmid replicon sequence identification tool through the openly accessed website (www.cge.cbs.dtu.dk). FASTA files of all the thirteen ESBL positive ST73 *E. coli* strains from the clinical samples of bacteraemia and UTI were used with the assembled input files option on the website. Gram negative *Enterobacteriaceae* with 80% threshold option were selected and the output was obtained.

2.15 Statistical analysis

SPSS software (version 19.0) was used to perform the statistical analysis of the data. Chi square test was used to determine statistical differences of the antibiotic resistance profiles, ESBL carriage, VAGs presence and ST prevalence between bacteraemia and UTI *E.coli* isolates. The normal distribution was confirmed and an independent t test was performed to test statistical differences in VAGs carriage of *E.coli* isolates from bacteraemia and UTI.

Chapter three

Comparison of *E. coli* populations from clinical samples of UTI and bacteraemia

3.1 Introduction

3.1.1 Population studies of *E. coli*

Numerous population studies of *E. coli* have been performed due its importance as a human pathogen (Hartl and Dykhuizen, 1984). Population studies have been performed to investigate the epidemiology of the emerging ESBL producing *E. coli* phenotypes to define their incidence and risk factors (Pitout, et al., 2004). The development of molecular techniques such as multi locus enzyme electrophoresis (MLEE) was the basis of many population genetic studies and established the link between virulence and phylogenetic groups (Clermont, et al., 2000; Johnson and Russo, 2002). Based on MLEE all *E. coli* strains can be divided into 4 main phylogenetic groups namely A, B1, B2 and D. Commensal *E. coli* strains are found in phylogenetic groups A or B1. *E. coli* strains which cause intestinal disease are found in A, B1 or D groups, while ExPEC strains in general are found in group B2, or to a lesser extent in group D (Jaureguy, et al., 2008).

Genotyping methods are being increasingly used to characterize bacterial pathogens in population genetics studies. Multi-locus sequence typing (MLST) is considered the main genotyping tool for identifying different strains of ExPEC (Tartof, et al., 2005). MLST identifies specific sequences within seven housekeeping genes that are conserved throughout the species. Each gene is given an allelic number and the combination of the seven allelic numbers produces a unique sequence type (ST) associated with the strain. The Achtman scheme is the main established MLST typing scheme for *E. coli* (Wirth et al., 2006). MLST played a major role in identifying the global emergence of the pandemic ST131 *E. coli* clone (Nicolas-Chanoine, et al., 2008). Many MLST based population studies have since been

performed and confirmed the clonality of ST131 *E. coli* which is highly associated with CTX-M-15 ESBL mediated antimicrobial resistance (Nicolas-Chanoine, et al., 2008).

The structure of the *E. coli* UTI population in the East Midlands area was previously investigated by our group (Croxall et al., 2011). One hundred and fifty random UTI *E. coli* isolates were collected and characterized in terms of MLST, antibiotic sensitivity, VAGs acquisition and in vitro invasion of epithelial cells. That study reported that *E.coli* isolated from UTI infections were resistant to front line antibiotics with increased resistance to ciprofloxacin and trimethoprim. The genotypic structure of the UTI population revealed a diverse population consisting of a total of 52 STs. ST131 complex represented the most dominant ST accounting for 22% of the population, followed by ST73 (11%) and ST69 (9%). Whole genome comparisons revealed that the dominant ST131 isolates were genetically homogeneous with between 10 and 60 SNPs difference between circulating UTI isolates in the region (Clark et al., 2012).

3.1.2 Aim

The population structure of UTI *E. coli* isolates is well characterized in the literature. Our group have contributed to defining the UTI *E. coli* population in the East Midlands area where a diverse population was observed consisting of 52 STs, only 26% of which were multi drug resistant ESBL carriers primarily represented by ST131 isolates while the remaining majority of STs were drug susceptible (Croxall, et al., 2011). Incidence of bacteraemia due to *E. coli* is increasingly reported worldwide. More importantly, bacteraemia due to multidrug resistant ESBL positive *E. coli* is also on the increase. Yet the population structure of *E. coli* causing bacteraemia is poorly defined. It would be beneficial to perform a thorough investigation of the molecular epidemiology of *E. coli* causing bacteraemia to determine if the UTI population

structure is mirrored in bloodstream infections. Clinical *E. coli* isolates from bacteraemia were collected from Nottingham University Hospital (NUH) with a contemporaneous collection of urine *E. coli* isolates from the period of March 2011 to July 2011, to contextualize our findings and provide a comprehensive snapshot of the epidemiology of the circulating ExPEC in the region. Specific aims of this chapter are:

A) Phenotypic characterization of *E. coli* isolates from clinical samples of bacteraemia and UTI in terms of antibiotic susceptibility profiles, ESBL carriage and VAGs carriage.
B) Genotypic characterization of the population structure of *E. coli* populations from bacteraemia and UTI by MLST.
C) Comparison between the *E. coli* isolates from bacteraemia and UTI to provide a comprehensive view of the existing ExPEC genetic epidemiology in bacteraemia and UTI.

3.2 Results

3.2.1 Increased levels of antibiotic resistance in clinical *E. coli* bacteraemia isolates compared to UTI isolates.

140 *E. coli* isolates from clinical cases of bacteraemia and 125 *E. coli* isolates from clinical cases of UTI collected from Nottingham University Hospital were tested for their antibiotic susceptibility according to the BSAC protocol as displayed in Figure 3.1.

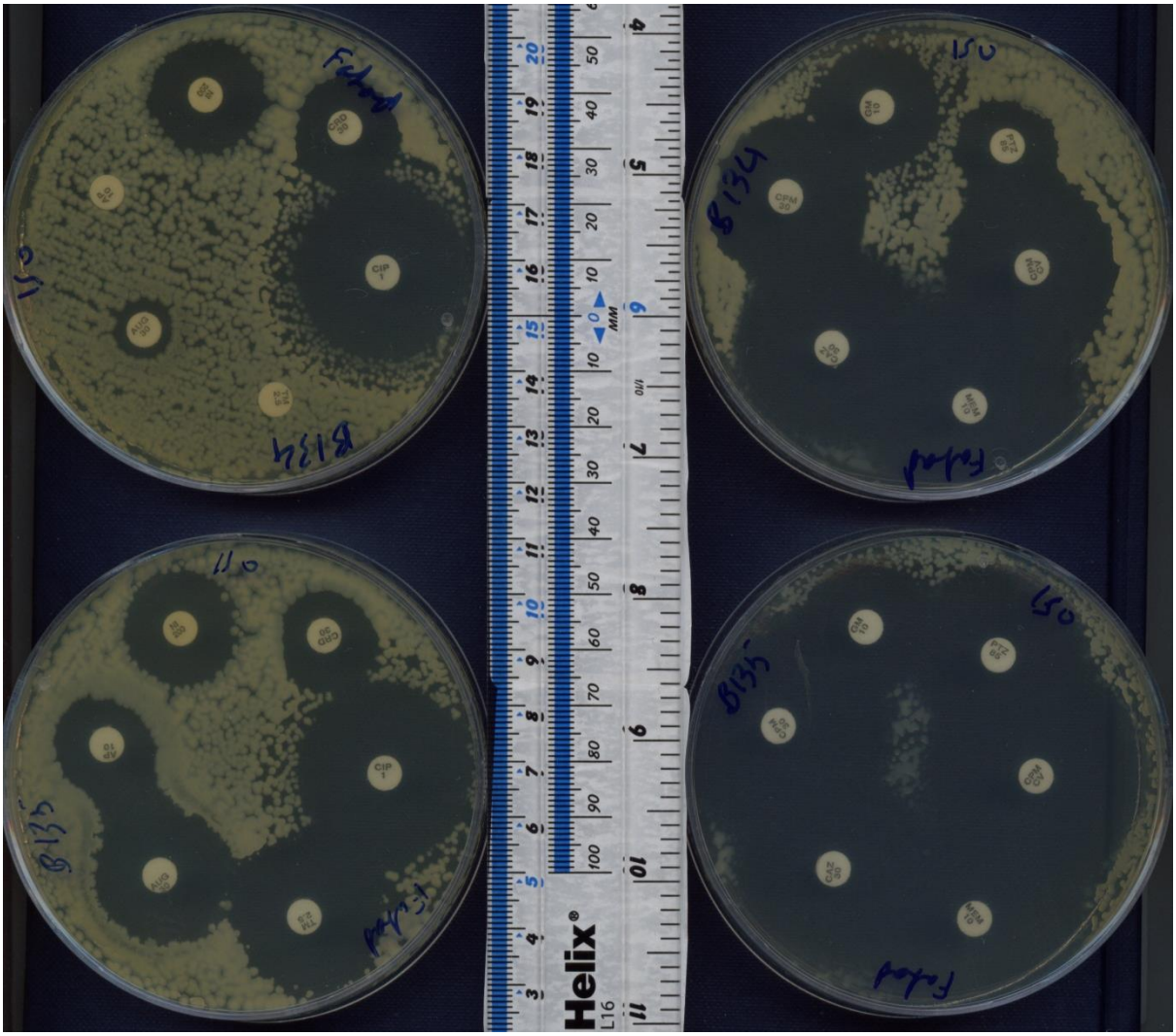


Figure 3. 1: Disc diffusion antibiotic susceptibility test. The figure displays an image of the disc diffusion susceptibility test of two bacteraemia *E. coli* isolates B134 (top two iso-sensitest

agar plates) and B135 (bottom two iso-sensitest agar plates). We notice the difference between the resistance profile of the B134 strain to ampicillin and trimethoprim displayed by the absence of clear zones around the discs (top left) compared to B135 (bottom left).

For all the antibiotics that were tested, the proportions of resistant bacteraemia *E.coli* isolates were higher than the UTI *E. coli* isolates (Figure 3.2). Differences in the percentages of resistance between the bacteraemia and UTI isolates were noticeable against ampicillin (63.6%/54.4%), ciprofloxacin (25.7%/8.8%) and cefradine (20%/11.2%), with statistical significant difference in resistance to ciprofloxacin ($P \leq 0.001$, 95% confidence interval is ±4.6) and cefradine ($P \leq 0.05$, 95% confidence interval is ±4.4).

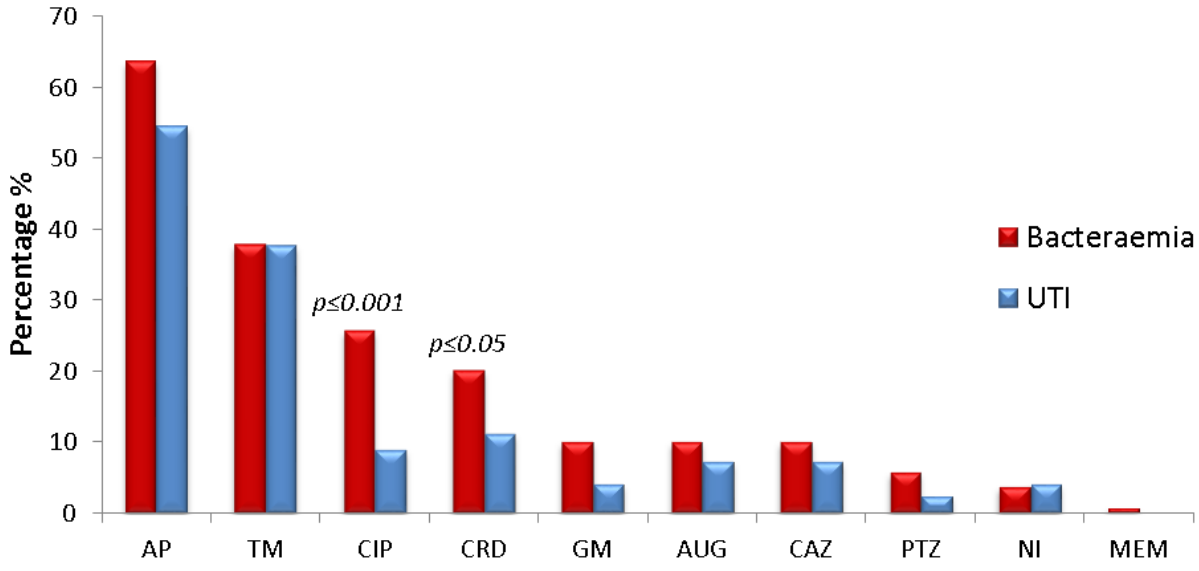


Figure 3. 2: Proportions of antibiotic susceptibility profiles of *E. coli* bacteraemia and UTI. The figure displays the percentages of bacteraemia and UTI *E. coli* isolates resistance to front line antibiotics gentamicin (GM), ceftazidime (CAZ), meropenem (MEM), piperacillin-tazobactam (PTZ), co-amoxiclav (AUG), trimethoprim (TM), ciprofloxacin (CIP), cefradine (CRD), nitrofurantoin (NI), and ampicillin (AP). Statistically significant differences in resistance to ciprofloxacin ($P \leq 0.001$) and cefradine ($P \leq 0.05$) are indicated.

The level of high antibiotic resistance of the bacteraemia *E.coli* isolates is further emphasized in Figure 3.3 which describes the number of isolates in the two populations displaying a multiple drug resistance (MDR) phenotype, exhibiting resistance to antibiotics belonging to two or more different groups of antibiotics. Our data shows a significantly higher number of MDR bacteraemia *E. coli* isolates than UTI *E. coli* isolates (*P*=0.01, 95% confidence interval is ±5.9).

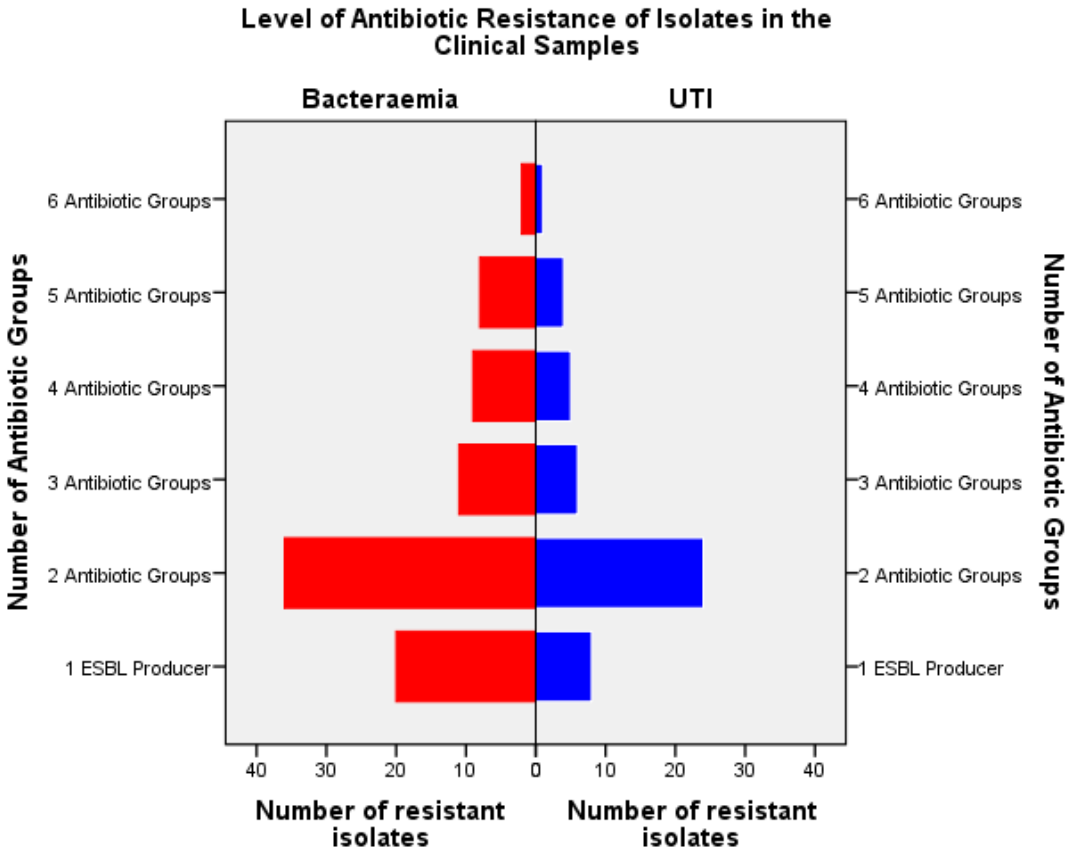


Figure 3. 3: Levels of multiple antibiotic drug resistance (MDR) in the clinical isolates. The figure shows the number of (MDR) phenotypes displayed by bacteraemia and UTI *E. coli* isolates. MDR phenotypes defined as exhibiting resistance to antibiotics belonging to two or more different classes of antibiotics. Data shows a significantly higher number of MDR bacteraemia *E. coli* isolates than UTI *E. coli* isolates (*P*=0.01). The graph also shows that the

phenotypic expression of ESBLs by the double disc sensitivity test was found to be significantly higher in bacteraemia *E. coli* isolates compared to UTI *E. coli* isolates (n=21/n=8, *P*=0.025, 95% confidence interval is ±3.9).

3.2.2 Significantly higher ESBL carriage in bacteraemia *E. coli* isolates compared to UTI *E. coli* isolates

PCR screening for ESBL carriage was performed on all *E. coli* isolates as displayed in Figure 3.4. Compared to the UTI *E. coli* population, bacteraemia *E. coli* isolates presented significantly higher percentages of ESBL carriage for bla_{SHV} (15.7%/5.6%, *P*=0.008, 95% confidence interval is ±3.7), $bla_{CTX\text{-}M}$ (29.3%/17.6%, *P*=0.025, 95% confidence interval is ±5.1) and bla_{OXA} (14.3%/6.4%, *P*=0.037, 95% confidence interval is ±3.7) but not bla_{TEM} (88.6%/84%, *P*=0.28, 95% confidence interval is ±4.1). In turn, the total ESBL carriage, excluding bla_{TEM}, in the bacteraemia population was significantly higher than the UTI *E. coli* population (59.3%/29.6%, $P \leq 0.001$, 95% confidence interval is ±6). Production of ESBLs was tested against all the *E. coli* isolates from the two clinical groups of bacteraemia and UTI with the use of ESBL combination discs tests (Figure 3.3). Consistent with the results of ESBL PCR screening, phenotypic production of ESBLs was significantly higher among bacteraemia *E. coli* isolates than UTI *E. coli* isolates (n=21/n=8, p=0.025, 95% confidence interval is ±3.9).

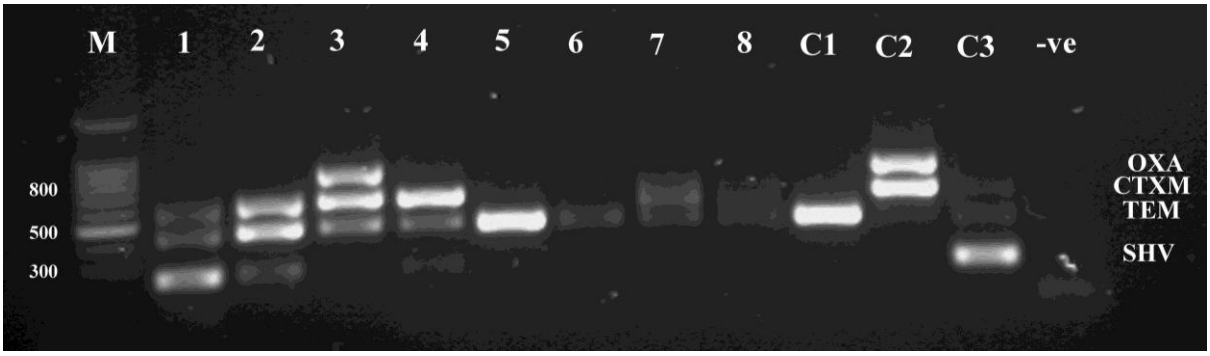


Figure 3. 4: ESBL screening multiplex PCR gel image. The figure displays the gel image of ESBL multiplex PCR for SHV (237bp), TEM (445bp), CTXM (593bp) and OXA (813bp). Lanes 1-8 represent the samples B36, B40, B134, B72, B14, B47, B51, and B62 respectively. Positive controls for each of the ESBL primers were used which are NCTC 13351 *E. coli* (C1), NCTC 13353 *E. coli* (C2) and NCTC 13368 *Klebsiella pneumoniae*. Negative control contained no DNA template.

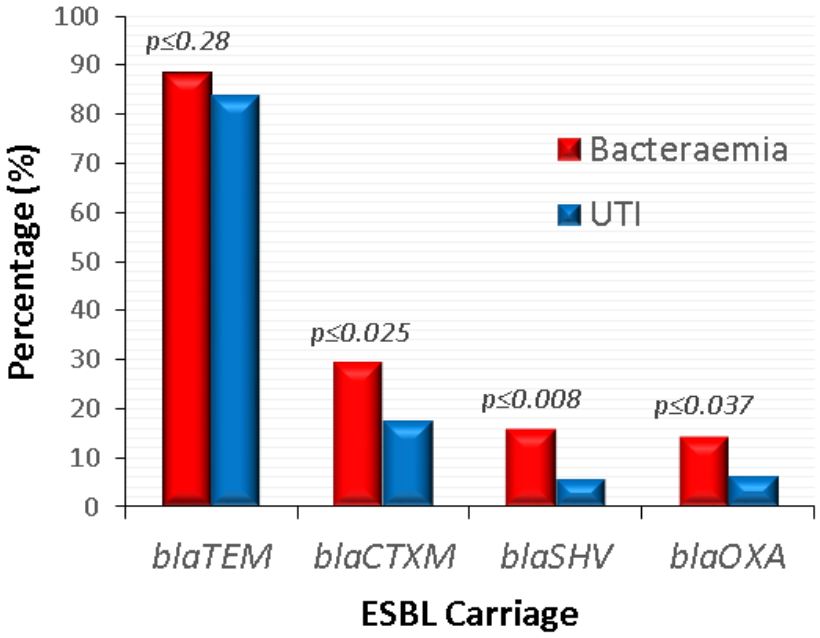


Figure 3. 5: Percentages of ESBL carriage of the bacteraemia and UTI clinical isolates. The figure displays the percentages of ESBL carriage of bacteraemia and UTI isolates. Significant difference in carriage of ESBL determinants bla_{SHV} (P=0.008), $bla_{CTX\text{-}M}$ (P=0.025) and bla_{OXA} (P=0.037) but not bla_{TEM} (P=0.28).

3.2.3 MLST shows a reduction in diversity in the bacteraemia population compared to the UTI population

MLST sequence types were determined for all *E. coli* isolates using the Achtman scheme (Wirth et al., 2006). The UTI population was represented by 63 different STs with diverse prevalence. The highest ST prevalence in the UTI population was ST73 (n=16, 12.8%), followed by a gradual decrease in the prevalence of the other STs such as ST131 (n=9, 7.2%), ST69 (n=9, 7.2%), ST95 (n=6, 4.8%), ST404 (n=6, 4.8%), ST127 (n=4, 3.2%), ST141 (n=4, 3.2%) and ST10 (n=3, 2.4%). Prevalence patterns of the STs of the bacteraemia *E. coli* population were noticeably different. Three main STs were obtained with ST131 being the most dominant (n=30, 21.4%) and significantly higher in prevalence than the UTI population ($P \leq 0.001$, 95% confidence interval is ±4.3), followed by ST73 (n=24, 17.1%) and ST95 (n=13, 9.2%). There was a marked decrease in the prevalence of the other STs in the bacteraemia population such as ST69 (n=5, 3.6%), ST12 (n=4, 2.9%), ST2278 (n=3. 2.1%), ST10 (n=2, 1.4%) and ST14 (n=2, 1.4%). Loss of diversity in the bacteraemia *E. coli* population was illustrated in Figure 3.6 as only eight of the most prevalent STs represent 59.2% of the total population, while the eight most prevalent STs in the UTI population represented only 45.6% of the total population.

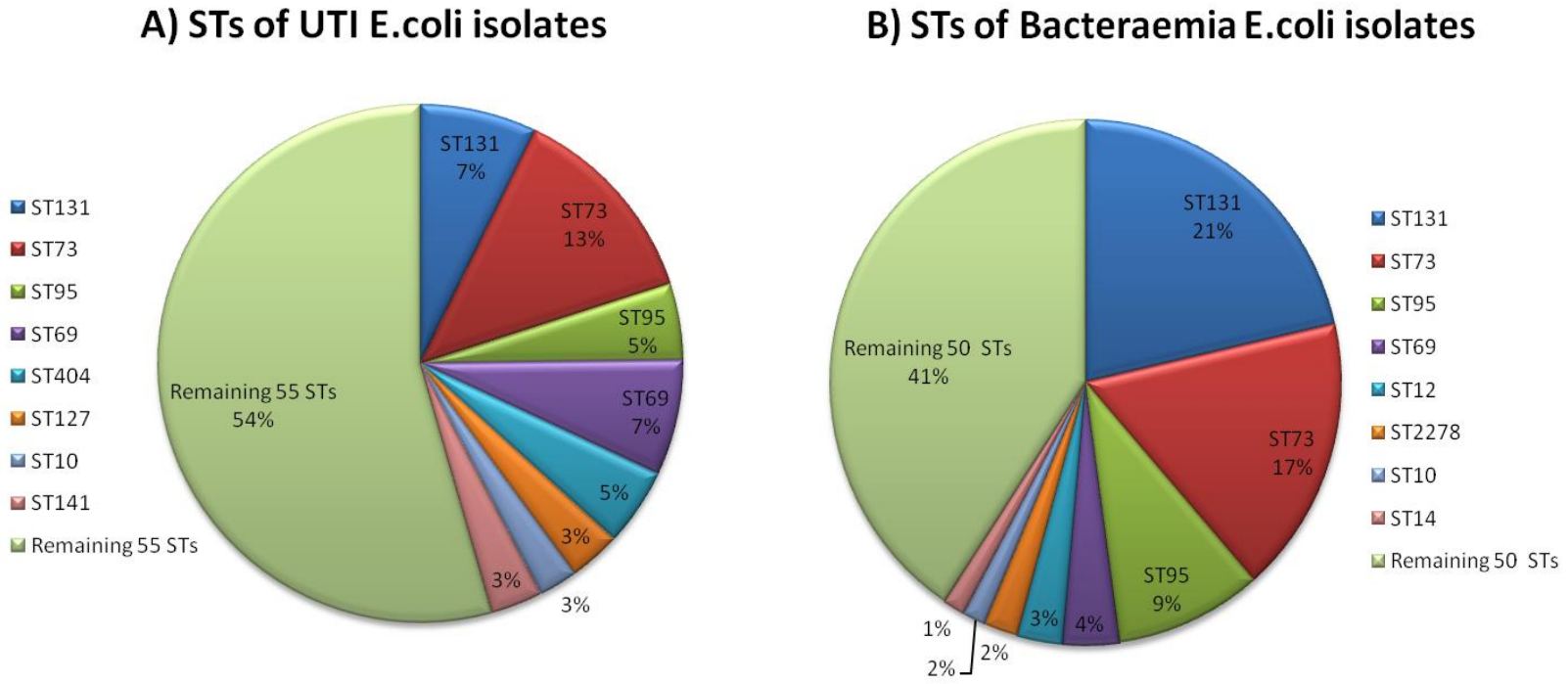


Figure 3. 6 Prevalence Patterns of the major STs in UTI and bacteraemia populations. (**A)** The UTI population was represented by 63 different STs with diverse prevalence, ST73 (n=16), ST131 (n=9), ST69 (n=9), ST95 (n=6), ST404 (n=6), ST127 (n=4), ST141 (n=4) and ST10 (n=3). **(B)** Prevalence patterns of the STs of the bacteraemia *E. coli* population were noticeably different. ST131 (n=30) is significantly higher in prevalence than the UTI population ($P \leq 0.001$), followed by ST73 (n=24), ST95 (n=13), ST69 (n=5), ST12 (n=4), ST2278 (n=3), ST10 (n=2) and ST14 (n=2).

3.2.4 Reduction in diversity of the bacteraemia population is associated with ESBL carriage

Minimum spanning trees (MST) were created using the Phyloviz 1.0 software (Francisco, et al., 2012) to visualize the different STs of the two *E. coli* populations from bacteraemia and UTI in relation to each other according to similarities and differences in their MLST allelic profiles. In the MST (Figures 3.7 and 3.8), each circle represents one ST, and the size of the circle reflects the number of isolates belonging to this particular ST within the bacteria population. The lines between the circles represent how different their allelic profiles are, where a line labelled by number one means the two linked STs differ in at least one of the seven alleles which is named a single locus variant (SLV), and a cluster of STs linked by SLVs are referred to as a clonal complex (Francisco, et al., 2012). The UTI *E. coli* population MST (Figure 3.7) was found to be diverse with evenly distributed high prevalent STs, ST73 (n=16), ST131 (n=9), ST69 (n=9), ST95 (n=6) and ST404 (n=6). They are displayed in almost equal sizes and slightly larger than the remaining STs of the UTI population. In contrast, the loss of diversity of the bacteraemia *E. coli* population was more apparent in their MST (Figure 3.8) with the three dominant STs ST131 (n=30), ST73 (n=24) and ST95 (n=13) clearly larger than the other STs.

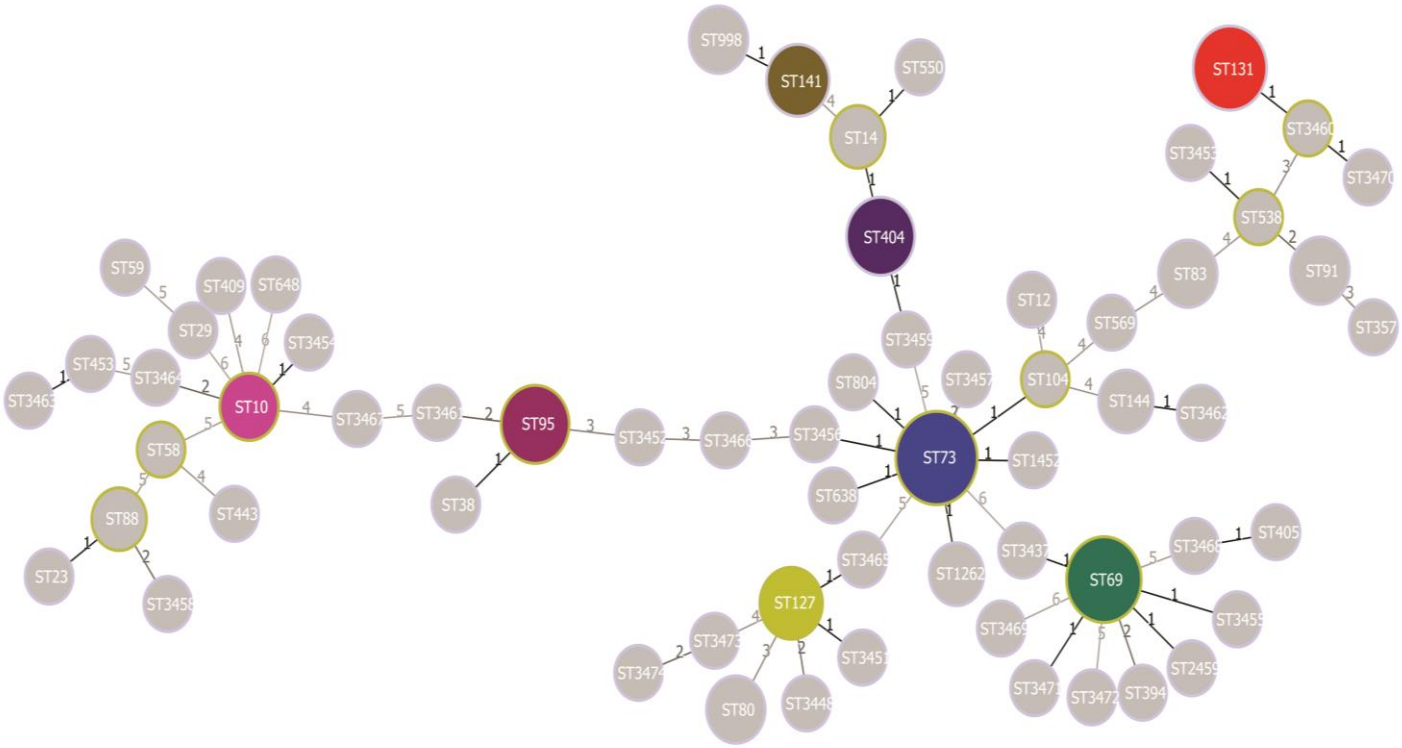


Figure 3. 7: Minimum Spanning Tree of UTI *E. coli* isolates. Each circle represents one ST, and the size of the circle reflects the number of isolates belonging to this particular ST within the bacteria population. The lines between the circles represent how different their allelic profiles are, where a line labelled by number one means the two linked STs differ in at least one of the seven alleles which is named a single locus variant (SLV), and a cluster of STs linked by SLVs are referred to as a clonal complex. Diversity of the UTI population is displayed with the main STs (ST73 n=16, ST131 n=9, ST69 n=9, ST95 n=6 and ST404 n=6) highlighted in different colours.

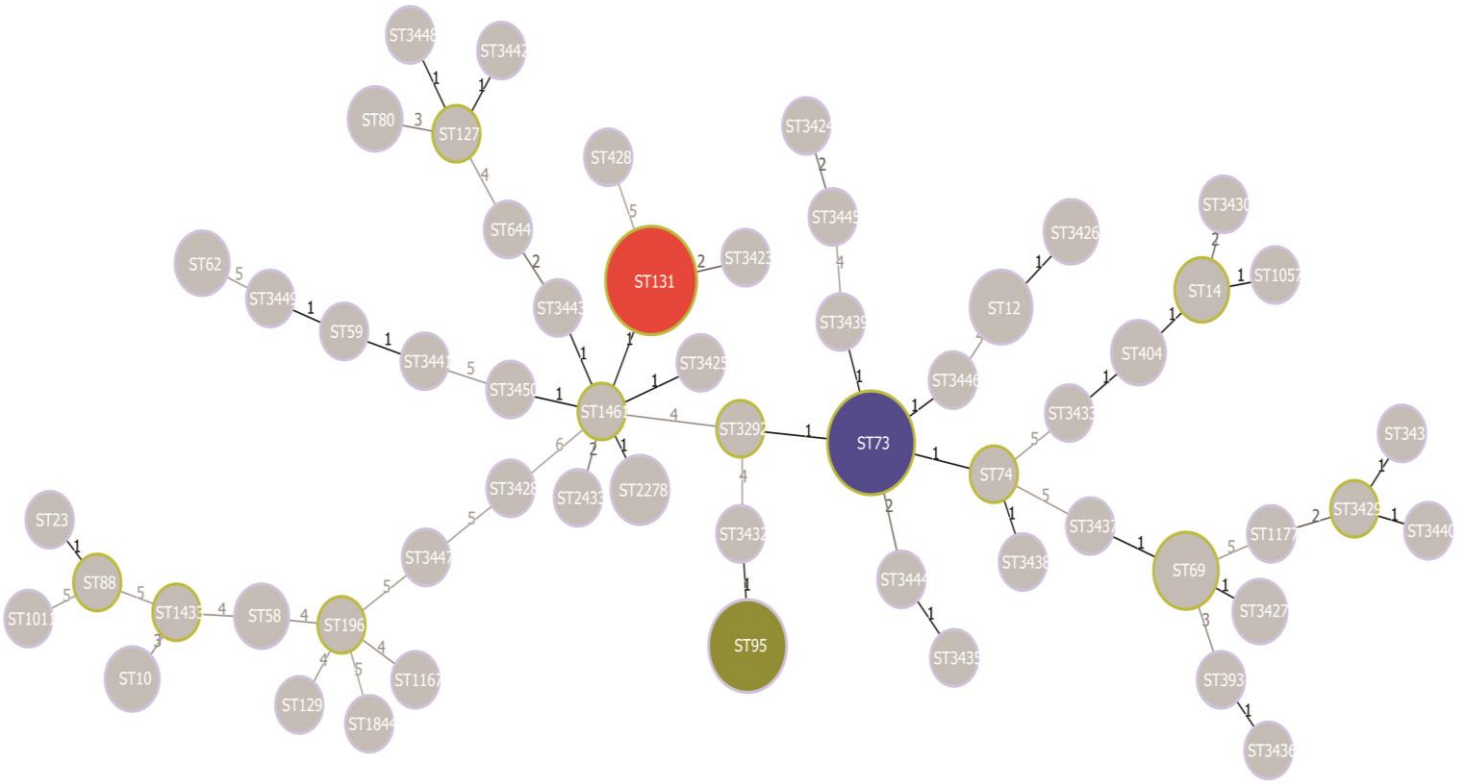


Figure 3. 8: Minimum spanning tree of bacteraemia *E. coli* isolates. Each circle represents one ST, and the size of the circle reflects the number of isolates belonging to this particular ST within the bacteria population. The lines between the circles represent how different their allelic profiles are, where a line labelled by number one means the two linked STs differ in at least one of the seven alleles which is named a single locus variant (SLV), and a cluster of STs linked by SLVs are referred to as a clonal complex. Loss of diversity of the bacteraemia population is displayed with the three dominant STs (ST131 n=30, ST73 n=26 and ST 95 n=16) highlighted in different colours.

The prevalence of ESBL genes was mapped onto the UTI and bacteraemia population MSTs (Figures 3.9 and 3.10). ESBL carriage in the UTI population was focussed on a small number of STs. Nineteen STs out of the total sixty three STs (30.1%) comprising the UTI population contained ESBL positive isolates (Fig 3.9). At the ST level, the aforementioned eight predominant STs within the UTI population exhibited differing levels of ESBL carriage, with the predominant ST73 group composed of 18.7% (n=3 of 16) of ESBL positive isolates. The remaining predominant STs of the UTI population exhibited ESBL positive isolates at the following levels: ST131 (44.4%, n=4 of 9), ST69 (33.3%, n=3 of 9), ST95 (50%, n=3 of 6), ST404 (0%), ST127 (0%), ST141 (0%) and ST10 (0%).

In stark contrast, thirty STs out of the total fifty eight STs (51.7%) of the bacteraemia population contained ESBL positive isolates, which is significantly higher than the number of ESBL positive STs of the UTI population (51.7%/30.1%, p=.016, 95% confidence interval is ±6, Figure 3.10). At the ST level there was variation in the proportions of ESBL positive isolates within the bacteraemia population compared to the UTI population amongst the predominant STs. Sequence types ST131 (50%, n=15 of 30), ST73 (50%, n=12 of 24), ST12 (75%, n=3 of 4), ST10 (100%, n=2), ST14 (50%, n=1 of 2), ST2278 (33.3%, n=1 of 3) all had higher ESBL carriage levels in the bacteraemia isolates, whilst ST95 (46.1%, n=6 of 13) and ST69 (20%, n=1 of 5) showed comparable or lower levels than their UTI counterparts.

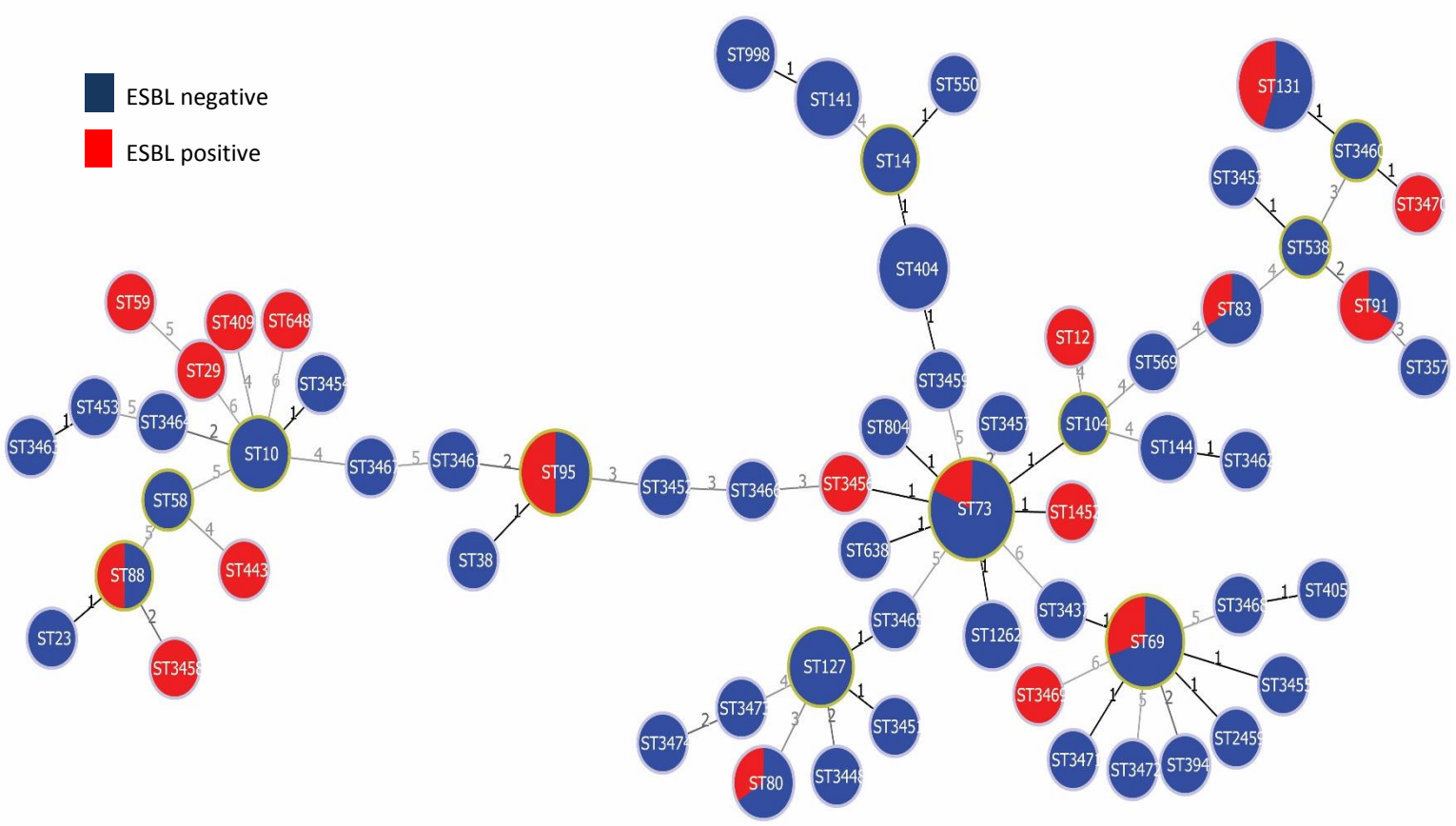


Figure 3. 9: UTI *E. coli* population minimum spanning tree with ESBL carriage. The figure presents ESBL carriage by the UTI *E. coli* isolates mapped on the different STs. We notice that only Nineteen STs out of the total sixty three STs (30.1%) of the UTI population were ESBL positive (Alhashash, et al., 2013).

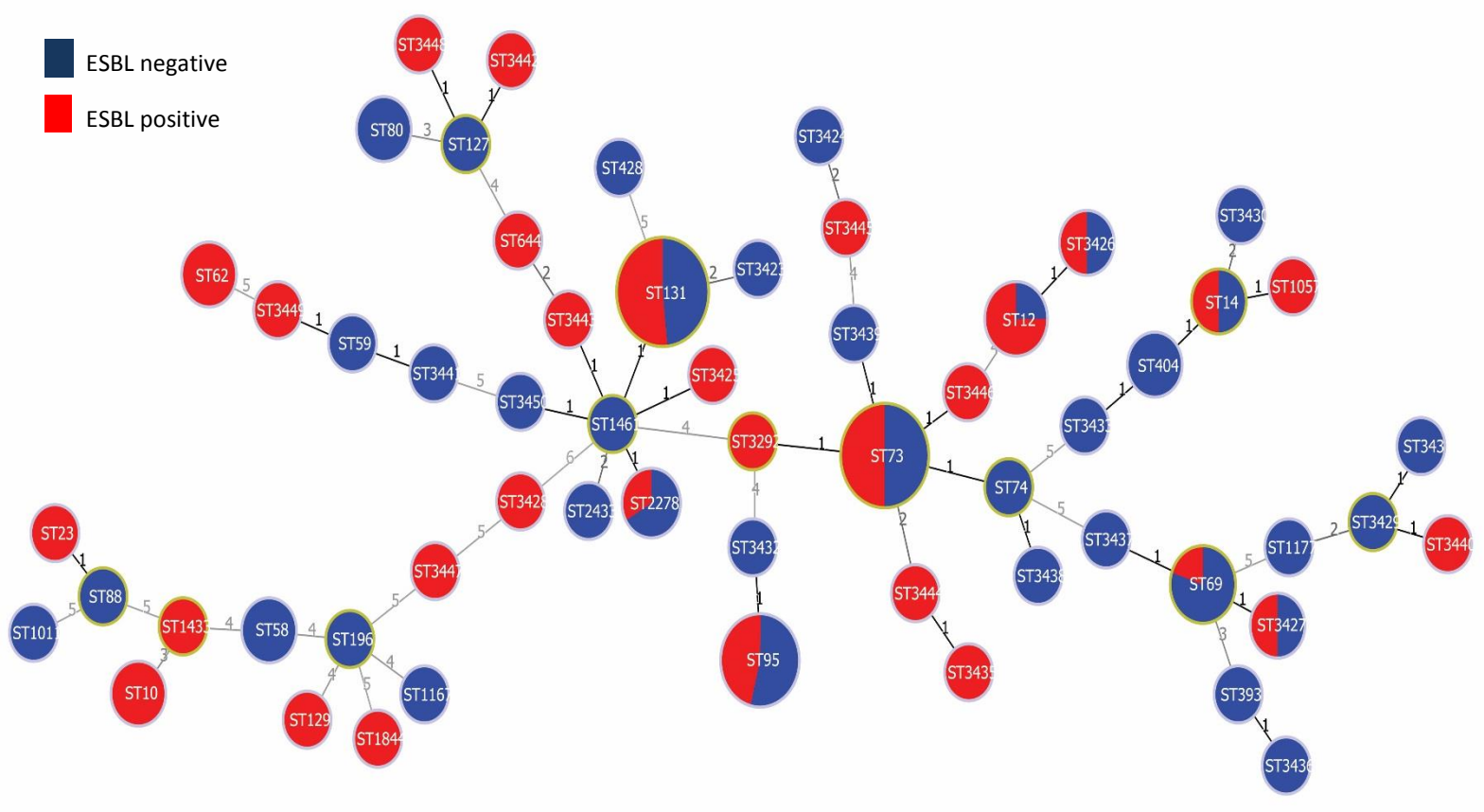


Figure 3. 10: Bacteraemia *E. coli* population minimum spanning tree with ESBL carriage. The figure presents ESBL carriage by the bacteraemia *E. coli* isolates mapped on the different STs. Thirty STs out of the total fifty eight STs (51.7%) of the bacteraemia population were ESBL positive which is significantly higher (51.72%, 30.1%, p=.016) than the ESBL carriage of UTI STs (Alhashash, et al., 2013).

3.2.5 Reduction in diversity in the bacteraemia population is not associated with selection for bacteria with a specific VAG profile

VAGs of all the *E. coli* isolates from the collected samples of bacteraemia and UTI were screened by multiplex PCR (Figure 3.11).

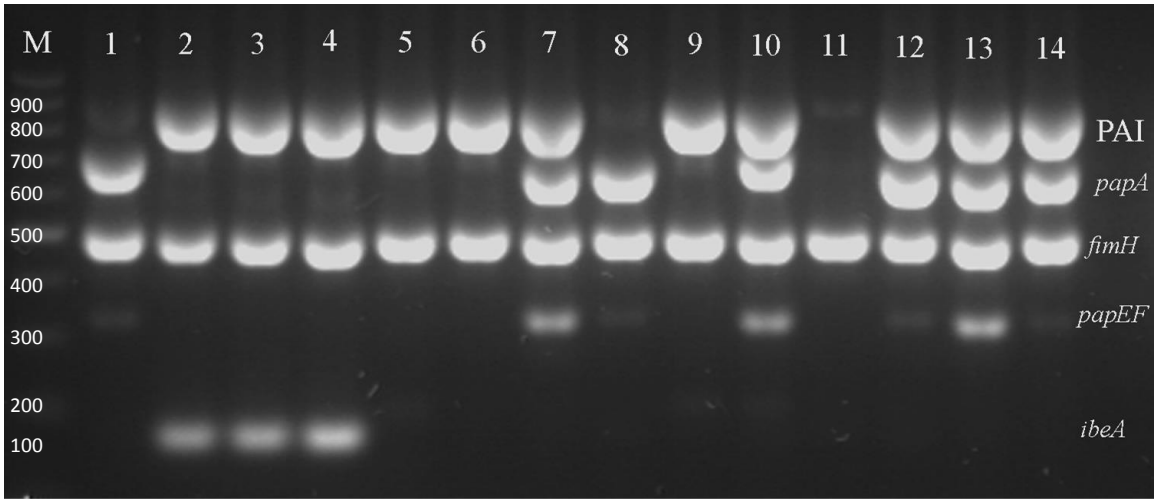


Figure 3. 11: VAGs multiplex PCR gel image for screening of bacteraemia and UTI populations. The figure shows an example of the VAG multiplex PCR screening for the bacteraemia and UTI populations. 100bp DNA ladder was used (M). Lanes 1-14 represents samples U9-U22. This gel image represents pool1 of the VAGs PCR screening for the genes PAI, *papA*, *fimH*, *papEF*, *ibeA*.

All *E.coli* isolates from the two clinical groups of bacteraemia and UTIs were carrying similar high percentages of *fimH* (97.1%/97.6%), *fyuA* (92.1%/91.2%), *kpsMTII* (83.6%/79.2.6%), PAI (76.4%/73.6%) and *traT* (70.7%/64%) as demonstrated in Figure 3.12. Significantly higher acquisition of *iutA*, *papA*, *papEF*, *papG* allele II and *papG* II,III were observed in the bacteraemia *E.coli* population compared to UTI (P=0.006 95% confidence interval is ±5.3, P=0.016 95% confidence interval is ±5.8, P=0.001 95% confidence interval is ±5.7, $P \leq 0.001$ 95% confidence interval is ±5.5 and P=0.006 95% confidence interval is ±5.9 respectively). Significantly higher acquisition of *sfaS* in the UTI population was observed (p=0.005, 95% confidence interval is ±4.5).

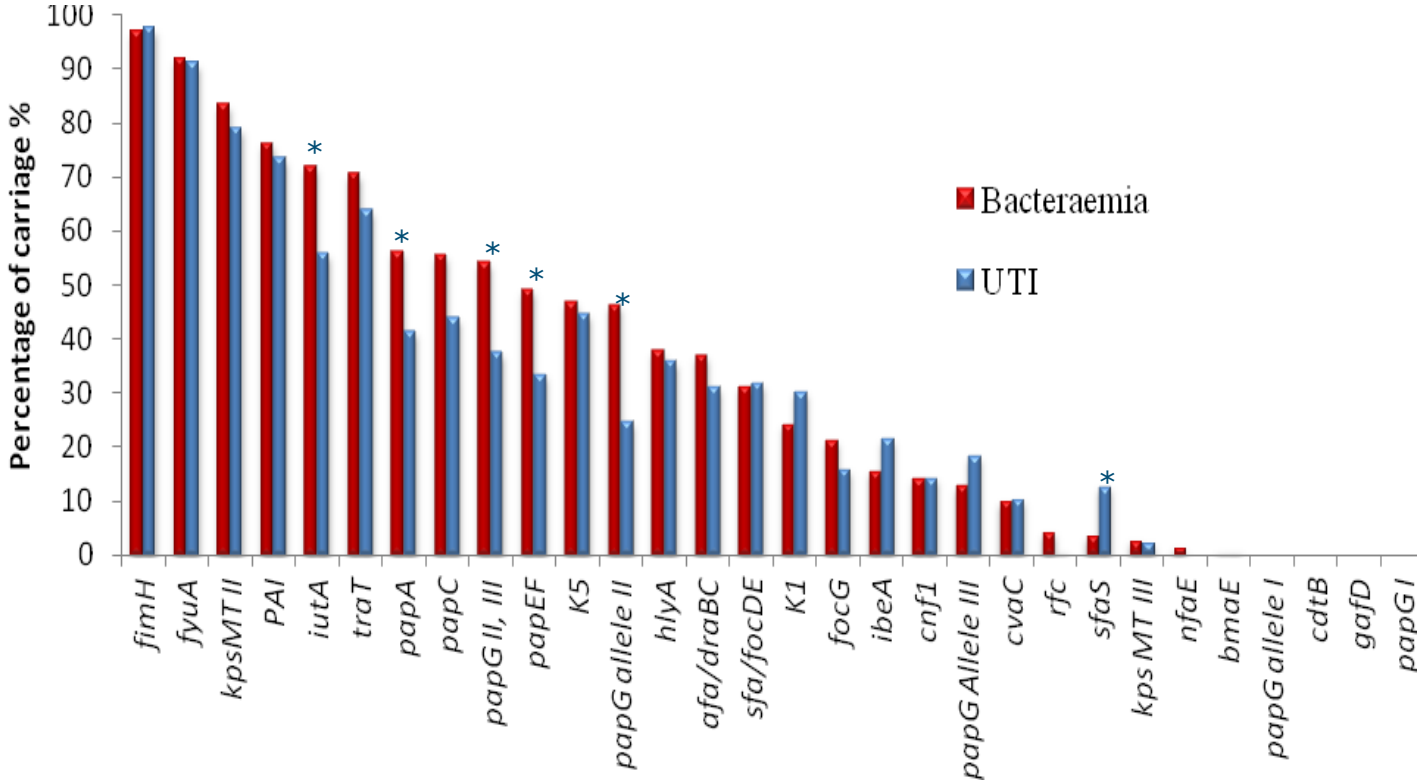


Figure 3. 12: Percentages of VAGs carriage of the bacteraemia and UTI populations. The figure presents VAGs multiplex PCR screening results of all the isolates from the clinical samples of bacteraemia and UTI. (*) Significantly higher acquisition of *iutA*, *papA*, *papEF*, *papG* allele II and *papG* II,III were observed in the bacteraemia *E.coli* population compared to UTI (P=0.006, P=0.016, P=0.001, $P\leq$0.001 and P=0.006 respectively). Significantly higher acquisition of *sfaS* in the UTI population was observed (p=0.005).

The *papA*, *papEF*, *papG* allele II *papG* II,III and *iutA* VAGs were analysed with reference to the main STs of the two clinical populations for their potential to be more associated with the bacteraemia population but inconsistent results were obtained (Figure 3.13). Bacteraemia ST131 isolates exhibited higher carriage than the UTI ST131 isolates for *papA* (41%/11%), *papEF* (34%/0%), *papG* allele II (34%/11%) and *papG* II, III (34%/11%). Slightly higher carriage by bacteraemia ST73 isolates than UTI ST73 isolates was observed for *papA* (70%/62%), *papG* allele II (74%/56%) and *iutA* (77%/62%). However, lower carriage by bacteraemia ST95 isolates than UTI ST95 isolates was observed for *papA* (71%/83%), *papG* II,III (71%/83%) and *papG* allele II (71%/83%) except *iutA* (64%/33%). Other STs such as ST12 isolates from bacteraemia (n=4) and UTI (n=1) exhibited (100%) carriage of *papA* , *papEF* and *papG* II,III, while ST10 isolates from bacteraemia (n=2) and UTI (n=3) possessed none of these VAGs. Variable VAGs carriage patterns were observed in the rest of the STs. Overall, the VAGs data of the two clinical subsets of bacteraemia and UTI were found equally distributed with no statistically significant difference (p=0.675). One interesting finding is that ST73 *E. coli* isolates from both the bacteraemia and UTI populations have almost identical carriage of all the VAGs tested (Figure 3.13).

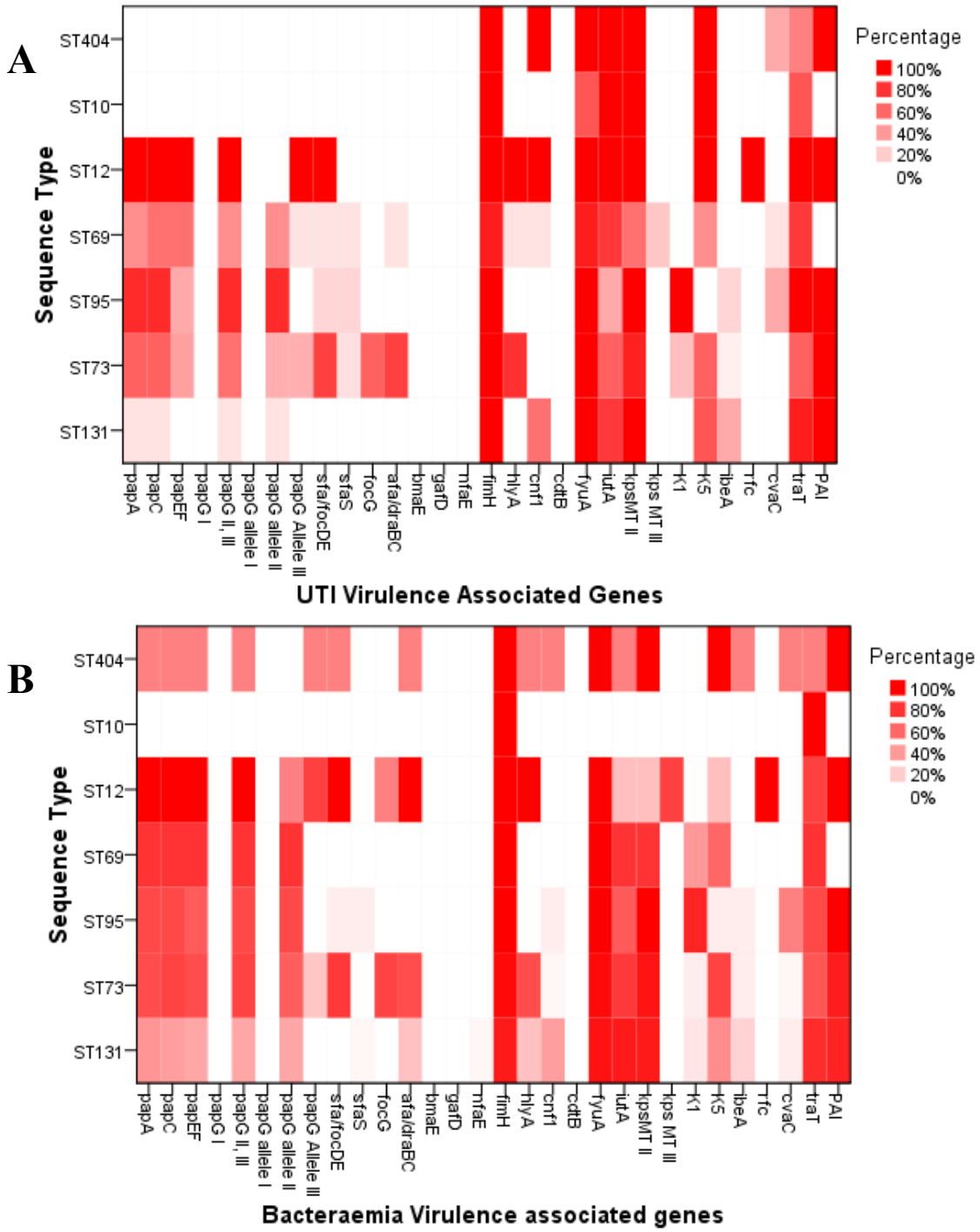


Figure 3. 13: Percentages of VAG carriage of the main STs in the bacteraemia and UTI populations. The figure displays a heat map of the percentages of VAGs carriage of the main STs of the UTI (A) and the bacteraemia (B) sample.

3.3 Discussion

An earlier published study by our research group in the year 2011 reported that *E.coli* isolated from UTI infections were resistant to front line antibiotics with increased resistance to ciprofloxacin and trimethoprim (Croxall, et al., 2011b). Similar levels of antibiotic resistance were observed in *E. coli* UTI isolates in this study, but more importantly, higher levels of antimicrobial resistance were observed in bacteraemia *E. coli* isolates in all the antibiotics tested with higher significant resistance to ciprofloxacin and cefradine. Resistance to ciprofloxacin by bacteraemia *E.coli* isolates has been reported in many studies which concluded that these elevated levels of resistance were observed in community acquired *E.coli* infections due to drug misuse in outpatients and increased prescription which selected for resistant mutants (Cooke, et al., 2010; Sahuquillo-Arce, et al., 2011).

When levels of multi-drug resistance, measured by resistance to different classes of antibiotics, were compared (Figure 3.3) there was a clear increase in the levels of multidrug resistance in the bacteraemia population. The demonstration of a multi-drug resistance phenotype in ExPEC is primarily associated with carriage of ESBLs. Not only was carriage of CTX-M, OXA, and SHV significantly higher in the bacteraemia population, but so too was phenotypic demonstration of the ESBL phenotype using the double disc diffusion method. This raises great concerns as ESBL *E.coli* bacteraemia is being increasingly associated with mortality. Significantly high mortalities due to ESBL producing *E.coli* bacteraemia were recorded compared to mortalities due to non ESBL producing *E. coli* bacteraemia (Melzer and Petersen, 2007). In fact, infection with an ESBL producing *E.coli* isolate was one of the predictors of mortality due to community acquired *E.coli* bacteraemia (Kang, et al., 2010).

New information about the structure of *E. coli* populations from bacteraemia and UTI in the region were obtained by MLST. The previous study by our research group described a diverse UTI *E.coli* population with ST131 being the main ST. The UTI *E. coli* population in the current study was also found to be diverse, but with ST73 as the main ST along with other STs namely ST131, ST69, ST95, ST404 and ST10. Different prevalent STs were observed in the bacteraemia *E. coli* population with a reduction in diversity and three dominant STs, ST131, ST73 and ST95 that are commonly associated with human disease (Manges, et al., 2008). An overview of the minimum spanning trees of the two populations mapped with ESBL carriage revealed that 51.72% of the bacteraemia STs contain ESBL positive isolates compared to 30.16% of the UTI STs. These data provide strong support for the hypothesis that antimicrobial resistance is a driving force for the reduction in diversity within the bacteraemia population of ExPEC. Such an observation has recently been made in a hospital population of *Enterococcus faecium* isolates (Willems, et al., 2012), where the circulation of a restricted population of highly antimicrobial isolates of low level genomic diversity were identified as causing an outbreak in a hospital.

The two populations of bacteraemia and UTI were screened for VAG carriage with the aim of identifying a specific group of VAGs associated with bacteraemia isolates that may explain this reduction of diversity. VAG screening revealed high levels of carriage by all the *E. coli* isolates of *fimH* (type 1 fimbriae), *fyuA* (yersinibactin siderophore receptor), *kpsMTII* (group II capsule), PAI (pathogenicity associated island marker) and *traT* (serum resistance), which is in agreement with a previous study by our research group investigating ExPEC isolates from community and hospital acquired UTI (Croxall, et al., 2011). This is expected as the aforementioned VAGs are well described ExPEC associated VFs (Bien et al., 2012). Significantly higher carriage of *sfaS*, which encodes for the S fimbriae adhesin was observed

in the UTI isolates compared to the bacteraemia isolates. S fimbriae is reported to be associated with binding to distal tubule and glomerular epithelium of the kidney leading to upper urinary tract infections and pyelonephritis (Johnson and Stell, 2000; Jacobsen et al., 2008). Loss of this VF by the bacteraemia *E. coli* isolates may be beneficial as they will not attach to the kidney epithelium and be free to gain access to the bloodstream.

A set of VAGs consisting of *iutA* (aerobactin siderophore receptor), *papA* (P fimbriae major subunit), *papEF* (P fimbriae associated), *papG* allele II (P fimbriae internal regions) and *papG* II,III (P fimbriae flanking regions) were found significantly more in the bacteraemia *E.coli* population compared to UTI. Aerobactin siderophore and P fimbriae adhesin are documented as bacteraemia virulence factors and are associated with upper UTI infection and kidney damage (Ron, 2010; Bien et al., 2012). ST404 displayed a good example of a bacteraemic ST carrying the documented bacteraemia virulence factor P fimbriae adhesin which was absent in all the UTI ST404 isolates. The carriage of the aforementioned set of VAGs was compared across all the STs in the *E.coli* populations of the two clinical samples of bacteraemia and UTI with the aim of identifying bacteraemia specific VAGs but varied profiles were observed with no association with specific STs. One example is *papG* allele II which was described as significantly prevalent in *E. coli* bacteraemia originating from upper UTI and a crucial trait for reaching the kidney (Moreno *et al.*, 2005, Wright *et al.*, 2007), was found completely absent in the bacteraemia and UTI ST10 and ST404 isolates. Similarly, previous work by our group found that the most prevalent ST131 in the urine population had variable VAGs profiles and failed to identify virulence specific genes that associate with a particular ST (Croxall, et al., 2011).

In conclusion, this study has described a difference in the epidemiology of *E. coli* populations of bacteraemia and UTI in terms of their phenotypic and genotypic characteristics. Significantly high numbers of ESBL positive and MDR *E.coli* isolates were observed in clinical cases of bacteraemia. Furthermore, 21.4% of the bacteraemia *E.coli* isolates were ST131. Interestingly ST73 *E. coli* in the UTI population are on the increase and are the most prevalent ST compared to the previous study of our group. They are also one of the three dominant STs of bacteraemia. When VAG carriage in ST73 isolates was compared in the two populations, they were almost identical while the ESBL carriage is significantly higher in the bacteraemia ST73 isolates (50%) compared to the UTI (18%).

We provide strong evidence that the presence of drug resistance driven by ESBL carriage appears to be a driving pressure resulting in a reduction in population diversity in the bacteraemia isolates. This observation carries significance on several fronts, not least of all due to the fact that numerous evolutionary bottleneck selection episodes in bacteria have resulted in lineages of genetically monomorphic pathogens which become human host restricted and more pathogenic, with classical examples in *Yersinia, Salmonella*, *Bacillus*, and *Francisella* (Achtman, 2012). There is a pressing need now to understand the population genomics of this set of isolates to help us identify any other factors which may be driving this selection effect, and to fully gauge the true extent of the diversity in the predominant bacteraemia sequence types, similar to the work which has identified circulation of genetically homogeneous strains of ST131 in health care facilities (Clark, *et al.*, 2012).

Chapter Four

Comparative genome analysis of ST73 isolates from clinical samples of UTI and bacteraemia

4.1 Introduction

4.1.1 Importance of comparative genome analysis

E. coli isolates from different geographical regions and hosts have heterogeneous genomes and may differ in size by up to 1 Mbp (Bergthorsson and Ochman, 1998). This diversity in the genome of *E. coli* species is due to deletion or acquisition of genetic elements by horizontal gene transfer. It was found that horizontal gene transfer events were observed in about 18% of all the open reading frames (ORFs) of MG 1655 *E. coli* (Lawrence and Ochman, 1998). In fact, horizontal gene transfer is largely accountable for the evolution of different bacterial pathotypes, as many VAGs are located on mobile genetic elements such as plasmids, bacteriophages and PAIs (Hacker and Kaper, 2000). Whole genome sequencing has allowed for the ability to comprehensively characterize the genetic diversity and evolution of many related strains which was very difficult before (Metzker, 2009). The characteristics of *E. coli* species of being both a commensal and a versatile pathogen made it the perfect candidate for such studies (Tenaillon et al., 2010). For example, whole genome comparisons of uroseptic *E. coli* CFT073, EHEC strain EDL933 and commensal *E. coli* MG1655 revealed large diversity between their genomes, as they only shared 39.2% of their combined nonredundant proteins. Furthermore, when the CFT073 *E. coli* genome was compared to two other UPEC strains large differences were observed in their PAIs. This indicates that the different ExPEC strains had common vertically evolved genes in their backbone, but had different horizontal gene transfer events dictating their PAIs (Welch et al., 2002). Another application of comparative whole genome analysis is in the investigation of disease outbreaks and bacterial transmission mechanisms. One example is the investigation of a putative disease outbreak of methicillin resistant *Staphylococcus aureus* (MRSA) in a special care baby unit in 2011 (Harris et al., 2013). Twelve infants were infected in a 6 months period and due to the high discriminatory

ability of whole genome analysis it was confirmed that MRSA carriage by a member of the staff at the special baby care unit was the source of the infection which allowed for the infection to persist.

A more related example of the importance of whole genome sequencing in outbreaks investigation is the identification of the hybrid *E. coli* O104:H4 strain in the German outbreak in 2011 (Bielaszewska et al., 2011). It resembled enterohaemorrhagic *E. coli* (EHEC), Shiga toxin encoded by *stx2* prophage. On the other hand, by comparative genome analysis O104:H4 *E. coli* strain was closely related to enteroaggregative *E. coli* (EAEC), particularly strain 55989 responsible for chronic diarrhoea in Central Africa (Mossoro et. al., 2002; Bielaszewska et al., 2011). One of the plasmids of O104:H4 strain called (pAA) is associated with typical (EAEC) adherence that is regulated by the specific (EAEC) virulence regulator AggR. Another (EAEC) related virulence factor released by O104:H4 strain is Pic which is a mucinase that enhances intestinal colonization (Richter et al., 2014). Furthermore, O104:H4 also contain an ESBL plasmid encoding a *bla*-CTXM-15 and *bla*-TEM (Yamaichi et. al., 2014).

4.1.2 Pangenome analysis

To be able to characterize and accurately compare and determine the entire gene contents of multiple genomes, the pangenome approach was developed (Tettelin et al., 2005). The pangenome, (or whole genome) consists of a core genome which includes all the genes present in all the strains in the analysis, a dispensable (or accessory) genome which includes the genes present in some of the strains in the analysis and a group of strain specific genes (Tettelin et al., 2008). This means that each time when a new genome is included in the analysis new strain specific genes are added and the size of the pangenome increases. The core genome is believed to contain the genes that encode for basic biological and phenotypic functions while the

accessory genome encodes for the species diversity and provides selective advantages such as niche adaptation, antibiotic resistance and virulence factors. Therefore pangenome analysis is useful in understanding the bacterial population structure, adaptation and evolution as well as providing targets for vaccines and antibiotics (Tettelin et al., 2008). Pangenome analysis can be used to determine the gene pool of genomes from the same species (Tettelin et al., 2005), or from different species (Gordienko et al., 2013). It can also analyse differences in the gene content of strains from the same species but of different pathotypes or groups (Rasko et al., 2008).

4.1.3 Application of genomics to study *E. coli* ST131

Population genetics analysis revealed that the recent epidemic spread of ST131 *E. coli* isolates was driven by a predominant subclone named *H30*, which was found associated with high carriage of the *fimH30* allele encoding type 1 fimbriae (FimH) (Johnson et al., 2013). Price et al. (2013) constructed a single nucleotide polymorphism (SNP) whole genome phylogeny of 105 ST131 *E. coli* isolates from 5 different countries. The CTX-M-15 producing and fluoroquinolone resistance phenotypes were overlaid on the phylogenetic tree to investigate the evolutionary origins of this dominant pathogen. The resulting phylogenetic tree revealed a subclade consisting of strains carrying the *FimH30* allele, designated *H30*, as well as carrying *gyrA* and *papC* alleles and O type. 95% of the fluoroquinolone resistant and all the ESBL producing isolates fell within this *H30* clade. To further investigate their history of evolution, all the genomes of *H30* isolates were aligned separately against ST131 reference strain NA114 and a high resolution phylogeny was produced. The resulting tree revealed that the acquisition of *fimH30* allele occurred before fluoroquinolone resistance acquisition followed by a large clonal expansion. This fluoroquinolone resistant subclone was named *H30*-R to differentiate it from the ancestral fluoroquinolone sensitive *H30* clade. Interestingly, 91% of the CTX-M-15

producing isolates formed a subclade within the *H30*-R and was named *H30*Rx referring to its extensive resistance. This study concluded that the highly prevalent pandemic CTX-M-15 producing ST131 isolates occur due to the clonal expansion of the *H30*-Rx subclone (Price et al., 2013). The global dissemination of ST131 *E. coli* was also confirmed in another study performed by Petty et al. (2014). 95 *E. coli* ST131 strains isolated between 2000 and 2011 from six different geographical locations were investigated in terms of whole genome analysis, virulence associated genes (VAGs) and CTX-M gene carriage (Petty et al., 2014). Phylogenetic analysis revealed a single distinct lineage of ST131 *E. coli* from the other ExPEC in the study. Within the ST131 *E. coli* lineage three closely related sublineages were identified named A, B and C. The most prevalent sublineage C (79% of the strains) was characterized by fluoroquinolone resistant ST131 *E. coli* isolates, the majority of which were CTX-M-15 positive confirming the *H30*-Rx subclone, and were found to be geographically dispersed (Petty et al., 2014).

4.1.4 Rational and aim

We proposed in Chapter 3 that carriage of ESBL genes by some *E. coli* strains may offer a selective advantage accounting for their dominance and progression to bacteraemia. To further investigate this hypothesis, we aimed to perform comparative genome analysis of a group of ESBL positive and ESBL negative *E. coli* strains belonging to the same dominant sequence type from both the bacteraemia and UTI groups in our collection. These detailed comparative genomics aimed to determine whether the increased prevalence of certain STs and the increased ESBL carriage in the bacteraemia strains were due to the circulation of successful circulating clones as in ST131. The comparative genome analysis also aimed to determine whether the bacteraemia strains were associated with specific bacteraemia associated loci or is ESBL carriage the major determinant for their progression to more serious disease. ST131 was the

most dominant bacteraemia ST, but it was not selected in our investigation. This is because ST131 is well-defined and characterised in the literature (Johnson et al., 2013; Price et al., 2013). Previous work by our pathogen research group at NTU has also contributed to defining this pandemic ST131 clone (Clark et al., 2012). Instead we chose to study ST73 *E. coli* strains that are also dominant in the bacteraemia population in Nottingham and the most prevalent ST in the UTI population. Therefore, it was interesting to investigate whether the high prevalence of ST73 was also due to a successful circulating clone as in ST131. Sequencing of isolates belonging to ST73 was elected for two more reasons. First, there is little research in the literature describing clinical *E. coli* ST73 isolates and any research in this field will provide new knowledge on this widely isolated ST. Secondly and more importantly, ST73 fitted with the hypothesis that ESBL acquisition favoured selection or advantage towards bacteraemia as 50% of the bacteraemia ST73 strains were ESBL positive compared to 18% of the UTI ST73 strains. Specific aims in this investigation were:

A) Construction of SNP based whole genome phylogenetic tree to determine the relatedness of the investigated blood and urine ST73 *E. coli* strains.
B) Comparative whole genome analysis of all the ST73 *E. coli* strains to identify bacteraemia specific loci or ESBL carriage specific loci.
C) Plasmid profiling of all the ESBL positive ST73 *E. coli* isolates to identify plasmid similarity.

4.2 Materials and Methods

4.2.1 Strains

A total of twenty-two strains of ST73 *E.coli* isolates from clinical samples of bacteraemia and UTI were selected for sequencing. The group consisted of ten ESBL positive bacteraemia *E. coli*, two ESBL negative bacteraemia *E. coli*, three ESBL positive UTI *E. coli* and seven ESBL negative UTI *E. coli* all belonging to sequence type ST73. The selected numbers of strains represented the ratio of positive to negative ESBL carriage in the two populations. Details of the selected *E. coli* ST73 isolates are provided in table 4.1A and Table 4.1B. In addition, a suitable reference *E. coli* genome was needed to perform many of the analyses in this chapter. We chose the *E. coli* CFT073 genome (NCBI Reference Sequence: NC_004431.1) as a reference genome since it belonged to sequence type ST73 and is a urosepsis isolate (Welch et al., 2002).

Table 4. 1A: Bacteraemia ST73 *E. coli* isolates selected for sequencing

Strain	ESBL type*	GM	CPM	CAZ	MEM	PTZ	AUG	TM	CIP	CRD	CRD	AP
B10	CTXM-15		R						R			R
B14	Negative	R										
B18	CTXM-15			R			R			R		R
B29	CTXM-15		R						R			R
B36	SHV+CTXM-15			R		R						R
B40	SHV+CTXM-15		R					R				R
B72	CTXM-15						R		R	R		R
B73	CTXM-15		R						R			R
B84	SHV+CTXM-15		R						R			R
B91	CTXM-15			R			R			R		R
B102	Negative	R										
B134	CTXM-15+OXA			R		R	R	R				R

*Isolates were screened by PCR for ESBL carriage as described in section 2.4. CTXM positive PCR products were sent and sequenced by Source Bioscience LifeSciences. Antibiotic susceptibility profiles were performed as described in section 2.3: Resistant (R), gentamicin (GM), cefpodoxime (CPM) ceftazidime (CAZ), meropenem (MEM), piperacillin-tazobactam (PTZ), co-amoxiclav (AUG), trimethoprim (TM), ciprofloxacin (CIP), cefradine (CRD), nitrofurantoin (NI), and ampicillin (AP).

Table 4. 1B: UTI ST73 *E. coli* isolates selected for sequencing

Strain	ESBL type*	GM	CPM	CAZ	MEM	PTZ	AUG	TM	CIP	CRD	CRD	AP
U1	Negative					R						
U7	Negative											
U21	Negative											R
U24	Negative											
U30	Negative											
U36	Negative							R				R
U42	CTXM-15		R							R		R
U48	Negative											
U50	CTXM-15						R			R		R
U76	SHV+CTXM-15		R			R	R	R	R			R

*Isolates were screened by PCR for ESBL carriage as described in section 2.4. CTXM positive PCR products were sent and sequenced by Source Bioscience LifeSciences. Antibiotic susceptibility profiles were performed as described in section 2.3: Resistant (R), gentamicin (GM), cefpodoxime (CPM) ceftazidime (CAZ), meropenem (MEM), piperacillin-tazobactam (PTZ), co-amoxiclav (AUG), trimethoprim (TM), ciprofloxacin (CIP), cefradine (CRD), nitrofurantoin (NI), and ampicillin (AP).

4.3 Results

4.3.1 Sequences assembly by Velvet and PAGIT software

Improved quality genome assemblies were created for all twenty two ST73 *E. coli* isolates using Velvet and PAGIT (sections 6.2 and 6.3 of appendix). Assembly metrics are described in Table 4.2.

Table 4. 2: Details of the twenty two sequenced ST73 *E. coli* isolates genomes after PAGIT assembly.

Strain	Sequence size (bps)	Number of contigs	N50 (bps)
B10	5173276	106	108731
B14	5099552	158	113475
B18	5120683	125	122417
B29	5264174	168	101820
B36	5191523	152	125321
B40	5257611	165	103549
B72	5158804	110	134654
B73	5150717	156	123129
B84	5182704	137	134972
B91	5155911	197	79515
B102	5075956	160	87164
B134	5230535	154	116039
U1	5243352	151	123112
U7	5176031	145	126228
U21	5145668	162	113459
U24	5120446	147	110560
U30	5287542	160	139416
U36	5162072	138	114804
U42	5188710	155	106920
U48	5080928	112	113440
U50	5256879	145	117621
U76	5189037	140	133761

4.3.2 SNP based phylogeny

SNP analysis of the twenty-two ST73 genome sequences showed a high level of polymorphism of 15,278 SNPs between the strains. The diversity is clearly observed in Figure 4.1 where there is no apparent clustering of bacteraemia strains, or UTI strains but rather they are randomly distributed around the phylogenetic tree. This rules out the probability of an *E. coli* ST73 clone outbreak in our bacteraemia or UTI clinical samples. When the ESBL carriage was mapped to the tree, ESBL positive strains were also randomly distributed across the tree. This also confirms that the increase of ESBL carrier strains, like the increase in the number of ST73 in the bacteraemia population, is not due to a circulation of a successful strain.

When the relationship between the twenty-two strains in the phylogenetic tree was examined in terms of ESBL carriage, two potential groups were recognised on opposite sides of the tree as indicated by circles in Figure 4.1. The first potential group consisted of seven ESBL positive strains, five from bacteraemia (B29, B73, B36, B84 and B10) and two from UTI (U42 and U76). The second potential group consisted of four ESBL positive strains, three from Bacteraemia (B18, B72 and B134) and one from UTI (U50). These potential groups were investigated further in terms of the SNPs distances between themselves and their neighbouring strains.

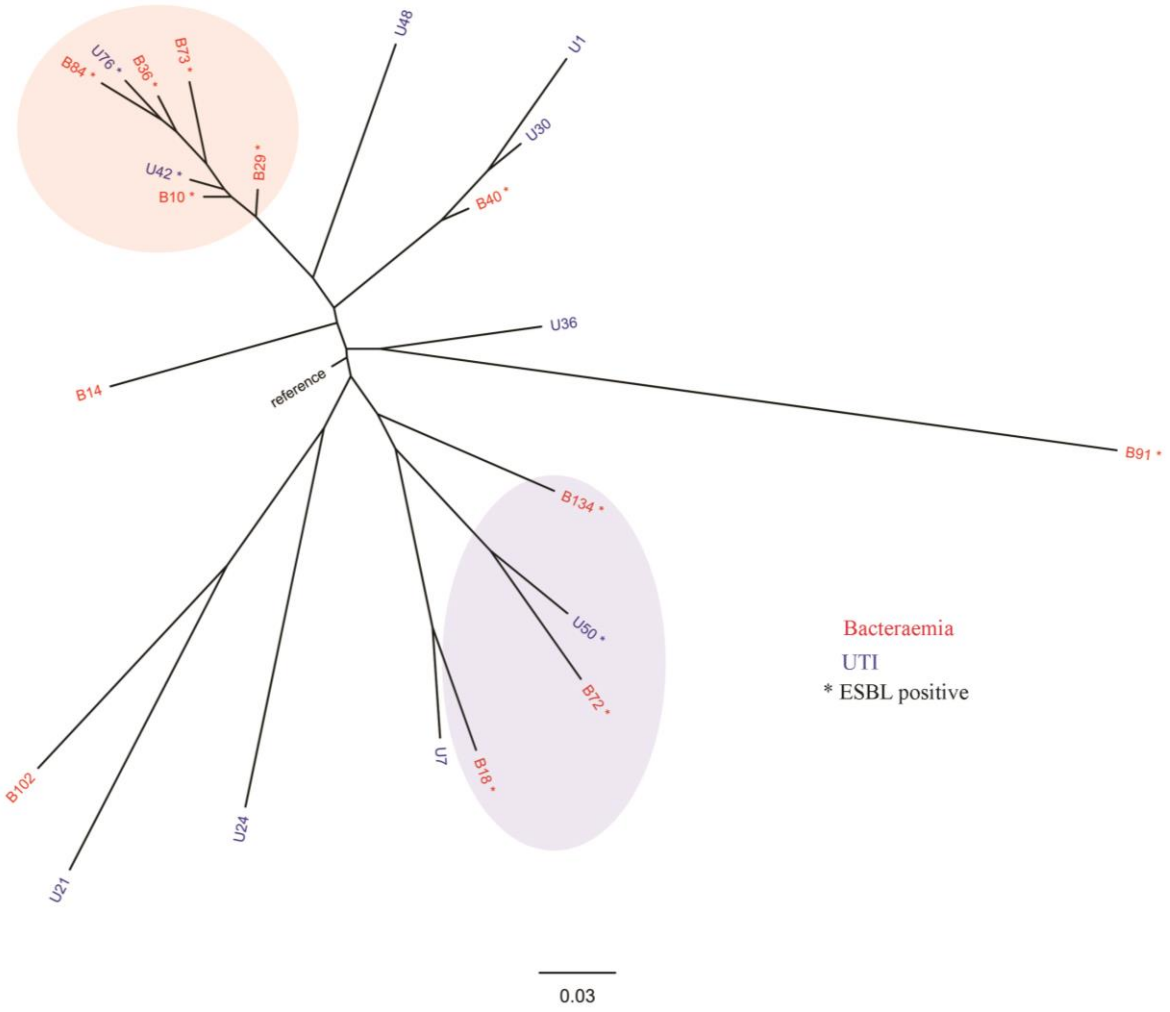


Figure 4. 1: SNPs based Phylogenetic tree of the twenty two *E. coli* ST73 against *E. coli* CFT073 reference genome using maximum likelihood method. The tree displays a diverse group of *E. coli* ST73 strains without clear clustering. Bacteraemia isolates coloured red, UTI isolates coloured blue, ESBL positive strains indicated by a star sign, yellow circles define potential ESBL positive group of strains and a blue circle define another potential ESBL positive group of strains.

4.3.2.1 Investigation of pairwise distance

Pairwise distance analysis was performed to determine the number of SNPs between all the strains and the reference genome in relation to each other. Table 4.3 presents the pairwise distance with colour scaling depending on the number of base differences between the strains where higher numbers correlate with darker colour. A rectangular layout of the SNP based phylogenetic tree is displayed in Figure 4.2. We noticed that the distances between the four strains (B134, U50, B72 and B18) were very large ranging between 1453 and 3119 SNPs, which shows a high level of heterogeneity between their genomes. The other group of the seven ESBL positive strains (B10, U42, B84, U76, B36 B73 and B29) was found to be slightly more related to each other as the SNPs distances between them were low ranging from 416 to 989 bases. In both cases MDR ST73 do not appear to be a successful circulating clone like ST131 observed in the previous study by our group where nine of the ten sequenced strains shared only 10 to 60 SNPs between them (Clark et al., 2012).

Table 4. 3: Pair wise distances matrix measured by the number of SNPs between the strains

	Ref	U48	U50	B84	B29	U42	B14	B40	B73	U24	U30	U1	U76	B18	U36	B72	U7	B10	B36	B91	B102	U21	B134
Ref																							
U48	1887																						
U50	1874	3109																					
B84	1800	2475	2901																				
B29	1079	1966	2411	1079																			
U42	1223	2046	2523	989	564																		
B14	1589	2596	2836	2658	2074	2210																	
B40	1232	2278	2443	2337	1663	1779	2206																
B73	1551	2011	2772	1133	790	836	2350	1896															
U24	2471	3679	3743	3510	3061	3187	3514	3196	3427														
U30	1598	2458	2696	2421	1932	2020	2412	816	2017	3340													
U1	2117	2916	3077	2898	2417	2513	2925	1276	2525	3792	1039												
U76	1724	2336	2870	728	979	841	2539	2209	1037	3468	2309	2776											
B18	2235	3497	2840	3230	2727	2869	3294	2868	3085	3831	3047	3520	3198										
U36	1252	2359	2712	2093	1874	1825	2395	2117	2216	3355	2424	2936	2068	3188									
B72	2182	3262	1453	3173	2664	2777	3145	2669	2997	3714	2854	3234	3119	3119	2936								
U7	2168	3444	2781	3153	2656	2795	3224	2814	3042	3676	3022	3432	3136	1364	3107	3008							
B10	1207	2032	2487	944	499	416	2175	1736	811	3166	2027	2479	840	2861	1833	2772	2788						
B36	1664	2282	2814	687	901	779	2501	2155	941	3377	2258	2732	626	3130	2023	3061	3051	773					
B91	3854	4848	5082	5025	4545	4617	4502	4531	4771	5711	4752	5201	4983	5483	4361	5221	5387	4611	4932				
B102	2828	3942	3505	3697	3192	3338	3786	3484	3545	3923	3616	4075	3695	3565	3727	3413	3471	3339	3618	5963			
U21	3028	4142	4013	4139	3587	3705	4134	3611	3949	4392	3936	4288	4123	4248	3913	3995	4141	3686	4090	6026	3206		
B134	1506	2948	2282	2697	2103	2217	2616	2206	2503	3427	2427	2853	2606	2690	2433	2634	2435	2221	2545	4757	3443	3798	

Table 4.3 displays the number of SNPs bases between the strains. The higher the number of SNPs between any two strains the more distantly related to each other.

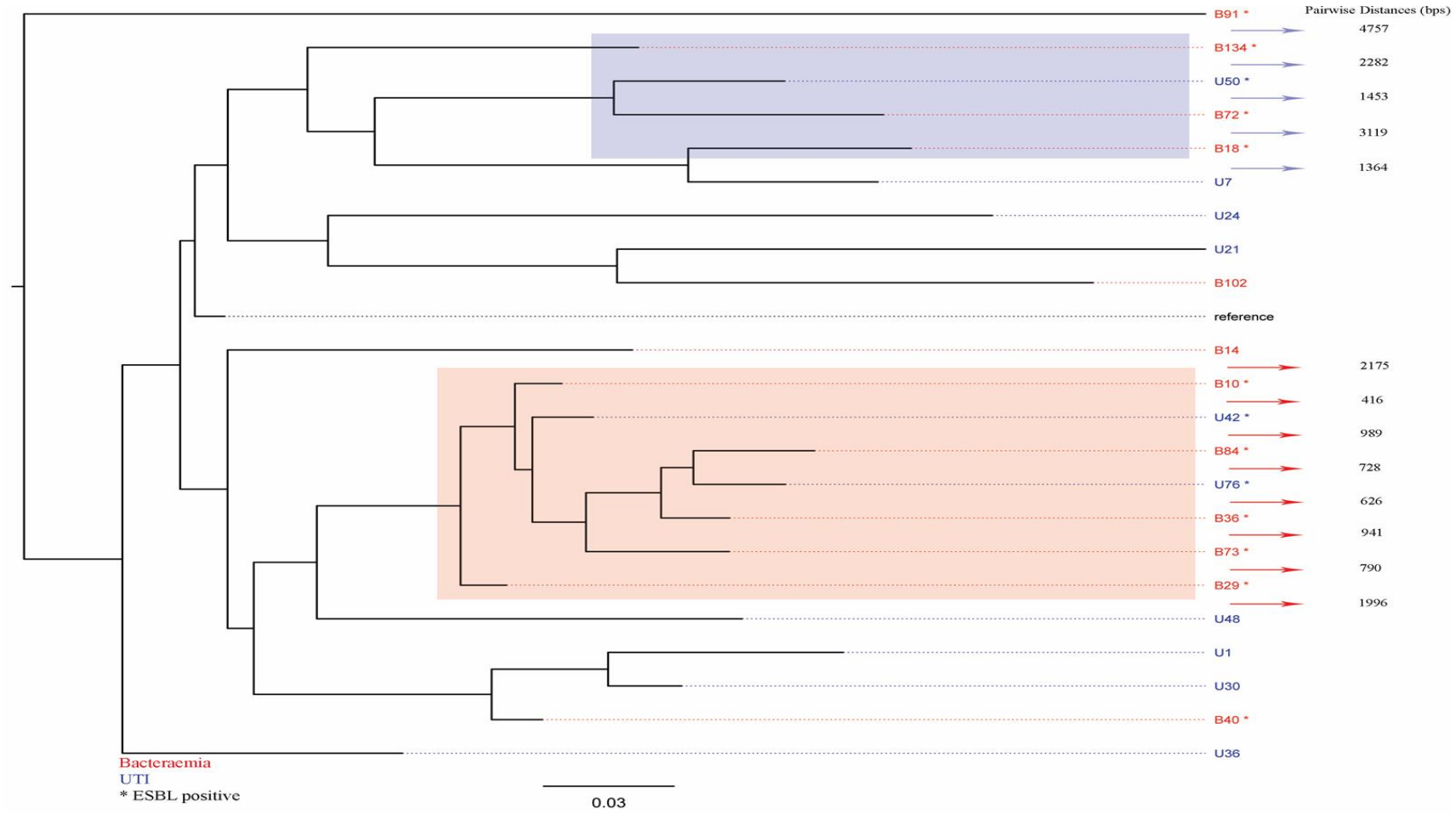


Figure 4. 2: pairwise distances of SNPs based phylogenetic tree of the 22 strains of *E. coli* ST73 against CFT073 *E. coli* reference genome.

Two groups of ESBL positive strains highlighted in different colours and the arrows represent the SNPs distance between neighbouring strains.

4.3.3 Comparative genome analysis

Comparisons between the genomes of all twenty-two ST73 *E. coli* strains aimed to investigate whether there were genomic regions or loci unique to the bacteraemia ST73 *E. coli* strains. Comparisons also aimed to investigate whether regions or loci were specific only to ESBL carrier ST73 *E. coli* strains from bacteraemia and UTI clinical samples. Comparative genome analysis was made using Gegenees (Agren et al., 2012).

4.3.3.1 Determination of the core and pangenome

The determination of the core and pangenome of all the twenty-two ST73 *E. coli* strains was performed by Gegenees software (Gegenees.org, 2014). The pangenome was constructed by fragment analysis approach where fragments of all the genomes were compared to each other by pairwise BLAST of their nucleotides (BLASTN) for similarities and only unique fragments were added gradually to the pangenome (Figure 4.3). The pangenome is represented on the top of the figure by a black horizontal line which consisted of 8,309,000 bps in 41545 fragments (each fragment consists of 200 bps) divided by many subsequences. The total number of genes in the pangenome was 10,095 genes reflecting the diversity of the ST73 strains. In Figure 4.3, several black lines are displayed below the pangenome line representing the twenty two genomes of ST73 *E. coli* strains in the analysis. Fragments of each genome in the analysis were compared to the pangenome on the top and when similar fragments were identified a dot is plotted over the genome line. The collection of dots similar to the pangenome form a peak. The green peaks on the left of each genome in Figure 4.3 represent the core genome in the analysis which consists of genomic regions shared by all the strains under investigation. The total fragments contained in the pangenome were 41545 and the core genome represented 59.8% of the pangenome (24,826 fragments). To the right of the core genome of each strain are the accessory genomic regions showing a high level of genomic diversity between the twenty-two

strains in the analysis (Figure 4.3). This is in line with published studies of *E. coli* pangenome analyses. Pangenome analysis of intestinal, ExPEC and commensal *E. coli* showed an open pangenome model where approximately 50% of the gene content represented the core genome and each new genome sequence contributed about 300 novel genes (Lloyd et al., 2007; Rasko et al., 2008). This suggests an open pangenome in ST73 *E. coli* in this project.

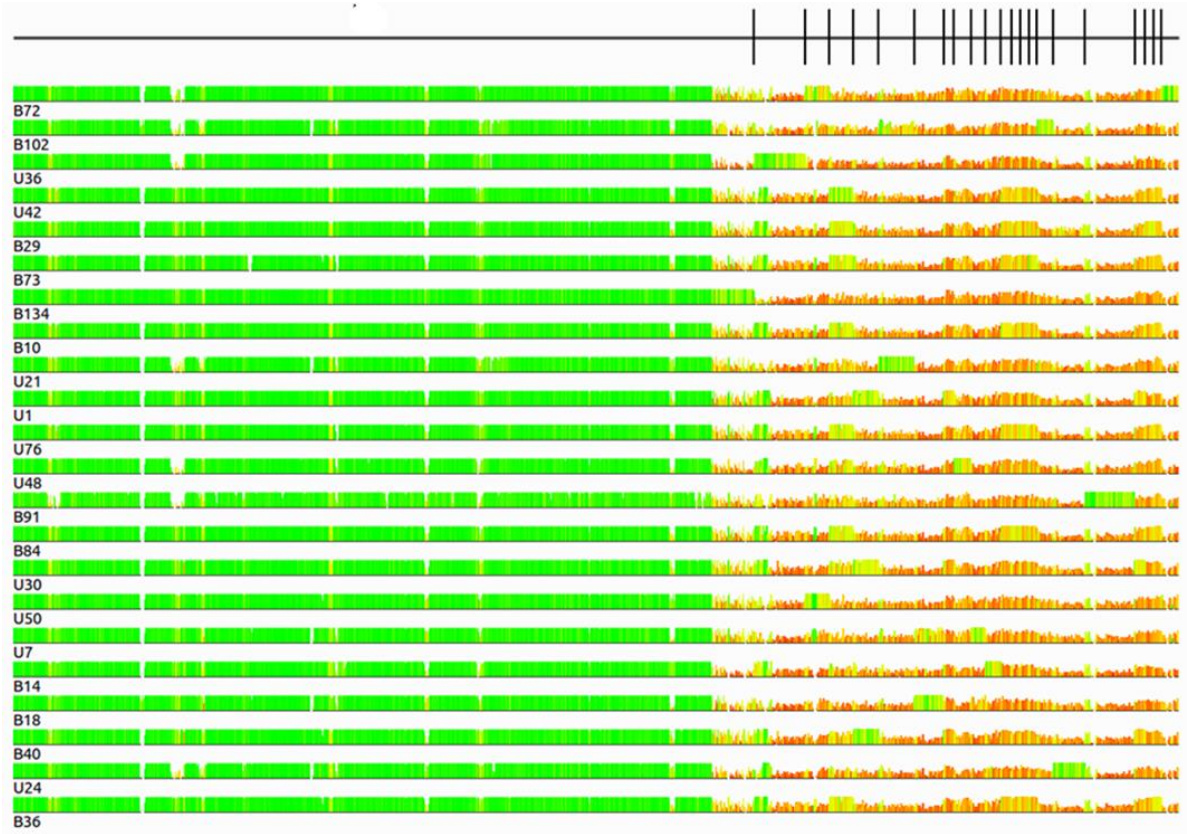


Figure 4. 3: Overview of the core and pangenome analysis by Gegenees software. The figure displays the pangenome at the top by a horizontal black line. Below it are black lines representing all the twenty two ST73 *E. coli* strains under study. The peaks on top of each genome are colour coded red, orange, yellow and green according to the similarity of the fragment compared to the pangenome where red represents least similarity and green represents exact similarity. The green peaks on the left of each genome line represent the core genome in the analysis which contains similar genomic regions shared by all the genomes in the study.

4.3.3.2 Investigation of bacteraemia and ESBL carrier specific loci

Using Gegenees software, all the genomes under investigation were fragmented and aligned using multithreaded pair-wise nucleotide comparisons (BLASTN). The alignment performed by Gegenees assigns an average score of the pair-wise comparisons of all the fragments against each other for each genome to determine their degree of similarity (Agren et al., 2012). These similarities were described in a matrix and were displayed in a heat plot (Figure 4.4). The visualisation of the whole genome alignments in the heat plot provided an overview of the phylogenetic relationship between the strains (Agren et al., 2012). We notice in Figure 4.4 that the same seven ESBL positive strains (B10, B29, B36, B73, B84, U42 and U76) which were identified earlier as being most similar to each other than the other strains were also found very similar in Gegenees alignment, indicated in the green colour and the high average BLASTN scores at the top of Figure 4.4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1: B10	100	96	96	96	96	96	95	92	92	92	91	91	91	92	91	90	89	90	94	94	94	85
2: B29	96	100	96	96	96	96	95	92	92	92	91	91	90	92	91	90	88	89	94	94	94	85
3: U76	97	97	100	97	97	96	96	93	93	93	92	92	91	93	92	90	89	90	95	94	94	86
4: B36	97	97	97	100	97	96	96	93	93	93	92	92	91	93	92	90	89	90	95	95	95	86
5: B84	96	97	97	97	100	96	96	93	93	93	92	92	92	93	92	90	89	91	95	94	94	86
6: U42	96	97	97	97	97	100	96	93	93	93	92	92	92	93	92	91	90	91	95	94	94	86
7: B73	96	97	97	97	97	96	100	93	93	93	91	92	91	93	92	90	89	90	95	95	94	86
8: B14	94	95	94	95	95	94	94	100	93	93	92	92	92	93	93	92	90	90	95	94	95	86
9: U48	95	95	95	95	95	95	94	93	100	93	93	92	91	93	93	91	90	92	94	94	94	88
10: U50	93	94	93	94	94	94	93	92	91	100	96	94	93	95	91	91	90	91	94	94	94	87
11: B72	92	92	92	92	93	92	91	91	91	95	100	93	92	93	91	91	91	91	93	92	92	87
12: U7	92	93	93	93	93	92	92	91	91	94	93	100	95	93	90	90	90	91	93	93	93	85
13: B18	92	92	92	92	93	92	92	92	90	93	93	96	100	94	91	91	91	90	93	93	93	85
14: B134	92	92	92	92	93	92	91	91	90	93	92	92	92	100	91	90	90	90	93	93	93	85
15: U36	92	92	92	92	92	92	91	91	91	91	91	90	90	92	100	90	89	89	92	92	92	85
16: U24	91	92	92	92	92	92	91	91	91	91	91	91	91	92	91	100	91	91	92	92	92	85
17: B102	91	92	92	92	92	92	91	91	91	92	93	92	93	93	91	92	100	94	92	92	92	86
18: U21	91	92	91	91	92	91	91	90	91	92	92	92	91	92	90	91	93	100	92	91	91	86
19: U30	93	93	93	93	94	93	93	92	91	92	91	91	90	92	90	89	88	89	100	96	96	84
20: B40	93	93	93	93	93	93	93	92	91	92	91	91	91	93	90	89	88	89	97	100	96	84
21: U1	93	93	93	93	94	93	93	92	91	92	91	91	91	93	91	90	89	89	97	97	100	84
22: B91	86	87	87	87	87	86	86	86	87	87	87	86	85	87	86	84	84	86	87	86	86	100

Figure 4. 4: Heat plot display of Fragmented all-all comparisons of Gegenees software. The figure displays the similarities of the aligned twenty two genomes by as a heat plot and BLASTN (200/100) score matrix. We notice the green colour group of strains at the top (U42, U76, B10, B29, B36, B73 and B84) with the high average scores which indicate their similarity. The target strains or group (green colour), the background strains or group (red colour) and the reference strain from the target group (black colour) are displayed on the left.

An important feature of the Gegenees software is the ability to analyse the alignment dataset to identify genomic regions specific or unique to a target group, also called target signature, when compared against a background group. For this purpose as seen in Figure 4.4, all the bacteraemia ST73 *E. coli* strains were selected as a target group to investigate bacteraemia specific signature, B134 strain was selected as a reference genome from the target group, and the ST73 UTI strains were selected as background group. Target specific loci or target signature analysis was performed by determining a biomarker score based on fragment comparisons of

the worst false positive in the background group against the worst false negative in the target group for each fragment in the reference genome (Agren et al., 2012). The output of the target signature analysis is displayed in Figure 4.5. The highest value of the biomarker scores is 1, which indicates perfect conservation and no cross reaction and displayed in green colour. In Figure 4.5 two grey blocks are displayed and each of them are separated by a black line which represents a total of 52821 fragments of the reference genome used as a coordinate axis starting from the top rectangle on the left and ends in the bottom rectangle on the right. Two biomarker stringency options were applied where the biomarker scores for each fragment can be plotted against the reference coordinate axis displayed in Figure 4.5. The highest stringency option (max/min or 100% target group specific) is displayed above the reference axis and presents genomic regions or loci only specific to the target group. No target signature or conserved genomic regions were found specific only to the ST73 bacteraemia strains. For that reason and to investigate further, it was useful to select a more relaxed stringency option (max/average or 80/20% target group specific) where a target signature may be identified but it may be absent or not conserved in all the bacteraemia ST73 strains. A large number of associated regions indicated by complete green columns were identified and they were defined as plasmid DNA shown on the right of the bottom rectangle (Figure 4.5). Small genomic regions in the middle were also identified. These highlighted regions were investigated further. These regions were found to be segments of unspecified sequences labelled with repeated N's and the plasmid highlighted regions were found associated with conjugation proteins shared by all the isolates and of no particular interest other than ESBL genes. They were highlighted because Gegenees does not only identify presence or absence of genes but also identifies group specific alleles of CDS. Not all the sequenced bacteraemia ST73 strains were ESBL positive, hence specific loci associated with ESBL plasmids were not found in the high stringency conditions.

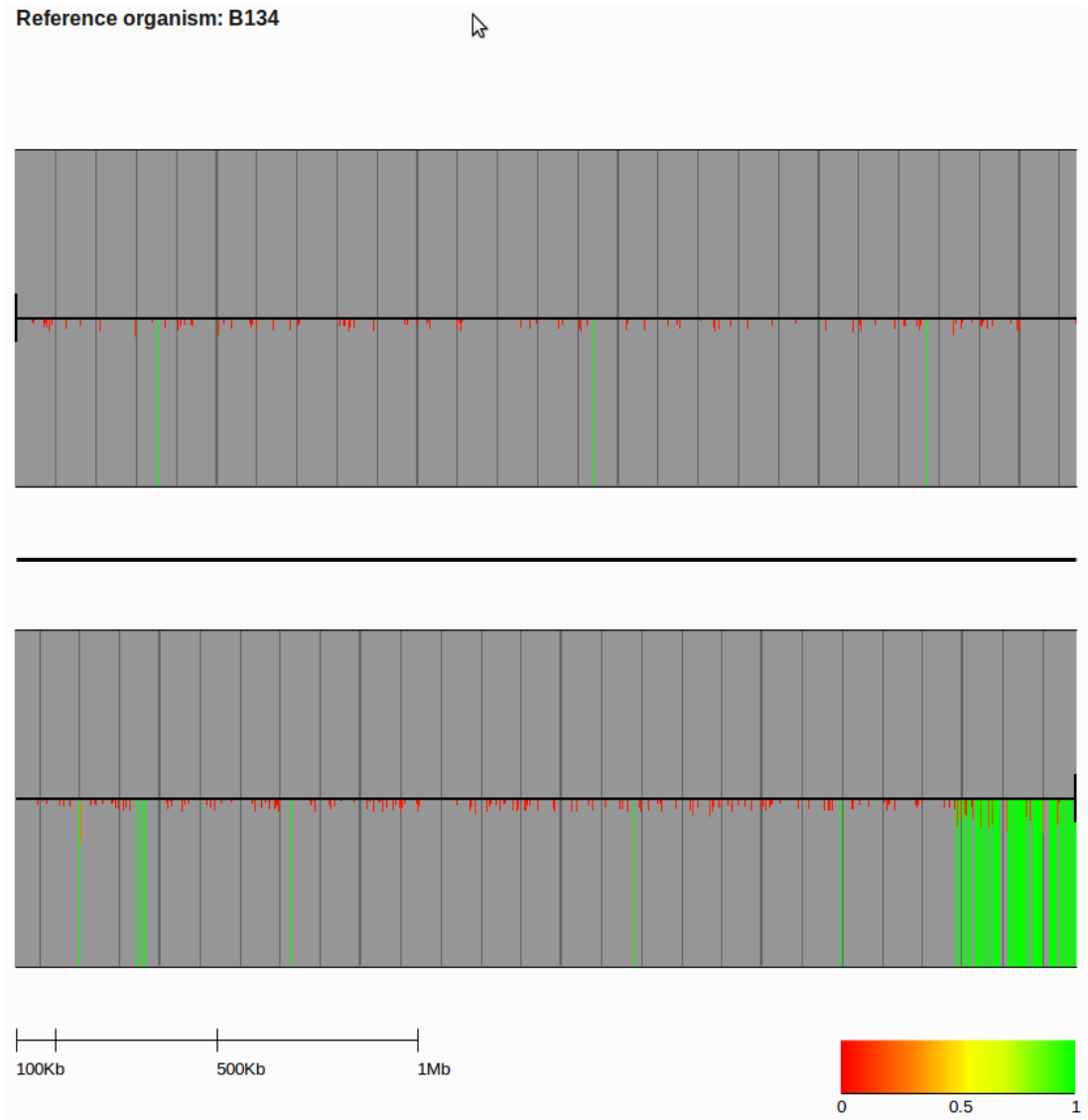


Figure 4. 5: Biomarker score overview of target specific loci (target signature) analysis by Gegenees software. The figure displays the output of the target specific loci analysis where all the bacteraemia strains were selected as a target group against a background of all the UTI strains. The black line in the middle of the two blocks represents the reference strain (B134) genome fragments starting from the top left and ends in bottom right. The reference genome size scale is displayed and a coloured biomarker scale is displayed where the highest score for completely unique region is 1 and coloured green. The high stringency option output is displayed above the reference coordinate axis in the two blocks with no target signature

identified. The less stringent option is displayed below the reference coordinate axis with large number of target specific regions or loci identified in the plasmid region.

The target signature analysis was repeated for the twenty-two ST73 *E. coli* strains where all the ESBL positive strains from bacteraemia and UTI samples were selected as a target group against a background group of all the ESBL negative strains from bacteraemia and UTI samples. A similar analysis output was obtained where no target specific signature was identified except for the plasmid regions as described above. This validates our finding that genome specific regions or loci were found highly associated with the plasmid region in both tests for bacteraemia target signature and ESBL carrier target signature.

4.3.4 Plasmid profiling by S1 PFGE

Detection and sizing of the CTX-M plasmids carried by the ST73 *E. coli* was performed (Figures 4.6A and 4.6B). Different plasmids were found to be carried by most of the tested ST73 strains and some carried more than one plasmid. This finding rules out the possibility that the increased prevalence of ST73 in the bacteraemia population, associated with increased antibiotic resistance, is due to the dissemination of a single successful plasmid. The different plasmids and their relative sizes are summarized in Table 4.4. The CTX-M negative ST73 strains showed no plasmid carriage by profiling with S1 PFGE (B14, U24 in Figure 4.6A and B102 in Figure 4.6B). Plasmids of the same sizes were found to exist in some strains, with an 82 kb plasmid carried by B72 and B134; a 48.5 kb plasmid carried by B18 and U50; and a 33.5 kb plasmid carried by (B18 and B72). The commonly occurring size of plasmid was 112 kb which was carried by B10, B29, B73, B84, U42 and U76 strains. Interestingly, these are the

same strains that were identified earlier by SNP based phylogeny (section 4.3.1) as being similar and grouping together (Figure 4.2).

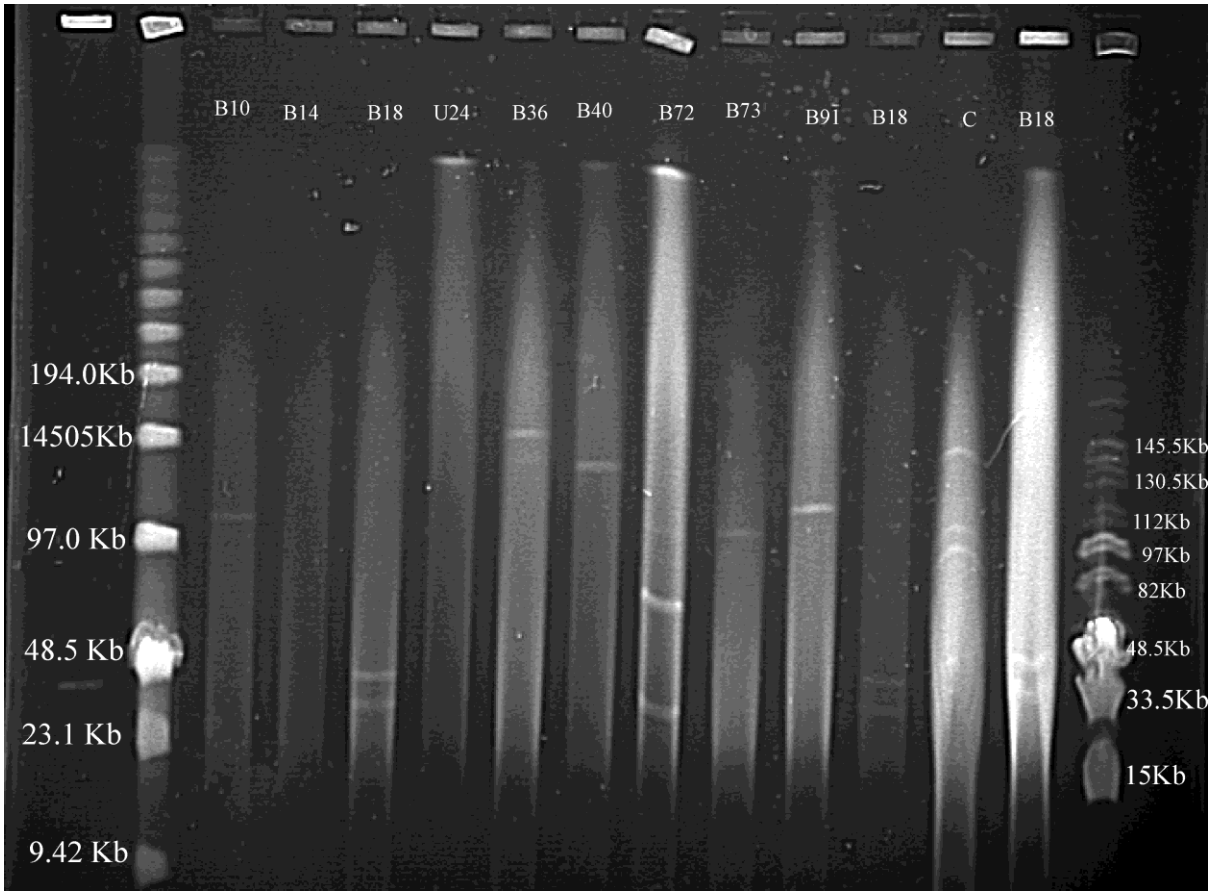


Figure 4. 6A: S1 nuclease PFGE gel image. The figure shows the different plasmids detected in the CTX-M carrier strains tested. On the left is low range DNA ladder and on the right is mid-range DNA ladder. C represents the positive control NCTC *E. coli* 13353.

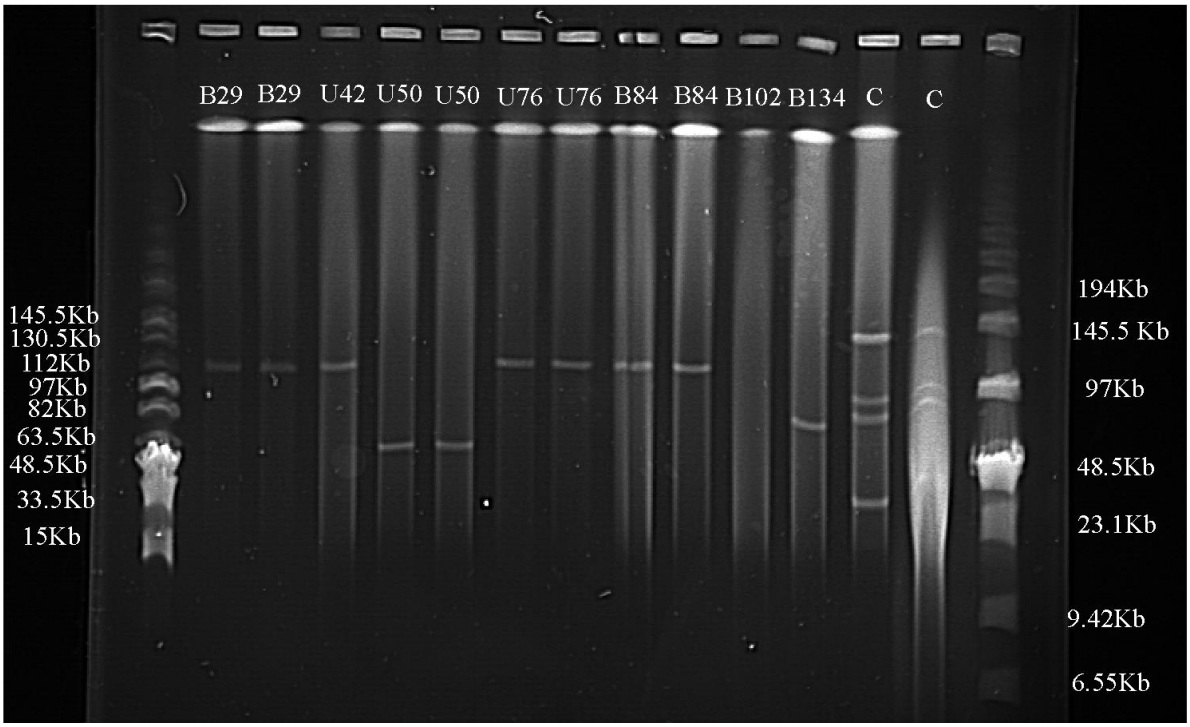


Figure 4. 6B: S1 nuclease PFGE gel image. The figure shows the different plasmids detected in the CTX-M carrier strains tested. On the left is low range DNA ladder and on the right is mid-range DNA ladder. C represents the positive control NCTC *E. coli* 13353.

Table 4. 4: List of the different plasmids and their size as determined by S1 PFGE

S1 PFGE plasmid size	Strain Name
33.5 kb	B18, B72
48.5 kb	B18, U50
82 kb	B72, B134
112 kb	B10, B29, B73, B84, U42, U76
120 kb	B91
140 kb	B40
145 kb	B36

4.3.5 Determination of plasmid types

After the detection of the different plasmids for all the 13 ESBL harbouring ST73 *E.coli* strains from clinical samples of bacteraemia and UTI by S1 PFGE, it was important to determine the plasmid type of these plasmids to complete their profiles. The Plasmid Finder tool (www.cge.cbs.dtu.dk) was used to identify the plasmid replicons or incompatibility groups (Inc) that are specific regions responsible for plasmid replication (Szczepanowski et al., 2005; Carattoli et al., 2014). The Plasmid finder tool also assigns the reference plasmid from the NCBI database used to identify the plasmid under investigation (Carattoli et al., 2014). The sequences of all thirteen ESBL carrying ST73 *E. coli* strains from the clinical samples of bacteraemia and UTI were uploaded to the website and the results are displayed in Table 4.5.

Table 4. 5: Plasmid Typing by Plasmid Finder tool

Strain	Plasmid Type (Inc)[a]	Identity (%)	Position in sequence	Accession Number in NCBI
B10	FIB(AP001918)	96%	5031346-5032023	AP001918
	Col156	98%	5118210-5118363	NC_009781
	FII	96%	5180958-5181206	AY458016
B18	Inconclusive			
B29	Col156	99%	5049007-5049160	NC_009781
	FIB(AP001918)	97%	5225976-5226657	AP001918
	FII	96%	5291977-5292228	AY458016
B36	FII	96%	5072669-5072918	AY458016
	Col156	99	5152675-5152828	NC_009781
	FIB(pB171)	81%	5209650-5210108	AB024946 (pB171)

B40	FIA	100%	5080891-5081278	AP001918
	FIB(AP001918)	97%	5153511-5154129	AP001918
B72	FII(pRSB107)	96%	5045924-5046185	AJ851089
B73	Col156	99%	5076791-5076944	NC_009781
	FII	96%	5102726-5102977	AY458016
	FIB(AP001918)	97%	5112600-5113281	AP001918
B84	FII	96%	5063637-5063888	AY458016
	FIB(AP001918)	97%	5127437-5128118	AP001918
	Col156	99%	5225448-5225601	NC_009781
B91	FII(29)	100%	4883713-4883971	CP003035 (pUTI89)
	FIB(S)	80%	5122047-5122580	FN432031
	Col156	99%	5150722-5150875	AP001918
B134	FIB(AP001918)	98%	4988225-4988906	AP001918
	FII(p14)	86%	5120388-5120627	JQ418538 (p14)
	FIA	100%	5212914-5213276	AP001918
U42	FIB(AP001918)	97%	5019853-5020534	AP001918
	Col8282	90%	5083549-5083746	DQ995353
	Col156	99%	5166911-5167064	NC_009781
	Col(MG828)	92%	5185730-5185991	NC_008486
U50	FII	100%	5002792-5003045	AY458016
U76	Col156	99%	4991211-4991364	NC_009781
	FII	96%	5037803-5038054	AY458016
	FIB(AP001918)	96%	5102247-5102925	AP001918

[a] The numbers between brackets after the Inc type refer to the NCBI accession number of the sequence used to Identify the plasmid type by Plasmid Finder tool (Carattoli et al., 2014)

The first thing we confirmed from Figure 5.4 was that the positions of the plasmid sequences were at the end of our genome assemblies which is where we expected them to be after PAGIT assembly. PAGIT reassembles the genomes according to the reference genome. Since the reference genome of CFT073 has no plasmid in the sequence, PAGIT reorders the unmatched sequences of plasmids at the end of the genomes. Some strains were found to have more than one Inc type in the same plasmid, which occurs in ExPEC (Karisik et al., 2006; Johnson and Nolan, 2010). We notice that the Inc type determined for all of the plasmids in our study belonged to IncFIA, IncFIB or IncFII which is very significant as these IncF types are well documented in the literature as encoding CTX-M-15 among other ESBLs conferring multidrug resistance (MDR) and virulence factors (Karisik et al., 2006; Villa et al., 2010; Carattoli, 2013). One of the most important types of plasmids found in our group was FII which had 96% homology in strains B10, B29, B36, B73, B84, U50, and U76. There was significant identity to pC15-1a which is an IncFII plasmid encoding CTX-M-15 ESBL first sequenced by Boyed et al. (2004) in Canada. It has also been found carrying CTX-M-15 in the United Kingdom (Karisik et al., 2006) and is strongly associated with ST131 producing CTX-M-15 ESBLs from many other countries in Europe (Novais et al., 2012). IncFII plasmids can also be found in the same plasmids with IncFIB regions as was the case in our strains as seen in Table 4.4 (TOBE et al., 1999; Karisik et al., 2006). Another well documented resistance plasmid type found in our strains was the FII(pRSB107) harboured by the strain B72 with 96% homology. It was characterized as containing IncFII, IncFIA and IncFIB replicons, MDR properties and virulence associated functions (Szczepanowski et al., 2005). Other studies in many countries have published the association of pRSB107 plasmids and homologous plasmids with the incidence and spread of CTX-M-15 and other CTX-M ESBLs (Carattoli et al., 2007; Novais et al., 2007; Coque et al., 2008;). Homologous plasmids to FIB(AP001918) plasmid were also

associated with hospital SHV ESBL infections in Poland (Zienkiewicz et al., 2007) and this plasmid was identified with 97% homology in SHV carrier strains B40, B84 and U76 (Table 4.1A and Table 4.1B).

Referring to Tables 4.3 and 4.4, the plasmid types of the thirteen ESBL positive ST73 *E. coli* strains were compared to those detected with S1 PFGE. Interestingly, five *E. coli* ST73 strains were found to have the same plasmid type determined by Plasmid Finder tool with the same size from S1 PFGE. These strains were B10, B84, U76, B73 and B29 which have a 112 kb plasmid with exactly the same plasmid type combination of the significant resistant plasmid FII in addition to FIB(AP001918) and col156. Although the exact sequence of the plasmid is not available for confirmation, with the same size and exact plasmid type (with homology more than 95% by Plasmid Finder tool), this 112 kb plasmid is very likely to be identical in these strains. These five strains were members of the group of seven similar strains mentioned in sections 4.3.1 and 4.3.2 with the exception of B36, which contains an 145 kb plasmid, and U42 which has a 112 kb without an IncFII replicon. These plasmid types and the other resistance plasmids from plasmid Finder were combined with the different plasmid sizes from S1 PFGE in Table 4.5 and then mapped on to the SNP phylogenetic tree in Figure 4.7.

Table 4. 6: Main plasmid types from Plasmid Finder with S1 PFGE profiles

S1 PFGE plasmid size	Strain	Plasmid Type from Plasmid Finder tool
33.5 kb	B18	Not determined
	B72	FII(pRSB107)
48.5 kb	B18	Not determined
	U50	FII
82 kb	B72	FII(pRSB107)
	B134	FII(p14) and FIA
112 kb	B10, B29, B73, B84, U76	FII, FIB(AP001918) and Col156
120 kb	B91	FII(29) and FIB(S)
140 kb	B40	FIA and FIB(AP001918)
145 kb	B36	FII and FIB(pB171)

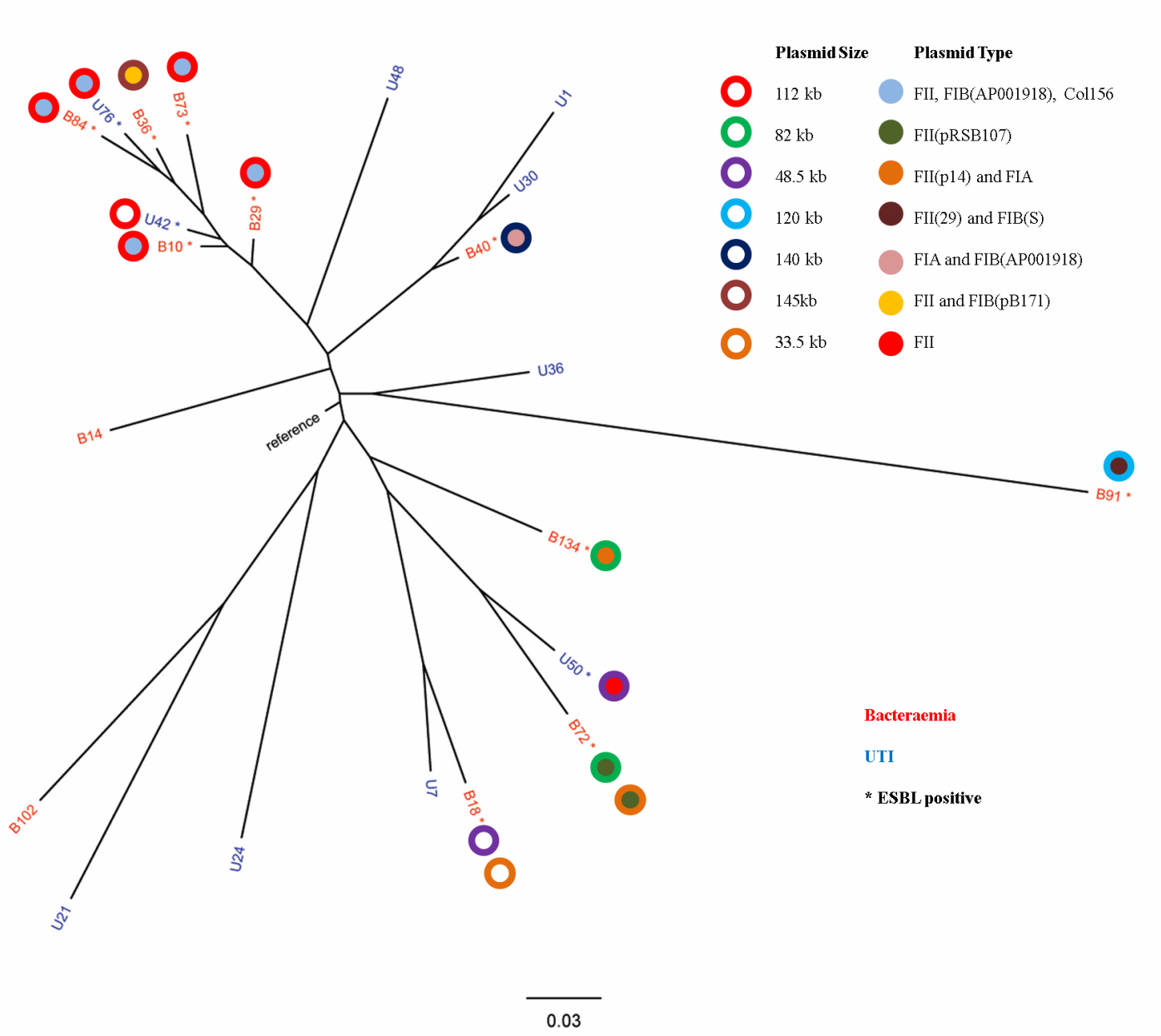


Figure 4. 7: SNP phylogenetic tree with the plasmid profiles of the ESBL carrier strains. The Figure displays the SNPs phylogenetic tree of the sequenced strains in this study with CFT073 strain as a reference. The plasmid profiles of the ESBL carrier strains (*) are displayed where the open circles represent the different plasmid sizes detected by S1 PFGE while the filled circles describe the different replicons identified by the Plasmid Finder web tool. Bacteraemia strains labelled red while UTI strains are blue.

Figure 4.7 shows the ESBL encoding plasmids were diverse among the thirteen ESBL carrier strains, containing 7 different sized plasmids with different backbones consisting of combinations of IncF replicons. Figure 4.7 illustrated more clearly the similarity between five (B10, B29, B73, B84, U76) of the seven strains identified in sections 4.3.1 and 4.3.2 earlier as they carry a plasmid of the same size and IncF replicon, a combination not found anywhere else in the tree. The remaining two strains (B36 and U42) shared a 112 kb plasmid with the rest of the group but with different variants of the incompatibility region consisting of FII/FIA and FIB respectively.

4.3.6 Serum resistance profiles

At this stage of the study no unique loci were identified specific either to the genomes of bacteraemia ST73 *E. coli* or ESBL carrier ST73 *E. coli*. Plasmid profiling also revealed diverse plasmids carried by ST73 *E. coli* of the two clinical groups. Serum resistance was then investigated with the aim of identifying a phenotype specific for the bacteraemia or ESBL positive ST73 *E. coli*. The ability to resist complement mediated killing reflected by the serum resistance properties is an important virulence determinant for ExPEC in UTI and bacteraemia infections (Buckles et al., 2009; Miajlovic et al., 2014). All twenty-two selected strains were tested for their ability to survive the bactericidal effect of normal human serum after three hours incubation at 37ºC. The results are displayed in Figure 4.8. We noticed that all strains were resistant to the killing effect of normal human serum which calls for further investigation. Serum resistance properties of UTI causing ExPEC are well investigated in the literature and provided strong evidence for the role of O Antigens, Capsule, lipopolysaccharide (LPS), outer membrane proteins, enterobacterial common antigen (ECA) and colonic acid (Buckles et al.,

2009; Phan et al., 2013; Miajlovic et al., 2014; Sarkar et al., 2014). A comprehensive study was performed by Phan et al. in 2013 to identify all the genes required to produce resistance to human serum in the *E. coli* ST131 strain EC958. The study utilized transposon directed insertion site sequencing (TraDIS) and fifty-six mutants were found associated with serum resistance with different significance. The main significant genes for serum resistance defined by the study were involved in O antigen biosynthesis (*rml*A), lipid A core biosynthesis and LPS (*waa*A, *rfa*A, *nag* and *pgm*), biosynthesis of ECA (*wec*A), colonic acid biosynthesis (*wca*F and *fbp*) and outer membrane proteins (*tol*A, *tol*Q and *lpp*).

The above genes were investigated for their presence in the genome of our ST73 reference strain CFT073 which is well documented as being a serum resistant prototype isolated from the blood of uroseptic patient (Miajlovic et al., 2014). Although serum resistance mechanisms are not required to be present at the same time in a single bacterium to confer resistance (Cross et al., 1986; Phan et al., 2013), all the genes mentioned above were found in the reference genome CFT073. More importantly, when this group of genes was investigated in our sequenced ST73 strains from the clinical samples of UTI and bacteraemia, all of them were found in the genomes with no exception. Furthermore, another well documented serum resistance mechanism is a major outer membrane protein encoded by *tra*T which is harboured on large IncF plasmids (Moll et al., 1980; Montenegro et al., 1985; Koga et al., 2014). This *tra*T gene was screened as part of the VAGs screening multiplex PCR (section 3.3.5) but was found to be asbent from B14, B102 and U24.

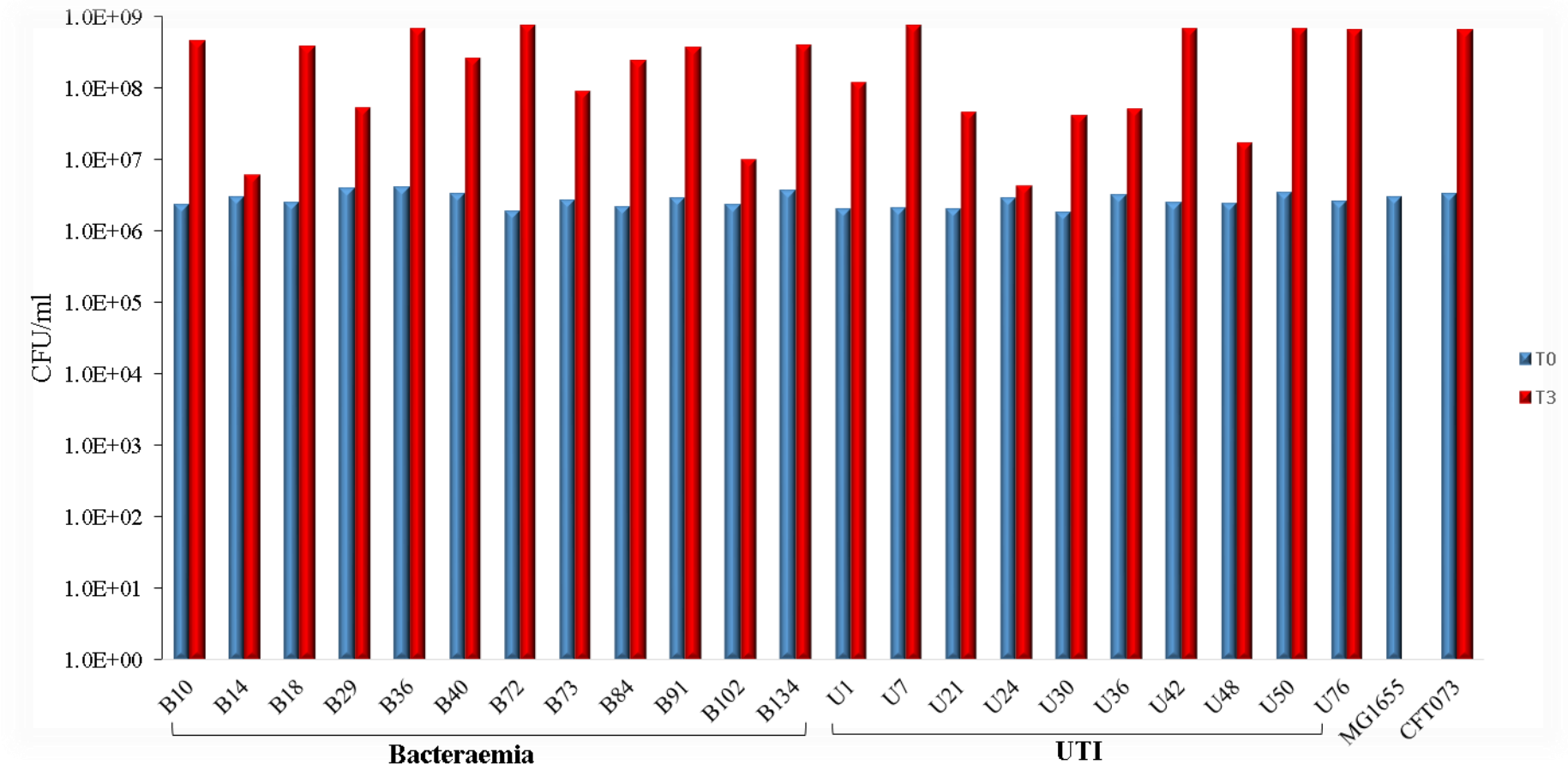


Figure 4. 8: Serum resistance properties of the sequenced ST73 *E. coli* isolates from UTI and Bacteraemia. The figure displays the viable count of the tested isolates at the start of the procedure (blue, T_0) and after three hours incubation (red, T_3). *E.* coli CFT073 was used as a resistant control strain and *E. coli* MG1655 as a sensitive control.

4.4 Discussion

In the previous chapter, one hundred and forty unrelated clinical *E. coli* isolates from bacteraemia and one hundred and twenty five clinical *E. coli* isolates from UTI were collected from Nottingham University hospital in a five month period of 2011. Compared to the UTI population, the population structure of bacteraemia revealed a reduction in diversity with ST131, ST73 and ST95 the three dominant STs. The *E. coli* isolates of these dominant STs were highly associated with ESBL carriage and multiple drug resistance and no significant variation in their VAGs carriage. This has led to the conclusion that ESBL carriage and MDR provided a selective advantage for these strains to emerge as successful dominant bacteraemic isolates (Alhashash et al., 2013). To further investigate this hypothesis, twenty-two randomly selected ST73 *E. coli* strains from clinical samples of bacteraemia and UTI with different ESBL carriage profiles were sequenced. A SNP based phylogenetic tree was obtained and revealed a heterogeneous group of ST73 *E. coli* isolates reflected by a total of 15278 high fidelity SNPs between them. The ST73 *E. coli* strains from bacteraemia and UTI did not cluster and were randomly distributed throughout the tree, which ruled out the presence of a specific outbreak strain. Seven ESBL positive *E. coli* strains, five from the bacteraemia samples (B10, B29, B36, B73 and B84) and two from UTI samples (U42 and U76) grouped together on one side of the tree and were further investigated by measuring the number of SNPs between them. They differed by 416 to 989 SNPs which indicates that they are distantly related ruling out recent transmission or linkage between cases. This finding of our local *E. coli* ST73 isolates is counter to the picture with the globally disseminated clone of *E. coli* ST131. Although collected from disperse geographical locations worldwide and different clinical sources, whole genome SNP phylogenetic analysis of these ST131 *E. coli* isolates reveal little diversity and clustering under a single lineage or sublineage (Price et al., 2013; Petty et al., 2014). Furthermore the SNP analysis between two clades of 85 ST131 *E. coli* isolates from a global data set were

distinguished only by 2,900 SNPs (Petty et al., 2014). In contrast, the analysis of the current study of a local 22 clinical *E. coli* ST73 isolates were distributed around the phylogenetic tree and the high diversity is confirmed by 15,278 SNPs between them. Before the recognition of the globally disseminated MDR *E. coli* ST131 (Price et al., 2013; Petty et al., 2014), our pathogen research group investigated the *E. coli* population in UTI in the region and ST131 *E. coli* isolates were the dominant strains with the highest antimicrobial resistance and CTX-M-15 carriage (Clark et al., 2012). Ten unrelated isolates of these ST131 *E. coli* from the hospital and the community samples were sequenced. SNP based phylogeny of the strains against the ST131 reference strain NA114 displayed clear clustering where only 1324 SNPs were found between them and strain specific SNPs were as low as 10 to 60 SNPs, which confirms the circulation of the pandemic clone of *E. coli* ST131 (Clark et al., 2012). This a clear contrast to the local dominant CTX-M-15 positive *E. coli* ST73 isolates investigated in this project.

For the purpose of our study, comparative genome analysis of all twenty-two sequences of ST73 *E. coli* strains from the clinical samples of bacteraemia and UTI was performed to investigate whether there are bacteraemia specific loci present or absent which define the strains and contribute to their passage and prevalence as a bacteraemic group other than being ESBL carriers. No genomic regions or unique loci were identified specific to the bacteraemia strains. The pangenome analysis showed a very heterogeneous population where the accessory genome represents almost 50% of the pangenome. In contrast, monomorphic genomes analysis can produce a core genome representing up to 80% of the pangenome, a result obtained by analysis of six strains of pathogenic *Streptococcus agalatiae* (Tettelin et al., 2005). No ctx-m genes were identified in the stringent option bacteraemia specific loci analysis but this can be explained since two of the bacteraemia strains were ESBL negative and they may not be from UTI origin associated with ESBL carriage. Causes for *E. coli* bacteraemia other than UTI

origins are documented in the literature such as trauma, burns, respiratory infection, intra-abdominal infections, soft tissue infections, surgical wounds, intra-venous catheters and mechanical ventilators (Russo and Johnson, 2003; Rodriguez-Bano et al., 2008; Wu et al., 2011). Indeed, when the less stringent conditions were used in the analysis ESBL and plasmid regions were identified as associated. The two comparative genome analyses, the high resolution phylogeny and pangenome analysis, clearly confirmed that the increased number of bacteraemic ST73 *E. coli* and the increased ESBL carriage of these bacteraemic ST73 *E. coli* are not a result of a sudden expansion of a successful clone as observed with ST131 (Price et al., 2013).

The importance of serum resistance as a mechanism in establishing disease in UTI and bacteraemia by ST73 *E. coli* was emphasised as all twenty-two sequenced strains were serum resistant. Carriage of the significant serum resistance genes was also confirmed in all the sequences. A parallel study was performed in our research group to investigate urosepsis by *E. coli* (McNally et al., 2013). Isolates of *E. coli* from UTI of five patients were collected and compared to *E. coli* isolates collected from the blood of the same patients when they developed bacteraemia. All the strains from UTI and bacteraemia belonging to the same patient were found to be of the same sequence type except for one patient with three different strains. When serum resistance phenotypes were tested, there was no significant difference in the sensitivity to human serum between isolates from blood and urine of the same patient. Furthermore, the strains belonging to ST73 complex were also found to be serum resistant. Another study by our group found that all the ST131 *E. coli* isolates from UTI and bacteraemia were highly serum resistant (Alqasim et al., 2014). This and the results from the current study support the association of serum resistance with ST73 *E. coli* in general and with other dominant clinical STs.

As there was no association found between specific genomic regions or loci with bacteraemia, our focus was turned to investigate the ESBL carrying plasmids in our sequenced ST73 strains. S1 PFGE revealed the presence of diverse plasmids distributed between our strains in terms of the number of plasmids present and their sizes. When the plasmid types were determined by Plasmid Finder tool, all the plasmids were of IncFIA, IncFIB and IncFII which are well documented as being associated with CTX-M carriage and multiple drug resistance (Karisik et al., 2006; Villa et al., 2010; Carattoli, 2013). In addition to their replicons, these IncF plasmids have F transfer region for conjugation and promote their own transmissibility, which explains their dissemination (Johnson and Nolan, 2009). One particular CTX-M resistant plasmid was identified (FII\FIBAP001918, 112 kb) and matched between five strains, four from bacteraemia (B10, B29, B73 and B84) and one from UTI (U76) with similarity in their SNP phylogeny. This may suggest the circulation of this small group taking in consideration the different collection time between them and clinical source but requires further strain collection and sequencing to confirm. There is a noticeable distribution of plasmid types across opposite ends of the phylogenetic tree (Figure 4.7) where similar plasmids are shared with relatively close strains. A similar conclusion was suggested in Enterotoxigenic *E. coli* (ETEC) (Shepard et al., 2011; von Mentzer et al., 2014). In fact, a very recent study was conducted investigating the plasmid role in the evolution of ST131 *E. coli* sublineages which demonstrated that many lineages acquired a particular plasmid which was then spread in a clonal lineage (Lanza et al., 2014) in a way similar to the evolution of the *H30*-Rx clone (Price et al., 2013).

Our investigation revealed that at the phylogenetic level, ST73 *E. coli* from bacteraemia and UTI are a diverse group not a successful circulating clone like ST131. At the gene content level, there was no bacteraemia specific loci associated with the bacteraemia ST73 *E. coli*, there was no ESBL positive specific loci associated with the ESBL carrying ST73 *E. coli* and the core genome was 50% of the pangenome genome which is different from the ST131 pandemic clone. Furthermore, at the plasmid level, diverse sizes and types of plasmids are distributed among the ESBL positive ST73 *E. coli* . This, in addition to displaying a highly serum resistant phenotype, leads to the conclusion that the increased number of ST73 *E. coli* isolates observed in the bacteraemia population and the increased ESBL carriage of these strains is totally a random event and that any ST73 *E. coli* can acquire a plasmid and become a successful bacteraemic strain.

Chapter 5

Conclusions and Future directions

5.0 Conclusion and future directions

The increase of bacteraemia caused by *E. coli* infection is associated with increased ESBL carriage and multidrug resistance, and has become a major health concern worldwide. Hence, identifying the *E. coli* bacteraemia population structure is of great importance. This PhD project aimed to identify and characterize the population of *E. coli* causing bacteraemia in terms of antibiotic susceptibility, ESBL and virulence associated genes (VAGs) carriage and analysis of the population structure using MLST. To put the *E. coli* bacteraemia population in context, the bacteraemia *E. coli* population was compared to a concurrent *E. coli* population isolated from UTI. In total 140 *E.coli* clinical isolates from bacteraemia and 125 clinical *E. coli* isolates from UTI were collected between March and July 2011 from Nottingham University Hospital (NUH).

In general, an increased level of antibiotic resistance was observed in the bacteraemia *E. coli* population when compared to the UTI *E. coli* isolates. In particular, a significantly higher resistance to ciprofloxacin and cefradine was observed in the bacteraemia *E. coli* population. This requires attention, as resistance to these antibiotics are associated with drug misuse by outpatients and increased prescription (Cooke et al., 2010; Sahuquillo-Arce, et al., 2011). More importantly, multidrug resistant *E. coli* were significantly more common in the bacteraemia *E. coli* population compared with the UTI *E. coli*. This multiple drug resistance, observed in the bacteraemia *E. coli* population, was associated with significantly higher ESBL carriage (CTX-M, OXA and SHV) than the UTI *E. coli* isolates. This is of great concern as ESBL positive bacteraemic *E. coli* are reported to be associated with increased mortality (Cooke, et al., 2010; Sahuquillo-Arce, et al., 2011).

The carriage of virulence associated genes (VAGs) was screened in both the bacteraemia and UTI populations. Both populations displayed high proportion of carriage of *fimH*, *fyuA*, *kpsMTII*, PAI and *traT*, which are well described ExPEC associated virulence factors (VFs) (Bien et al., 2012). A significantly higher number of genes encoding iron acquisition system protein (*iutA*) and P fimbriae adhesin (*papA, papEF, papG* allele II and *papG* allele II,III) were observed in the bacteraemia *E. coli*. This was expected, as aerobactin siderophore (*iutA*) is a bacteraemia associated VF and P fimbriae is essential for upper UTI infections (Ron, 2010; Bien et al., 2012). On the other hand, S fimbriae which is encoded by the *sfaS* gene was present significantly more frequently in the UTI isolates. S fimbriae is associated with upper UTIs and pyelonephritis (Johnson and Stell, 2000; Jacobsen et al., 2008). As it is beneficial for the urine *E. coli* strains to attach to the kidney epithelium and cause disease, we could argue that loss of S Fimbriae by the bacteraemia *E. coli* strains is also beneficial as they are not confined to the kidney and are free to gain access to the blood stream.

The population structure of the bacteraemia and UTI *E. coli* isolates was determined using MLST. The UTI *E. coli* population was found to be genetically diverse, consisting of a total of 63 sequence types (STs) with almost equal distribution of the common STs. The most prevalent STs were ST73 (12.3%), ST131 (7.2%), ST69 (7.2%) and ST95 (4.8%). This indicated a change in the structure of the UTI *E. coli* population compared to the previous work by our group where ST131 complex most prevalent (Croxall et al., 2011). The MLST results revealed that there was less genetic diversity in the bacteraemia population when compared to the UTI population. Within the bacteraemia population there were three dominant STs that represented 48% of the total isolates in the population. ST131 was the most dominant (21.4%) followed by ST73 (17.1%) and ST95 (9.2%). The increased prevalence of ST131 in the bacteraemia population warrants extra attention and molecular testing for identification purposes is advised

before prescribing antibiotics, as these strains are associated with ESBL carriage and multidrug resistance.

To understand the decreased genetic diversity in the bacteraemia population, when compared to the UTI population, comparisons were carried out to identify any variation in the carriage of ESBL and VAGs. This comparison revealed that there was a significant association between ESBL carriage and the bacteraemia STs, with 51.7% of the bacteraemia STs ESBL positive compared to only 30.1% of the UTI STs. The dominant bacteraemia STs (STs 131, 73 and 95) presented with a high proportion ESBL carriage, with a significantly higher ESBL carriage found in the bacteraemia ST73 isolates (50%), when compared to the UTI ST73 isolates (18%). In contrast, there was no significant difference observed for VAG carriage, which were equally distributed between the *E. coli* isolates from clinical bacteraemia and UTI samples. Attempts of assigning a set of VAGs specific to bacteraemia STs, which could explain a reason for their dominance, failed to identify any loci significantly associated with those dominant STs.

Our findings revealed a strong association between multiple drug resistance due to ESBL carriage and the decrease in the genetic diversity of the bacteraemia population. ESBL carriage may provide a selective advantage for the success and progression of the bacteraemia isolates. Further genomic analysis was needed to identify any other factors that could explain the favoured selection for the dominant bacteraemia STs. This genomic analysis would also enable the identification and analysis of the diversity across the genome. Because of the fact that *E. coli* ST131 is a well-defined pandemic circulating clone (Johnson et al., 2013; Price et al., 2013; Petty et al., 2014), we aimed to investigate if this clonlity is observed in other dominant multi drug resistant (MDR) STs. We elected to genome sequence and analyse the poorly defined

ST73 clinical *E. coli* isolates for the purpose of our project. To our knowledge, comparative genome analysis specific to ST73 *E.coli* isolates from two contemporaneous clinical samples are lacking in the literature. Twenty two ST73 *E. coli* isolates from UTI and bacteraemia were selected (10 ESBL positives from bacteraemia, 2 ESBL negatives from bacteraemia, 3 ESBL positives from UTI and 7 ESBL negatives from UTI), which represent the ratio of positive to negative ESBL carriage in the two populations. This genomic analysis was targeted at determining the phylogenetic relationship between the strains and identifying the presence or absence of loci specific for both bacteraemia and ESBL carriage. Plasmid profiling was also performed to identify presence or absence of a successful circulating drug resistance plasmid.

The high resolution phylogenetic tree representing the twenty two ST73 *E. coli* isolates revealed that the isolates were genetically diverse and there was no specific grouping according to the clinical source. This indicates that the increased prevalence of ST73 isolates the bacteraemia and UTI population is not due to the circulation of a successful clone, as was previously the case with ST131 in the same geographical region (Clark et al., 2012). When the SNP distances were determined, the high heterogeneity of the twenty-two genome sequences was confirmed, with a total number of 15,278 SNPs and the closest strains ranging from 416 to 989 SNPs difference. In comparison ST131 from the UTI population observed previously in the region by our group (Clark et al., 2012) had 1324 SNPs between them and strain specific SNPs ranged between 10 and 60 only. It is clear that our high resolution phylogeny is completely different from the globally identified pandemic ST131 subclone *H30*Rx (Price et al., 2013; Petty et al., 2014). Large numbers of ST131 *E. coli* isolates from global geographic locations and different clinical sources are found to cluster phylogenetically in a single lineage or closely related subclades with few SNPs between them (Price et al., 2013; Petty et al., 2014).

In contrast, our ST73 *E. coli* isolates showed high level of diversity presented by heterogeneous phylogeny and 15278 high fidelity SNPs between local 22 strains.

Comparative genome analysis of genomic sequences of all the twenty two ST73 *E. coli* from the bacteraemia and UTI clinical samples were performed using the Gegenees software (Agren et al., 2012). The comparative genomic analysis was carried out with the aim of identifying specific loci associated with the bacteraemia isolates and the ESBL positive isolates. No unique loci or genomic regions were found in the bacteraemia isolates. This indicates that bacteraemia ST73 *E. coli* isolates do not carry specific genomic loci which make them more likely to cause bacteraemic, therefore in theory any UTI ST73 *E. coli* strain can cause bacteraemia. Similarly, no unique loci or genetic regions were found specific to all the ESBL carrying isolates from UTI or bacteraemia. This indicated that the acquisition of ESBL plasmid by ST73 *E. coli* isolates from bacteraemia and UTI is not due to genetic predisposition, but rather by a completely random event. The limitation to these conclusions is that they apply only to ST73 isolates. A larger sample size of ST73 *E. coli* isolates from bacteraemia and UTI would further support our findings. Our investigation would be more informative and more conclusive if we broaden the field and analysed many different STs at the same time, perhaps including the low prevalent STs such as ST10 isolates which were found ESBL negative in the UTI population and ESBL positive in the bacteraemia. This will provide more understanding to what drives different STs to become successful pathogenic STs.

Construction of a pangenome for our ST73 data set confirmed the high level of genomic variation of the ST73 *E. coli* isolates. The pangenome revealed that the accessory genome, which encodes for the strain diversity and selective advantages for virulence factors, antibiotic

resistance and niche adaptation (Tettelin et al., 2008), accounted for about 50% of the pangenome (Figure 4.5). The plasticity and mosaic characteristics of the ST73 reference genome CFT073 *E. coli* is well documented in the litrature (Welch et al., 2002). Nevertheless, these studies produced core genomes of only 50 to 60% of the pangenome of CFT073 *E. coli*; this is a result of the inclusion of other distant pathotypes, such as intestinal pathogenic *E. coli* and commensal *E. coli* isolates (Lloyd et al., 2007; Rasko et al., 2008;). To our knowledge, our project produced for the first time a pangenome for isolates from the same ST of *E. coli* isolated from two clinical sources from the same geographical region and over the same time period and yet the genomic contents were very diverse, which is completely different to the monomorphic genomic characteristics of ST131 (Clark et al., 2012; Price et al., 2013). These results were not successful in identifying genomic traits that can be used to differentiate between the bacteraemia and UTI isolates. To identify a defining characteristic future work may have to investigate variation in gene expression, whilst isolates are exposed to the two clinical environments. This future work could be carried out using expression studies such as RNA-sequencing, which could identify if a difference exists between the bacteraemia and UTI isolates, or between the ESBL plasmid positive and negative isolates.

Plasmid profiles of all the ESBL positive ST73 *E. coli* were determined and diverse plasmids were obtained. This indicated that there was no circulation of a particular resistance plasmid. All the plasmids detected in our study were typed to IncFIA, IncFIB and IncFII, which are associated with multiple drug resistance and CTX-M carriage. (Karisik et al., 2006; Villa et al., 2010; Carattoli, 2013). These types of plasmids can promote their own transmission by their F transfer region through conjugation, which may explain the presence of similar plasmids shared by different strains in our sample. One plasmid was of particular interest which have the size and the same type named (FII\FIBAp001918) by the plasmid finder tool (Carattoli et al., 2014),

was found carried by five closely related strains; 4 from bacteraemia (B10, B29, B73, and B84) and one from UTI (U76), located on one side of the phylogenetic tree (Figure 4.7). This plasmid warrants further characterization and analysis by sequencing. Taking in consideration that this plasmid was shared between isolates from different clinical sources, analysis may reveal the start of a successful circulating group of isolates in a clonal lineage. Sequencing and Analysis of the other ESBL plasmids in this group may identify specific VAGs carried on these plasmids, which may have a role in their increased prevalence in bacteraemia isolates.

This project reported an increase and dominance of ST73 *E. coli* in the urine population compared to previous reports (Gibreel et al., 2011; Clark et al., 2012). This project also revealed that ST73 *E. coli* was a dominant ST in the bacteraemia population. Our study reported that this increase of bacteraemic ST73 *E. coli* is highly associated with increased ESBL carriage and multi drug resistance. The comparative genome analysis showed that there were no unique loci specific to the bacteraemia ST73 *E. coli* and no unique loci specific to ESBL carrying ST73 *E. coli* originating from blood or urine. Furthermore all ST73 *E. coli* appear to have intrinsic ability to be serum resistant. In conclusion, the evidence gathered from this project infers that it is possible for any ST73 *E. coli* from UTI to potentially acquire an ESBL resistance plasmid and become highly pathogenic or bacteraemic, both of which have associations with increased mortality. This project provided the latest comprehensive snapshot of the population structure of clinical *E. coli* isolates from UTI and bacteraemia clinical samples and offered new insights on the adaptation of a common sequence type to become a dominant successful pathogen.

Chapter Six

Appendix

6.0 Appendix

6.1 Full scanned copy of the publication which contained information and data in this PhD thesis:

Title:

Multidrug-Resistant *Escherichia coli* Bacteremia

Authors:

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A.

Journal:

(2013). *Emerg. Infect. Dis*. 19 (10): 1699-1701

Acknowledgments

We thank Tobias Bleicker, Sebastian Brünink, and Monika Eschbach-Bludau for technical assistance; Thomas Seifert, Sonja Matthee, and Conrad Mathee for invaluable help; and Anna-Marie Corman for assistance with geographic information processing.

This study was supported by the European Union FP7 projects EMPERIE (contract number 223498) and ANTIGONE (contract number 278976); the German Research Foundation (DFG grant DR 772/3-1, KA1241/18-1, to C.D.); the National Health Laboratory Service Research Trust (to W.P.); and the South African National Research Foundation (to W.P.).

Ndapewa Laudika Ithete, Samantha Stoffberg, Victor Max Corman, Veronika M. Cottontail, Leigh Rosanne Richards, M. Corrie Schoeman, Christian Drosten, Jan Felix Drexler,[1] and Wolfgang Preiser[1]

Author affiliations: University of Stellenbosch/National Health Laboratory Service, Tygerberg, South Africa (N.L. Ithete, S. Stoffberg, W. Preiser); University of Bonn Medical Centre, Bonn, Germany (V.M. Corman, C. Drosten, J.F. Drexler); University of Ulm, Ulm, Germany (V.M. Cottontail); Durban Natural Science Museum, Durban, South Africa (L.R. Richards); and University of KwaZulu Natal, Durban (M.C. Schoeman)

DOI: http://dx.doi.org/10.3201/eid1910.130946

References

1. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367:1814–20. http://dx.doi.org/10.1056/NEJMoa1211721
2. Reusken CB, Lina PH, Pielaat A, de Vries A, Dam-Deisz C, Adema J, et al. Circulation of group 2 coronaviruses in a bat species common to urban areas in Western Europe. Vector Borne Zoonotic Dis. 2010;10:785–91. http://dx.doi.org/10.1089/vbz.2009.0173
3. Annan A, Baldwin HJ, Corman VM, Klose SM, Owusu M, Nkrumah EE. Human betacoronavirus 2c EMC/2012–related viruses in bats, Ghana and Europe. Emerg Infect Dis. 2013;19:456–9. http://dx.doi.org/10.3201/eid1903.121503
4. Geldenhuys M, Weyer J, Nel LH, Markotter W. Coronaviruses in South African bats. Vector Borne Zoonotic Dis. 2013;13:516–9. http://dx.doi.org/10.1089/vbz.2012.1101
5. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, et al. Genomic characterization of severe acute respiratory syndrome–related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J Virol. 2010;84:11336–49. http://dx.doi.org/10.1128/JVI.00650-10
6. Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, et al. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J Virol. 2007;81:1574–85. http://dx.doi.org/10.1128/JVI.02182-06
7. Falćon A, Vazquez-Moron S, Casas I, Aznar C, Ruiz G, Pozo F, et al. Detection of alpha and betacoronaviruses in multiple Iberian bat species. Arch Virol. 2011;156:1883–90. http://dx.doi.org/10.1007/s00705-011-1057-1
8. Wacharapluesadee S, Sintunawa C, Kaewpom T, Khongnomnan K, Olival KJ, Epstein JH. Group C betacoronavirus in bat guano fertilizer, Thailand. Emerg Infect Dis. 2013; Epub ahead of print. http://dx.doi.org/10.3201/eid1908.130119
9. Anthony SJ, Ojeda-Flores R, Rico-Chavez O, Navarrete-Macias I, Zambrana-Torrelio C, Rostal MK, et al. Coronaviruses in bats from Mexico. J Gen Virol. 2013; 94(Pt 5):1028–38. http://dx.doi.org/10.1099/vir.0.049759-0
10. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. J Virol. 2007;81:2805–16. http://dx.doi.org/10.1128/JVI.02095-06

[1]These authors contributed equally to this article.

Address for correspondence: Jan Felix Drexler, Institute of Virology, University of Bonn Medical Centre, 53127 Bonn, Germany; email: drexler@virology-bonn.de

Multidrug-Resistant *Escherichia coli* Bacteremia

To the Editor: Extraintestinal pathogenic *Escherichai coli* (ExPEC) bacteria have the ability to cause diverse and serious diseases, such as urinary tract infections (UTIs) and bacteremia (*1–3*); incidence of bacteremia is increasing globally (*4*). The emergence of multidrug resistance in *E. coli* is also becoming a global concern, with particular emphasis on *E. coli* sequence type (ST) 131, which is being increasingly reported in UTIs. Drug resistance is mediated by extended-spectrum β-lactamases (ESBLs), mainly of the CTX-M family, particularly CTX-M-15 and 14, and less frequently of the SHV and OXA families (*5,6*). Few studies are available regarding the characterization of *E. coli* strains causing bacteremia.

We characterized 140 *E. coli* isolates from bacteremia patients treated at Nottingham University Hospital (Nottingham, UK) over a 5-month period, with the aim of developing an epidemiologic profile of the population of ExPEC that causes bacteremia. For context, we compared the isolates with 125 *E. coli* isolates from urine samples collected during the same period. Cases were selected to include isolates from a diverse patient group: patient ages ranged from 1 month to 90 years; patient sex was evenly divided between male and female; infections were community- and hospital-associated; and suspected sources of infection varied. Antimicrobial drug susceptibility tests, PCR detection of ESBL genes multilocus sequence typing using the Achtman scheme (http://mlst.ucc.ie/mlst/dbs/Ecoli), and virulence-associated gene (VAG) carriage screening by PCR were performed on isolates as described (*7*).

Significantly more bacteremia *E. coli* isolates than urine *E. coli* isolates were resistant to ciprofloxacin (25.7%

vs. 8.8%; p≤0.001) and cefradine (20.0% vs. 11.2%; p≤0.05). These results were reflected in the number of isolates in the 2 populations displaying a multidrug- resistance phenotype (resistance to antimicrobial drugs belonging to ≥2 classes); a significantly higher number of multidrug-resistant bacteremia *E. coli* isolates than multidrug-resistant urine isolates were found (50.7% vs. 32%; p = 0.01). PCR screening for ESBL carriage showed significantly higher ESBL carriage in bacteremia *E. coli* isolates than urine isolates for bla_{SHV} (15.7% vs. 5.6%; p = 0.008), $bla_{CTX\text{-}M}$ (29.3% vs. 17.6%; p = 0.025), and bla_{OXA} (14.3% vs. 6.4%; p = 0.037). Total ESBL carriage for bacteremia isolates was also significantly higher than for urine isolates (59.3% vs. 29.6%; p≤0.001).

Multilocus sequence types were determined for all *E. coli* isolates. A total of 63 STs were found among the urine isolates (Figure, panel A); the highest prevalence was ST73 (n = 16, 12.8%), followed by ST131 (n = 9, 7.2%), ST69 (n = 9, 7.2%), ST95 (n = 6, 4.8%), ST404 (n = 6, 4.8%), ST127 (n = 4, 3.2%), ST141 (n = 4, 3.2%), and ST10 (n = 3, 2.4%). Prevalence patterns of STs among bacteremia *E. coli* isolates were noticeably different (Figure, panel B). Three main STs were obtained. ST131 dominated (n = 30, 21.43%) and was significantly higher in prevalence than for the urine isolates (p≤0.001). ST73 (n = 24, 17.14%) and ST95 (n = 13, 9.29%) were the other 2 primary STs found. The 8 most prevalent STs in the bacteremia isolates represented 59.29% of the total population, whereas the 8 most prevalent STs in the urine isolates represented 45.6% of the total population. This finding is suggestive of selection of a smaller number of dominant STs in bacteremia.

ESBL carriage was mapped onto minimum-spanning trees for the 2 isolate groups. ESBL carriage among urine isolates was focused on a small number of STs; 19 (30.16%) of the 63 STs contained ESBL-positive isolates (Figure, panel A). The predominant ST73 group contained 18.75% ESBL-positive isolates; the other predominant STs exhibited ESBL-positive isolates at the following levels: ST131 (44.44%), ST69 (33.33%), ST95 (50%), and ST10 (0%). In contrast, 30 (51.72%) of the 58 STs among bacteremia isolates contained ESBL-positive isolates, significantly higher than for the urine isolates (p = 0.016). At the ST level, predominant STs had higher ESBL carriage in the bacteremia isolates than in the urine isolates: ST131 (50%), ST73 (50%), ST12 (75%), ST10 (100%), ST14 (50%), ST2278 (33.33%). ST95 (46.15%) and ST69 (20%) showed comparable levels. These results suggest that ESBL drug resistance is selecting for dominant ExPEC bacteremia strains.

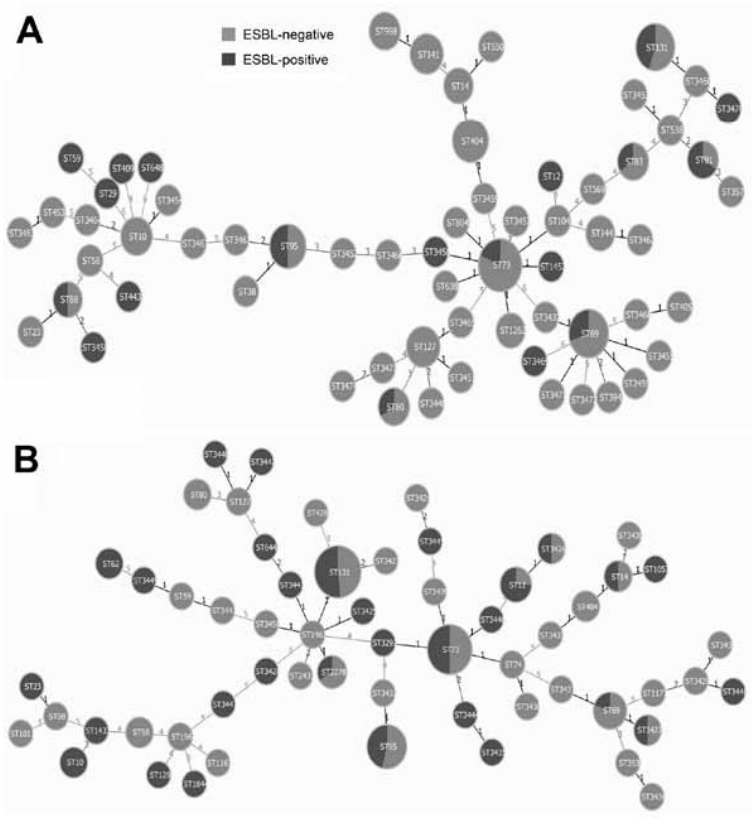


Figure. Minimum-spanning trees showing carriage of extended-spectrum β-lactamases (ESBL) in *Escherichia coli* isolates from urine samples (A) and samples from patients with bacteremia (B). Each circle represents 1 sequence type (ST), and the size of the circle reflects the number of isolates belonging to this particular ST within the bacteria group. Lines between the circles represent how different their allelic profiles are; a line labeled 1 means the linked STs differ in ≥1 of the 7 alleles, which is named a single locus variant (SLV). A cluster of STs linked by SLVs is a clonal complex. Nineteen (30.16%) of 63 STs found among the urine isolates were ESBL positive, in comparison to 30 (51.72%) of 58 for the bacteremia isolates.

To investigate whether the differences in ST observations between bacteremia and urine isolates could be attributable to differences in virulence genes, VAGs of all isolates were screened by multiplex PCR. VAGs were found equally distributed across the 2 populations, with no statistically significant difference ($p = 0.675$). Comparison of serum resistance levels between urine and blood isolates also showed no phenotypic differences.

In conclusion, we found high levels of ESBL carriage and multidrug resistance in ExPEC isolates that cause bacteremia. A comparison with urine isolates provided evidence that ESBL-mediated drug resistance appears to be the selective pressure in the emergence of dominant STs in bacteremia. Future research should focus on identifying if prolonged antimicrobial drug treatment in bacteremia patients is leading to this selection.

Acknowledgments

We thank staff in the clinical microbiology laboratory at Nottingham University Hospitals for their assistance in isolate collection and Nicholas Gleadall and Kuan Min Chen for technical assistance in performing multilocus sequence typing PCRs.

This work was supported by a Kuwait Government and Kuwait Civil Service Commission scholarship awarded to F.A. and an East Midlands Development Agency iNET grant awarded to A.M. and M.D.

Fahad Alhashash, Vivienne Weston, Mathew Diggle, and Alan McNally

Author affiliations: Nottingham Trent University, Nottingham, United Kingdom (F. Alhashash, A. McNally); and Nottingham University Hospitals National Health Service Trust, Nottingham (V. Weston, M. Diggle)

DOI: http://dx.doi.org/10.3201/eid1910.130309

References

1. Wiles T, Kulesus R, Mulvey M. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. Exp Mol Pathol. 2008;85:11–9. http://dx.doi.org/10.1016/j.yexmp.2008.03.007
2. Ron EZ. Distribution and evolution of virulence factors in septicemic *Escherichia coli*. Int J Med Microbiol. 2010;300:367–70. http://dx.doi.org/10.1016/j.ijmm.2010.04.009
3. Russo TA, Johnson J. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect. 2003;5:449–56. http://dx.doi.org/10.1016/S1286-4579(03)00049-2
4. de Kraker ME, Jarlier V, Monen JC, Heuer OE, van de Sande N, Grundmann H. The changing epidemiology of bacteremias in Europe: trends from the European Antimicrobial Resistance Surveillance System. Clin Microbiol Infect. 2012; [Epub ahead of print]. http://dx.doi.org/10.1111/1469-0691.12028
5. Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V, Demarty R, Alonso MP, Canica MM, et al. Intercontinental emergence of *Escherichia coli* clone O25:H4–ST131 producing CTX-M-15. J Antimicrob Chemother. 2008;61:273–81. http://dx.doi.org/10.1093/jac/dkm464
6. Lau SH, Kaufmann ME, Livermore DM, Woodford N, Willshaw GA, Cheasty T, et al. UK epidemic *Escherichia coli* strains A–E, with CTX-M-15 β-lactamase, all belong to the international O25:H4–ST131 clone. J Antimicrob Chemother. 2008;62:1241–4. http://dx.doi.org/10.1093/jac/dkn380
7. Croxall G, Hale J, Weston V, Manning G, Cheetham P, Achtman M, et al. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. J Antimicrob Chemother. 2011;66:2501–8. http://dx.doi.org/10.1093/jac/dkr349

Address for correspondence: Alan McNally, Pathogen Research Group, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK; email: alan.mcnally@ntu.ac.uk

Transmission of Schmallenberg Virus during Winter, Germany

To the Editor: Schmallenberg virus (SBV), an orthobunyavirus, emerged in northern Europe in 2011 (*1*). SBV infection causes transient fever, diarrhea, and a reduced milk yield in adult ruminants but, most notably, stillbirths and severe malformations in lambs and calves (*2*). Insect vectors play an essential role in transmission; the viral genome has been detected in various field-collected biting midges (*Culicoides* spp.) (*3*,*4*).

During autumn 2012 and winter 2012–2013, blood samples were taken at several times from individual sheep on a farm located in the German federal state of Mecklenburg–Western Pomerania. The farm is surrounded by agricultural fields and meadows. Approximately 1,000 ewes and their lambs, a dog, and some cats were kept on the farm; most of the animals are outdoors year-round. Only dams with ≥2 lambs are housed in open stabling in December and January. The dung is regularly cleared away and stored ≈10 m from 1 of the stable entrances. Repellents or insecticides were not applied in the monitored period. Blood samples were taken in September 2012 and in January and February 2013 and analyzed by an SBV-specific real-time quantitative reverse transcription PCR (RT-qPCR) (*5*) and by an SBV antibody ELISA (ID Screen Schmallenberg virus Indirect; IDvet; Montpellier, France) by using the recommended cutoff of 50% relative optical density as compared with the positive control (sample-to-positive ratio [S/P]).

In September 2012, blood samples from 60 sheep tested negative by the SBV antibody ELISA. Moreover, fetal malformations of the brain, spinal cord, or skeletal muscle, which might have suggested a previous SBV-infection of the dam, were not

6.1 Supplementary information for Chapter Four – (Comparative genome analysis)

6.2 Sequences assembly by Velvet software

All the raw sequencing data of the selected twenty two ST73 *E.coli* isolates were received from Exeter Sequencing Service in a FASTQ file format. FASTQ files are common text files which include the sequencing reads of the nucleotides and their associated quality scores (Cock, et al., 2010). Velvet software was used to assemble the sequences (Ebi.ac.uk/~zerbino/velvet, 2014). Illumina sequencing produces short reeds and the genome assembly requires piecing together these random reads into contiguous sequences or contigs (Zerbino and Birney, 2008). The size of our FASTQ files were large of about 1.9 Gigabytes (gb) in average as it included a lot of information about the nucleotide in the sequence. Scripts were used to shorten and shuffle the forward and reverse FASTQ files of each isolate into a Velvet recognizable version. Velvet is a de Bruijn graph assembler used for the production of *de novo* assembled genomes, which means it assembles the genomes without the use of a reference genome of a closely related strain (Swain, et al., 2012). Velvet works in two parts. First, the reads were converted into k nucleotides words or k-mers in a statistical table containing all the nucleotides present in the sequence with their quality information (Zerbino and Birney, 2008). This was obtained by running velvet h script of the software with k-mer size of 31. Second, the overlapping k-mers and the de Bruijn graph produced by velvet h were pieced together into long sequences or contigs to make the assembled genome. This was obtained by running velvet g script of the software with expected coverage the bps of 75 fold and 8 reads cut-off parameters. The output file of Velvet software is a FASTA format which is simply a text file containing the nucleotides sequence of the genome without any annotations. Assembled genome sequences were obtained for all the twenty two ST73 *E. coli* isolates and were used for further analysis.

De novo Velvet assembly was performed using a script involving these command lines:

Head -n 20000000 sampleR1.fastq > sampleshort.fastq

Head -n 20000000 sampleR2.fastq > sampleshort2.fastq

Perl shuffleSequences_fastq.pl sampleshort.fastq sampleshort2.fastq sampleinput.fastq

Velveth: sample assem 31 -shortPaired -fastq sampleinput.fastq

Velvetg: /home/alan/Desktop/sample assem -exp_cov 75 -cov_cutoff 8

6.3 Post assembly genome improvement toolkit (PAGIT)

PAGIT software produces a high quality sequence from *de novo* assembled genomes by utilizing a number of sub programmes namely ABACAS, IMAGE, iCORN and RATT in accordance of a closely related reference strain (Swain, et al., 2012). In the PAGIT assembly CFT073 *E. coli* genome sequence (NCBI Reference Sequence: NC_004431.1, genome size is 5231428 bps) was used as a reference genome because it belonged to sequence type ST73. The forward and reverse FASTQ files in addition to the FASTA file obtained from Velvet were also included. The unguided contigs produced from Velvet were reordered by ABACAS programme according to the reference genome and any sequences which are not mapped with reference such as plasmid sequences are reordered at the end. IMAGE programme reduced the gaps in the assembled genome by mapping the FASTQ files against the contigs (Swain, et al., 2012). Two important parameters of the quality of an assembled genome is a reduced total number of contigs and an increased N50 (which is the median contig length) up to 50 kb or more (Zerbino and Birney, 2008). IMAGE programme has managed to close the gaps and reduce the number of contigs in the original velvet assembly genomes from thousands to hundreds, and has increased the N50 from thousands to tens of thousands. Details of a much improved assembled genomes of all the twenty two ST73 *E. coli* isolates after PAGIT assembly are described in Table 4.2. In addition, errors in the assembled genome such as insertions, deletions or substitutions were corrected by iCORN programme through mapping against original sequencing reads. Finally, RATT transferred the annotations from the reference genome CFT073 to the final assembled genome (Swain, et al., 2012). The high quality genome assembly by PAGIT was performed on all the selected twenty two ST73 *E.coli* isolates. The final output file of PAGIT assembly was an EMBL file which is an annotation file format similar to GENBANK (GBK) file. A final FASTA file for each of the sequenced isolates was obtained from their final EMBL file through Artemis software (Carver, et al., 2012).

The PAGIT assembly script was written by Dr. Alan McNally. Firstly, (velv.fa), (strainfor.fq), (strainrev.fq), (reference.fasta) and (reference.embl) files were moved into PAGIT (or PAGIT.V1.64bit) folder. Now you open PAGIT folder and double click on dorun.sh file and choose display, and on top you write the exact strain name, the velv.fa name, the reference.fasta name and the reference.embl name and then save. Open the terminal type cd and run PAGIT from the terminal using these command lines:

cd Desktop

cd PAGIT (or PAGIT.V1.64bit)

source sourceme.pagit

./dorun.sh

6.4 Annotation of sequenced

Prokka software is a rapid more accurate unguided annotation tool compared to transfer annotation softwares which produce errors as they force the transfer of annotations (Seemann, 2014). Only FASTA files of the assembled genomes are required to start the Prokka script where it utilizes a built in database to annotate the genome and produces output files in GFF and GENBANK (GBK) formats which are compatible annotation formats ready for further analysis (Seemann, 2014). All the selected twenty two ST73 *E. coli* isolates genomes where annotated by Prokka software.

Prokka annotation was performed using the command line:

./prokka --outdir sample ID _prokka --genus Escherichia --locustag sample ID sample ID.fasta

6.5 Single nucleotide polymorphism (SNP) based phylogeny

6.5.1 Alignment of sequenced genomes

All the twenty two sequenced ST73 *E. coli* isolates were aligned against *E. coli* CFT073 as a reference genome using SMALT software which is developed and optimized by Wellcome Trust Sanger Institute (Sanger.ac.uk, 2014). SMALT provides a more accurate alignment of paired end sequencing reads produced by Illumina, Roche 454 or ABI Sanger compared to other alignment programmes such as BWA and BOWTIE (Wellcome Trust Sanger Institute, 2014a). SMALT alignment process involves first forming an index table of the reference genome consisting specific length words or hashing with equidistant spaces, then sequence reads are mapped against the hash index (Wellcome Trust Sanger Institute, 2014b). The input files used for SMALT software were the FASTA file of the reference genome *E. coli* CFT073 and the two FASTQ files of sequences for each strain. The words length used in the index was seventeen with equidistant steps of two. Each of the twenty two ST73 strains were mapped by SMALT software individually and the output file for each was obtained in a SAM file format. SAM files created by SMALT were large in size containing all the statistical information of the alignment including the position of each single read against the reference, the quality score of the bases and the quality score of the reads.

SAMtools software was used to extract the required information from SAM files generated by SMALT in the previous step. SAMtools is a software used to manipulate the data in SAM format files such as indexing, merging, sorting and other post processing requirements (Li et al., 2009). SAMtools scripts were applied on the SAM files to produce an index from the reference genome, generate BAM files which are readable by other programmes and sorting the BAM files information of the position of the reads according to the reference genome index with their specific read scores. Since we aim to obtain high mapping with high fidelity SNPs,

SAMtools scripts were also used to remove duplicate sequences generated by the adapters during multiplex sequencing. A last and very important SAMtool script was used to pileup the reads against the reference genome to produce a BCF file format containing the base variants of the reads along with the number of bases and the quality scores in these variants or SNPs. This last SAMtools scripts also included a command to pipe the output BCF file into bcfTools programme with all the SNPs and their quality scores information. BCF files can be recognized, converted to VCF files or indexed by bcfTools software (Evolution and Genomics, 2011). VCF files were produced from BCF files by scripts of bcfTools software and piped into perl vcfutils to utilize a varfilter script with maximum likelihood of one hundered which removed SNPs reads which were very low. These VCF files contained all the SNPs data for each strain against the reference genome *E. coli* CFT073 and they were filtered to obtain high fidelity SNPs using scripts performed by Alan McNally at NTU. The filtering criteria for obtaining high fidelity SNPs based phylogenetic tree was removing multiple alleles, removing reads with of quality scores of less than thirty, removing SNPs reads which were less than eight and most importantly the allele frequency was determined to be 75%.

```
/Desktop/smalt-0.5.8$ ls

./smalt_x86_64 index -k 17 -s 2 CFT073index CFT073.fasta

./smalt_x86_64 map -f sam -o B14.sam CFT073index B14_R1.fastq B14_R2.fastq

samtools faidx CFT073.fasta

samtools view -bS -t CFT073.fasta.fai -o B14.bam B14.sam

samtools sort B14.bam B14.sort

samtools rmdup B14.sort.bam B14.out

samtools mpileup -ugf CFT073.fasta B14.out | bcftools view -bvcg - > B14.raw.bcf
```

bcftools view B14.raw.bcf | perl Vcfutils.pl varFilter -D 100 > B14.vcf

bgzip B14.vcf

tabix -p vcf B14.vcf.gz

6.5.2 Construction of SNPs based phylogenetic tree

VCF files containing SNPs information of the all the twenty two sequenced ST73 *E. coli* strains obtained in the previous step were filtered as described above using scripts applied by Alan McNally at NTU and converted to FASTA file formats. Another FASTA format file of all the SNPs was generated from VCF files with removing off the common SNPs which were present in all the strains regardless of the reference genome in order to improve measurements of SNPs distances. The individual SNPs FASTA files were concatenated into one SNPs FASTA file and was viewed in SeaView software. SeaView programme is a sequence alignment editor and viewer of many alignment file formats such as NEXUS, CLUSTAL, MASE, MSF, FASTA and PHYLIP (Gouy et al., 2010). The SNPs FASTA files were opened in SeaView software to view the SNPs and saved as PHYLIP file format which the only format recognizable by the phylogenetic tree generating software called RAxML-HPC or randomized axelerated maximum likelihood for high performance computing (Stamatakis, 2006). The RAxML-HPC script was implied to produce a 100 phylogenetic trees with rapid bootstrapping which is a statistical method used to place confidence in repeated sampling or in this case the 100 phylogenetic trees (Felsenstein, 1985). RAxML-HPC script used also applied (GTR) or general time reversible model and gamma distribution which is a nucleotide substitution model for DNA evolution in which all the bases in the bacterial genome are considered unbiased and can mutate to produce sense and non-sense SNPs unlike the eukaryotic genome (Rodriguez et al., 1990). The last step in RAxML-HPC script was set to imply the maximum likelihood statistics

which estimates or infer the maximum likelihood of the generated phylogenetic trees to be true given the different evolutionary parameters (Felsenstein, 1981). The output file obtained from RAxML-HPC was best tree in RAxML format and it was used to view the SNPs based phylogenetic tree using Figtree software (tree.bio.ed.ac.uk/software, 2014). The phylogenetic tree was viewed and saved into the phylogenetic file format called NWK file.

RaxML maximum likelihood phylogeny was performed using the command line:

raxmlHPC -f a -m GTRGAMMA -x 12345 -# 100 -s Alan_ Escherichia _align.phy -n Alan _ Escherichia _raxml

6.5.3 Measurement of pairwise distance

The SNPs FASTA files generated in section 4.2.7.3 were used to compute the number of base differences between the ST73 *E coli* strains using MEGA software (Tamura et al., 2013). Pairwise distance measurement was selected from the Distance menu and the number of SNPs were determined. A table was generated and exported as spreadsheet file for further analysis.

6.6 Comparative Genome analysis

Whole genome comparisons were performed on all the twenty two ST73 *E. coli* from the clinical samples of bacteraemia and UTI using the software Gegenees (Gegenees.org, 2014). Gegenees software allows the comparative analysis on up to hundreds of complete or draft genomes by aligning fragments of all the genomes utilizing multithreaded blast approach. The alignment data can provide a phylogenetic view of the genomes, but more importantly for our study, it can be used to investigate the presence of genomic regions unique in a target group when compared to a background group (Agren et al., 2012). Gegenees software was also used to determine the core and pangenome.

6.6.1 Investigation of bacteraemia and ESBL carrier specific loci

Gegenees software was used to investigate the presence or absence of genomic regions or loci specific only to bacteraemia or to ESBL carrier strains of the sequenced ST73 *E. coli*. This was performed by utilizing the Fragmented all-all comparison option (Gegenees.org, 2014). First, all the twenty two GENBANK (GBK) files were imported in a new database in Gegenees software, FASTA file formats can also be used. Then, in the (Fragmented all-all comparison) widow, nucleotide comparison (BLASTN) was chosen with the highest resolution parameters for the alignment (200/100) which were 200 bps of sliding fragment window size and 100bps of progressive step size respectively. When the alignments and fragments comparison were performed and displayed, the bacteraemia strains were selected as target group against a background group of all the UTI strains in order to investigate for the presence of bacteraemia specific genomic regions or loci. The strain B134 was chosen as reference genome from the target group to refer the unique loci to with annotations. In the other test, all the ESBL carrier strains were selected as a target group from bacteraemia and UTI clinical samples against a

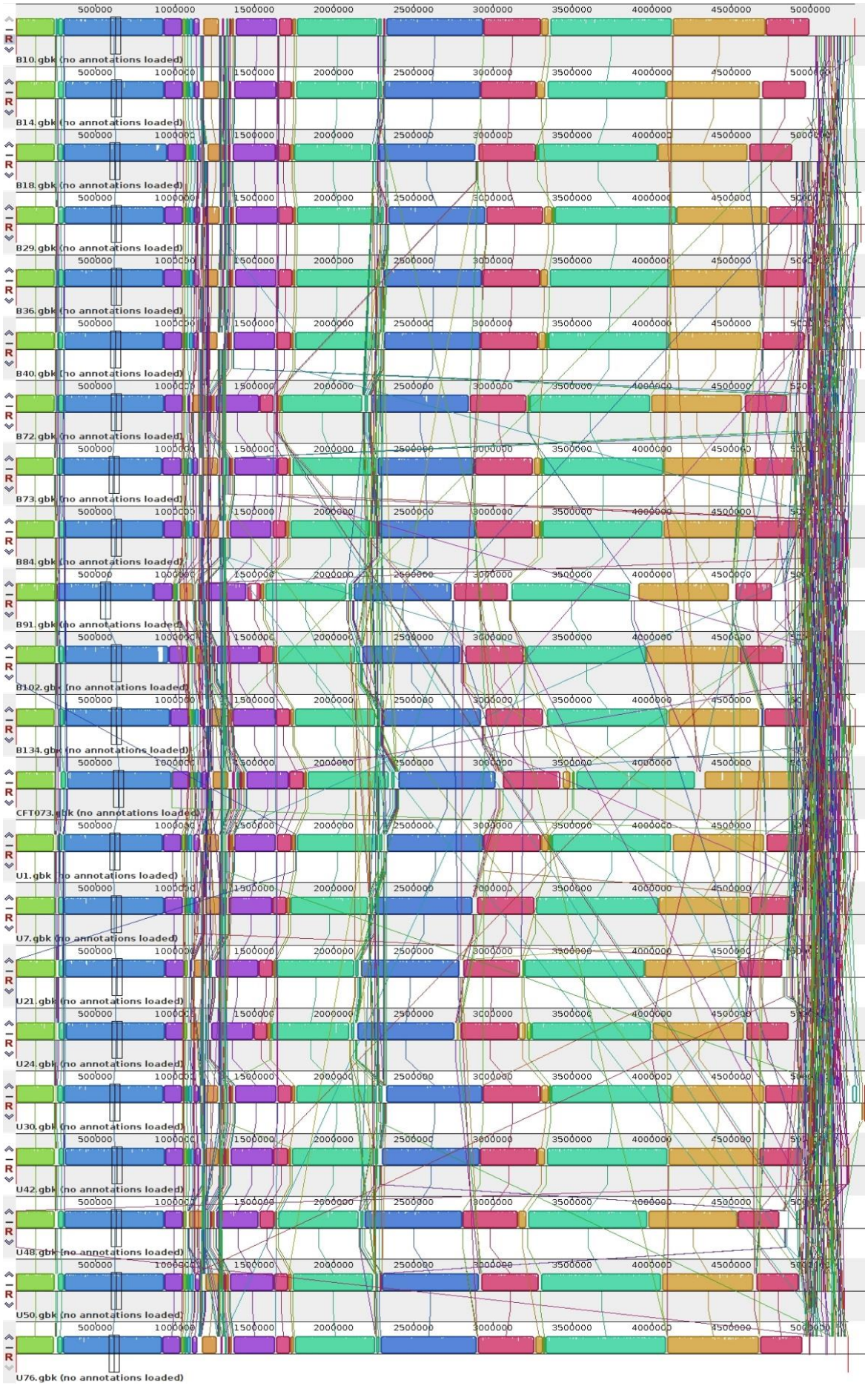
background group of ESBL absent strains in order to investigate ESBL carrier specific genomic regions or loci with B134 strain as a reference from the target group.

6.6.2 Determination of core and pangenome

An overview of the core and pangenome of all the twenty two ST73 *E. coli* strains was created by Gegenees software. All the genomes were fragmented then compared to each other and their BLASTN scores were determined. The pangenome was constructed progressively by analysing the fragments BLASTN scores and then each new unique fragment was deposited in the pangenome (Gegenees.org, 2014). In the (New pangenome) option, 1% threshold was selected for maximum distinctive selection of unique fragments.

6.7 Progressive Mauve alignment

Progressive Mauve software is capable to successfully align multiple sequences of closely related strains even if they have suffered genome rearrangements, loss or gain by recombination (Darling et al., 2010). In the Figure below, the twenty two sequenced genomes of ST73 *E. coli* strains from clinical samples of bacteraemia and UTI are shown. Blocks of the same colour connected by a line are homologous while the difference in their sizes and sometimes their loss displays the variation between the strains. More noticeably are the large numbers of homologous lines at the end of the genomes which represent the plasmids sequences region with multiple crosses between them indicating high variations and rearrangement events in their sequences.


500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
4500000
5000000
B10.gbk (no annotations loaded)
B14.gbk (no annotations loaded)
B18.gbk (no annotations loaded)
B29.gbk (no annotations loaded)
B36.gbk (no annotations loaded)
B40.gbk (no annotations loaded)
B72.gbk (no annotations loaded)
B73.gbk (no annotations loaded)
B84.gbk (no annotations loaded)
B91.gbk (no annotations loaded)
B102.gbk (no annotations loaded)
B134.gbk (no annotations loaded)
CFT073.gbk (no annotations loaded)
U1.gbk (no annotations loaded)
U7.gbk (no annotations loaded)
U21.gbk (no annotations loaded)
U24.gbk (no annotations loaded)
U30.gbk (no annotations loaded)
U42.gbk (no annotations loaded)
U48.gbk (no annotations loaded)
U50.gbk (no annotations loaded)
U76.gbk (no annotations loaded)

References

References

ACHTMAN, M. (2012) Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1590): 860-867.

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. *Antimicrobial Agents and Chemotherapy* 57(1): 490-497.

AGREN, J., SUNDSTRÖM, A., HÅFSTRÖM, T. AND SEGERMAN, B. (2012) Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLoS ONE* 7(6): e39107.

AJIBOYE, R., SOLBERG, O., LEE, B., RAPHAEL, E., DEBROY, C. AND RILEY, L. (2009) Global Spread of Mobile Antimicrobial Drug Resistance Determinants in Human and Animal *Escherichia coli* and *Salmonella* Strains Causing Community-Acquired Infections. *Clinical Infectious Diseases* 49(3): 365-371.

AL-HASAN, M. N., ECKEL-PASSOW, J. E. AND BADDOUR, L. M. (2010) Bacteremia complicating gram-negative urinary tract infections: A population-based study. *Journal of Infection* 60(4): 278-285.

ALHASHASH, F., WESTON, V., DIGGLE, M. AND MCNALLY, A. (2013) Multidrug-Resistant *Escherichia coli* Bacteremia. *Emerg. Infect. Dis.* 19(10).

ALQASIM, A., SCHEUTZ, F., ZONG, Z. AND MCNALLY, A. (2014) Comparative genome analysis identifies few traits unique to the *Escherichia coli* ST131 H30Rx clade and extensive mosaicism at the capsule locus. *BMC Genomics* 15(1): 830.

ANDERSON, G., DODSON, K., HOOTON, T. and HULTGREN, S. (2004) Intracellular bacterial communities of uropathogenic in urinary tract pathogenesis. *Trends in Microbiology* 12(9): 424-430.

ANDERSON, G. G., PALERMO, J., SCHILLING, J., ROTH, R., HEUSER, J. AND HULTGREN, J. (2003) Intracellular Bacterial Biofilm-Like Pods in Urinary Tract Infections. *Science* 301(5629): 105-107.

ANDREWS, J. M. AND HOWE, R. A. (2011) BSAC standardized disc susceptibility testing method (version 10). *Journal of Antimicrobial Chemotherapy* 66(12): 2726-2757.

ARMSTRONG, G. L., HOLLINGSWORTH, J. AND MORRIS, J. G (1996) Emerging Foodborne Pathogens: *Escherichia coli* O157:H7 as a Model of Entry of a New Pathogen into the Food Supply of the Developed World. *Epidemiologic Reviews* 18(1): 29-51.

ATEBA, C. N. AND MBEWE, M. (2013) Determination of the genetic similarities of fingerprints from *Escherichia coli* O157:H7 isolated from different sources in the North West Province, South Africa using ISR, BOXAIR and REP-PCR analysis. *Microbiological Research* 168(7): 438-446.

BACKHED, F., SODERHALL, M., EKMAN, P., NORMARK, S. AND RICHTER-DAHLFORS, A (2001) Induction of innate immune responses by *Escherichia coli* and purified lipopolysaccharide correlate with organ- and cell-specific expression of Toll-like receptors within the human urinary tract. *Cell Microbiol* 3(3): 153-158.

BARTON, B., HARDING, G. AND ZUCCARELLI, A. (1995) A General Method for Detecting and Sizing Large Plasmids. *Analytical Biochemistry* 226(2): 235-240.

BASTA, T., KECK, A., KLEIN, J. AND STOLZ, A. (2004) Detection and Characterization of Conjugative Degradative Plasmids in Xenobiotic-Degrading *Sphingomonas* Strains. *Journal of Bacteriology* 186(12): 3862-3872.

BAUDRY, P., NICHOL, K., DECORBY, M., LAGACÉ-WIENS, P., OLIVIER, E., BOYD, D., MULVEY, M., HOBAN, D. AND ZHANEL, G. (2009). Mechanisms of resistance and mobility among multidrug-resistant CTX-M-producing *Escherichia coli* from Canadian intensive care units: the 1st report of QepA in North America. *Diagnostic Microbiology and Infectious Disease* 63(3): 319-326.

BERGERON, C., PRUSSING, C., BOERLIN, P., DAIGNAULT, D., DUTIL, L., REID-SMITH, R., ZHANEL, G. AND MANGES, A. (2012). Chicken as Reservoir for Extraintestinal Pathogenic *Escherichia coli* in Humans, Canada. *Emerg. Infect. Dis.* 18(3): 415-421.

BERGTHORSSON, U. AND OCHMAN, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Molecular Biology and Evolution* 15(1): 6-16.

BERT, F., JOHNSON, J., OUATTARA, B., LEFLON-GUIBOUT, V., JOHNSTON, B., MARCON, E., VALLA, D., MOREAU, R. AND NICOLAS-CHANOINE, M. (2010). Genetic Diversity and Virulence Profiles of *Escherichia coli* Isolates Causing Spontaneous Bacterial Peritonitis and Bacteremia in Patients with Cirrhosis. *Journal of Clinical Microbiology* 48(8): 2709-2714.

BETTELHEIM, K. AND LENNOX-KING, S. (1976). The acquisition of *Escherichia coli* by new-born babies. *Infection* 4(3): 174-179.

BHAKD,I S. (1989) Potent leukocidal action of *Escherichia coli* hemolysin mediated by permeabilization of target cell membranes. *Journal of Experimental Medicine* 169(3): 737-754.

BIELASZEWSKA, M., MELLMANN, A., ZHANG, W., KÖCK, R., FRUTH, A., BAUWENS, A., PETERS, G. AND KARCH, H. (2011). Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infectious Diseases.* 11(9): 671-676.

BIEN, J., SOKOLOVA, O. AND BOZKO, P. (2012). Role of Uropathogenic *Escherichia coli* Virulence Factors in Development of Urinary Tract Infection and Kidney Damage. *International Journal of Nephrology* 2012: 1-15.

BLANCO, J., BLANCO, M., MORA, A., JANSEN, W., GARCÍA, V., VÁZQUEZ, AND BLANCO, J. (1998). Serotypes of *Escherichia coli* isolated from septicaemic chickens in Galicia (Northwest Spain). *Veterinary Microbiology* 61(3): 229-235.

BLANCO, J., MORA, A., MAMANI, R., LOPEZ, C., BLANCO, M., DAHBI, G., HERRERA, A., BLANCO, J., ALONSO, M., GARCIA-GARROTE, F., CHAVES, F., ORELLANA, M., MARTINEZ-MARTINEZ, L., CALVO, J., PRATS, G., LARROSA, M., GONZALEZ-LOPEZ, J., LOPEZ-CERERO, L., RODRIGUEZ-BANO, J. AND PASCUAL, A. (2011). National survey of *Escherichia coli* causing extraintestinal infections reveals the spread of drug-resistant clonal groups O25b:H4-B2-ST131, O15:H1-D-ST393 and CGA-D-ST69 with high virulence gene content in Spain. *Journal of Antimicrobial Chemotherapy* 66(9): 2011-2021.

BOUGUÉNEC, C. (2005). Adhesins and invasins of pathogenic *Escherichia coli*. *International Journal of Medical Microbiology* 295(6-7): 471-478.

BOUZA, E., JUAN, R., MUÑOZ, P., VOSS, A. AND KLUYTMANS, J. (2001). A European perspective on nosocomial urinary tract infections I. Report on the microbiology workload, etiology and antimicrobial susceptibility (ESGNI-003 study). *Clinical Microbiology and Infection* 7(10): 523-531.

BOYD. D,, TYLER. S,, CHRISTIANSON. S,, MCGEER, A., MULLER, M., WILLEY, B., BRYCE, E., GARDAM, M., NORDMANN, P. AND MULVEY, M. (2004). Complete Nucleotide Sequence of a 92-Kilobase Plasmid Harboring the CTX-M-15 Extended-Spectrum Beta-Lactamase Involved in an Outbreak in Long-Term-Care Facilities in Toronto, Canada. *Antimicrobial Agents and Chemotherapy* 48(10): 3758-3764.

BRISSE, S., DIANCOURT, L,. LAOUENAN, C., VIGAN, M., CARO, V., ARLET, G., DRIEUX, L., LEFLON-GUIBOUT, V., MENTRE, F., JARLIER, V. AND NICOLAS-CHANOINE, M. (2012). Phylogenetic Distribution of CTX-M- and Non-Extended-Spectrum-β -Lactamase-Producing *Escherichia coli* Isolates: Group B2 Isolates, Except Clone ST131, Rarely Produce CTX-M Enzymes. *Journal of Clinical Microbiology* 50(9): 2974-2981.

BROLUND, A., FRANZÉN, O., MELEFORS, Ö., TEGMARK-WISELL, K. AND SANDEGREN, L. (2013). Plasmidome-Analysis of ESBL-Producing *Escherichia coli* Using Conventional Typing and High-Throughput Sequencing. *PLoS ONE* 8(6): e65793.

BRUMFITT, W., GARGAN, R. AND HAMILTON-MILLER, J. (1987). Periurethral enterobacterial carriage preceding urinary infection. *The Lancet* 329(8537): 824-826.

BRZUSZKIEWICZ E, THURMER A, SCHULDES J, LEIMBACH A, LIESEGANG H, MEYER, F., BOELTER, J., PETERSEN, H., GOTTSCHALK, G. AND DANIEL, R. (2011). Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch Microbiol.* 193(12): 883 - 891.

BUCHHOLZ., U., BERNARD, H., WERBER, D., BÖHMER, M. M., REMSCHMIDT, C., WILKING, H., DELERÉ, Y., AN DER HEIDEN, M., ADLHOCH, C., DREESMAN, J., EHLERS, J., ETHELBERG, S., FABER, M., FRANK, C., FRICKE, G., GREINER, M., HÖHLE, M., IVARSSON, S., JARK, U., KIRCHNER, M., KOCH, J., KRAUSE, G., LUBER,

P., ROSNER, B., STARK, K. AND KÜHNE, M. (2011) German Outbreak of *Escherichia coli* O104:H4 Associated with Sprouts. *New England Journal of Medicine* 365(19): 1763-1770.

BUCKLES, E., WANG, X., LANE, M., LOCKATELL, C., JOHNSON, D., RASKO, D., MOBLEY, H. AND DONNENBERG, M. (2009) Role of the K2 Capsule in *Escherichia coli* Urinary Tract Infection and Serum Resistance. *The Journal of Infectious Diseases* 199(11): 1689-1697.

BUCKLES, E., WANG, X., LANE, M., LOCKATELL, C., JOHNSON, D., RASKO, D., MOBLEY, H. AND DONNENBERG, M. (2009) Role of the K2 Capsule in *Escherichia coli* Urinary Tract Infection and Serum Resistance. *The Journal of Infectious Diseases* 199(11): 1689-1697.

BUSH, K., JACOBY, G. A. AND MEDEIROS, A. A. (1995) A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrobial Agents and Chemotherapy* 39(6): 1211-1233.

CAGNACCI, S., GUALCO, L., DEBBIA, E., SCHITO, G. C. AND MARCHESE, A. (2008) European Emergence of Ciprofloxacin-Resistant *Escherichia coli* Clonal Groups O25:H4-ST 131 and O15:K52:H1 Causing Community-Acquired Uncomplicated Cystitis. *Journal of Clinical Microbiology* 46(8): 2605-2612.

CANTÓN, R. AND COQUE, T. M. (2006) The CTX-M β-lactamase pandemic. *Current Opinion in Microbiology* 9(5): 466-475.

CARATTOLI, A. (2013) Plasmids and the spread of resistance. *International Journal of Medical Microbiology* 303(6-7): 298-304.

CARATTOLI, A., GARCIA-FERNANDEZ, A., VARESI, P., FORTINI, D., GERARDI, S., PENNI, A., MANCINI, C. AND GIORDANO, A. (2007) Molecular Epidemiology of

Escherichia coli Producing Extended-Spectrum β-Lactamases Isolated in Rome, Italy. *Journal of Clinical Microbiology* 46(1): 103-108.

CARATTOLI, A., ZANKARI, E., GARCIA-FERNANDEZ, A., VOLDBY LARSEN, M., LUND, O., VILLA, L., MOLLER AARESTRUP, F. AND HASMAN, H. (2014) In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 58(7): 3895-3903.

CARVER, T., HARRIS, S. R., BERRIMAN, M., PARKHILL, J. AND MCQUILLAN, J. A. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4): 464-469.

CHAPPELL, L. (2012) Finding a needle in a haystack. *Nat Rev Micro* 10(7): 446-446.

CHEN, M., TOFIGHI, R., BAO, W., ASPEVALL, O., JAHNUKAINEN, T., GUSTAFSSON, L. E., CECCATELLI, S. AND CELSI, G. (2006) Carbon monoxide prevents apoptosis induced by uropathogenic *Escherichia coli* toxins. *Pediatric Nephrology* 21(3): 382-389.

CLARK, G., PASZKIEWICZ, K., HALE, J., WESTON, V., CONSTANTINIDOU, C., PENN, C., ACHTMAN, M. AND MCNALLY, A. (2012) Genomic analysis uncovers a phenotypically diverse but genetically homogeneous *Escherichia coli* ST131 clone circulating in unrelated urinary tract infections. *Journal of Antimicrobial Chemotherapy* 67(4): 868-877.

CLERMONT, O., BONACORSI, S. AND BINGEN, E. (2000) Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Applied and Environmental Microbiology* 66(10): 4555-4558.

CLERMONT, O., DHANJI, H., UPTON, M., GIBREEL, T., FOX, A., BOYD, D., MULVEY, M. R., NORDMANN, P., RUPPE, E., SARTHOU, J. L., FRANK, T., VIMONT, S., ARLET, G., BRANGER, C., WOODFORD, N. AND DENAMUR, E. (2009) Rapid detection of the

O25b-ST131 clone of *Escherichia coli* encompassing the CTX-M-15-producing strains. *Journal of Antimicrobial Chemotherapy* 64(2): 274-277.

COCK, P. J. A., FIELDS, C. J., GOTO, N., HEUER, M. L. AND RICE, P. M. (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research* 38(6): 1767-1771.

COLOMER-LLUCH, M., MORA, A., LOPEZ, C., MAMANI, R., DAHBI, G., MARZOA, J., HERRERA, A., VISO, S., BLANCO, J. E., BLANCO, M., ALONSO, M. P., JOFRE, J., MUNIESA, M. AND BLANCO, J. (2013) Detection of quinolone-resistant *Escherichia coli* isolates belonging to clonal groups O25b:H4-B2-ST131 and O25b:H4-D-ST69 in raw sewage and river water in Barcelona, Spain. *Journal of Antimicrobial Chemotherapy* 68(4): 758-765.

CONNELL, I., AGACE, W., KLEMM, P., SCHEMBRI, M., MARILD, S. AND SVANBORG, C. (1996) Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proceedings of the National Academy of Sciences* 93(18): 9827-9832.

COOKE, N. M., SMITH, S. G., KELLEHER, M. AND ROGERS, T. R. (2010) Major Differences Exist in Frequencies of Virulence Factors and Multidrug Resistance between Community and Nosocomial *Escherichia coli* Bloodstream Isolates. *Journal of Clinical Microbiology* 48(4): 1099-1104.

COQUE, T. M., NOVAIS, Â., CARATTOLI, A., POIREL, L., PITOUT, J., PEIXE, L., BAQUERO, F., CANTÓN, R. AND NORDMANN, P. (2008) Dissemination of Clonally Related *Escherichia coli* Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15. *Emerg. Infect. Dis.* 14(2): 195-200.

CORTES, P., BLANC, V., MORA, A., DAHBI, G., BLANCO, J. E., BLANCO, M., LOPEZ, C., ANDREU, A., NAVARRO, F., ALONSO, M. P., BOU, G., BLANCO, J. AND LLAGOSTERA, M. (2010) Isolation and Characterization of Potentially Pathogenic

Antimicrobial-Resistant *Escherichia coli* Strains from Chicken and Pig Farms in Spain. *Applied and Environmental Microbiology* 76(9): 2799-2805.

CROSS, A. S., KIM, K. S., WRIGHT, D. C., SADOFF, J. C. AND GEMSKI, P. (1986) Role of Lipopolysaccharide and Capsule in the Serum Resistance of Bacteremic Strains of *Escherichia coli*. *Journal of Infectious Diseases* 154(3): 497-503.

CROXALL, G., HALE, J., WESTON, V., MANNING, G., CHEETHAM, P., ACHTMAN, M. AND MCNALLY, A. (2011) Molecular epidemiology of extraintestinal pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. *Journal of Antimicrobial Chemotherapy* 66(11): 2501-2508.

CROXEN, M. A., LAW, R. J., SCHOLZ, R., KEENEY, K. M., WLODARSKA, M. AND FINLAY, B. B. (2013) Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. *Clinical Microbiology Reviews* 26(4): 822â€“880.

CUNNINGHAM, P. N., WANG, Y., GUO, R., HE, G. AND QUIGG, R. J. (2004) Role of Toll-Like Receptor 4 in Endotoxin-Induced Acute Renal Failure. *The Journal of Immunology* 172(4): 2629-2635.

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 extraintestinal pathogenic *Escherichia coli* producing extended-spectrum Î²-lactamases. *International Journal of Antimicrobial Agents* 42(4): 347-351.

DARLING, A. E., MAU, B. AND PERNA, N. T. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* 5(6): e11147.

DHAKAL, B. AND MULVEY, M. (2012) The UPEC Pore-Forming Toxin α-Hemolysin Triggers Proteolysis of Host Proteins to Disrupt Cell Adhesion, Inflammatory, and Survival Pathways. *Cell Host & Microbe* 11(1): 58-69.

DIAS, R. C., MARANGONI, D. V., SMITH, S. P., ALVES, E. M., PELLEGRINO, F. L., 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. *Microbial Drug Resistance* 15(4): 303-308.

EBI.AC.UK, 2014. Velvet: a sequence assembler for very short reads [Online]. Available at: https://www.ebi.ac.uk/~zerbino/velvet/ [Accessed: 25 September 2014].

ESS-WIKI.EXETER.AC.UK, 2014. Exeter Sequencing Service - Exeter Sequencing Service [Online]. Available at: http://ess-wiki.exeter.ac.uk/ess/Exeter_Sequencing_Service [Accessed: 25

ETO, D. S., JONES, T. A., SUNDSBAK, J. L. AND MULVEY, M. A. (2007) Integrin-Mediated Host Cell Invasion by Type 1-Piliated Uropathogenic *Escherichia coli*. *PLoS Pathogens* 3(7): e100.

EVOLUTION AND GENOMICS, 2011. bcfTools [Online]. Available at: http://evomics.org/resources/software/variation/bcftools/ [Accessed: 2 October 2014].

FAM, N., LEFLON-GUIBOUT, V., FOUAD, S., ABOUL-FADL, L., MARCON, E., DESOUKY, D., EL-DEFRAWY, I., ABOU-AITTA, A., KLENA, J. AND NICOLAS-CHANOINE, M. (2011) CTX-M-15-Producing *Escherichia coli* Clinical Isolates in Cairo

(Egypt), Including Isolates of Clonal Complex ST10 and Clones ST131, ST73, and ST405 in Both Community and Hospital Settings. *Microbial Drug Resistance* 17(1): 67-73.

FANG, H., ATAKER, F., HEDIN, G. AND DORNBUSCH, K. (2008) Molecular Epidemiology of Extended-Spectrum β -Lactamases among *Escherichia coli* Isolates Collected in a Swedish Hospital and Its Associated Health Care Facilities from 2001 to 2006. *Journal of Clinical Microbiology* 46(2): 707-712.

FEIL, E. J., LI, B. C., AANENSEN, D. M., HANAGE, W. P. AND SPRATT, B. G. (2004) eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology* 186(5): 1518-1530.

FELSENSTEIN, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6): 368-376.

FELSENSTEIN, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39(4): 783.

FOXMAN, B. (2003) Epidemiology of urinary tract infections: Incidence, morbidity, and economic costs. *Disease-a-Month* 49(2): 53-70.

FRANCISCO, A. P., VAZ, C., MONTEIRO, P. T., MELO-CRISTINO, J., RAMIREZ, M. AND CARRIÇO, J. A. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 13(1): 87.

FRANK, C., WERBER, D., CRAMER, J. P., ASKAR, M., FABER, M., AN DER HEIDEN, M., BERNARD, H., FRUTH, A., PRAGER, R., SPODE, A., WADL, M., ZOUFALY, A., JORDAN, S., KEMPER, M. J., FOLLIN, P., MÜLLER, L., KING, L. A., ROSNER, B., BUCHHOLZ, U., STARK, K. AND KRAUSE, G. (2011) Epidemic Profile of Shiga-Toxin-

Producing *Escherichia coli* O104:H4 Outbreak in Germany. *New England Journal of Medicine* 365(19): 1771-1780.

GARCIA, E. C., BRUMBAUGH, A. R. and MOBLEY, H. L. T. (2011) Redundancy and Specificity of *Escherichia coli* Iron Acquisition Systems during Urinary Tract Infection. *Infection and Immunity* 79(3): 1225-1235.

GEGENEES.ORG, 2014. Gegenees [Online]. Available at: http://www.gegenees.org/ [Accessed: 16 October 2014].

GIBREEL, T. M., DODGSON, A. R., CHEESBROUGH, J., FOX, A. J., BOLTON, F. J. AND UPTON, M. (2011) Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *Journal of Antimicrobial Chemotherapy* 67(2): 346-356.

GLODE, M. P., SUTTON, A., ROBBINS, J. B., MCCRACKEN, G. H., GOTSCHLICH, E. C., KAIJSER, B. AND HANSON, L. A. (1977) Neonatal Meningitis Due to *Escherichia coli* K1. *Journal of Infectious Diseases* 136(Supplement 1): S93-S97.

GOLUSZKO, P., MOSELEY, S. L., TRUONG, L. D., KAUL, A., WILLIFORD, J. R., SELVARANGAN, R., NOWICKI, S. AND NOWICKI, B. (1997) Development of experimental model of chronic pyelonephritis with Escherichia coli O75:K5:H-bearing Dr fimbriae: mutation in the dra region prevented tubulointerstitial nephritis. *Journal of Clinical Investigation* 99(7): 1662-1672.

GORDIENKO, E. N., KAZANOV, M. D. AND GELFAND, M. S. (2013) Evolution of Pan-Genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*. *Journal of Bacteriology* 195(12): 2786-2792.

GOUY, M., GUINDON, S. AND GASCUEL, O. (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* 27(2): 221-224.

GUTACKER, M., MATHEMA, B., SOINI, H., SHASHKINA, E., KREISWIRTH, B., GRAVISS, E. AND MUSSER, J. (2006) Single-Nucleotide Polymorphism-Based Population Genetic Analysis of *Mycobacterium tuberculosis* Strains from 4 Geographic Sites. *The Journal of Infectious Diseases* 193(1): 121-128.

GUYER, D. M., RADULOVIC, S., JONES, F. AND MOBLEY, H. L. T. (2002) Sat, the Secreted Autotransporter Toxin of Uropathogenic *Escherichia coli*, Is a Vacuolating Cytotoxin for Bladder and Kidney Epithelial Cells. *Infection and Immunity* 70(8): 4539-4546.

HACKER, J. AND KAPER, J. B. (2000) Pathogenicity Islands and the Evolution of Microbes. *Annu. Rev. Microbiol.* 54(1): 641-679.

HANCOCK, V., FERRIERES, L. and KLEMM, P. (2008) The ferric yersiniabactin uptake receptor *FyuA* is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology*, 154(1): 167-175.

HARRIS, S. R., CARTWRIGHT, E. J., TÖRÖK, M. E., HOLDEN, M. T., BROWN, N. M., OGILVY-STUART, A. L., ELLINGTON, M. J., QUAIL, M. A., BENTLEY, S. D., PARKHILL, J. AND PEACOCK, S. J. (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet Infectious Diseases* 13(2): 130-136.

HEDLUND, M., WACHTLER, C., JOHANSSON, E., HANG, L., SOMERVILLE, J. E., DARVEAU, R. P. AND SVANBORG, C. (1999) P fimbriae-dependent, lipopolysaccharide-independent activation of epithelial cytokine responses. *Molecular Microbiology* 33(4): 693-703.

HEALTH PROTECTION AGENCY. Polymicrobial bacteraemia and fungaemia in England, Wales, and Northern Ireland: 2010.

http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317132423063 (26 October 2012, date last accessed).

HENDERSON, J. P., CROWLEY, J. R., PINKNER, J. S., WALKER, J. N., TSUKAYAMA, P., STAMM, W. E., HOOTON, T. M. AND HULTGREN, S. J. (2009) Quantitative Metabolomics Reveals an Epigenetic Blueprint for Iron Acquisition in Uropathogenic *Escherichia coli*. *PLoS Pathogens* 5(2): e1000305.

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). *Journal of Antimicrobial Chemotherapy* 69(1): 91-100.

HOUDOUIN, V., BONACORSI, S., BIDET, P., BINGEN-BIDOIS, M., BARRAUD, D. AND BINGEN, E. (2006) Phylogenetic background and carriage of pathogenicity island-like domains in relation to antibiotic resistance profiles among *Escherichia coli* urosepsis isolates. *Journal of Antimicrobial Chemotherapy* 58(4): 748-751.

JACOBSEN, S. M., STICKLER, D. J., MOBLEY, H. L. T. AND SHIRTLIFF, M. E. (2008) Complicated Catheter-Associated Urinary Tract Infections Due to *Escherichia coli* and *Proteus mirabilis*. *Clinical Microbiology Reviews* 21(1): 26-59.

JACOBY, G. A. AND MEDEIROS, A. A. (1991) More extended-spectrum beta-lactamases. *Antimicrobial Agents and Chemotherapy* 35(9): 1697-1704.

JAKOBSEN, L., GARNEAU, P., BRUANT, G., HAREL, J., OLSEN, S. S., PORSBO, L. J., HAMMERUM, A. M. AND FRIMODT-MØLLER, N. (2011) Is *Escherichia coli* urinary tract

infection a zoonosis? Proof of direct link with production animals and meat. *Eur J Clin Microbiol Infect Dis* 31(6): 1121-1129.

JAKOBSEN, L., HAMMERUM, A. M. AND FRIMODT-MOLLER, N. (2010a) Detection of Clonal Group A *Escherichia coli* Isolates from Broiler Chickens, Broiler Chicken Meat, Community-Dwelling Humans, and Urinary Tract Infection (UTI) Patients and Their Virulence in a Mouse UTI Model. *Applied and Environmental Microbiology* 76(24): 8281-8284.

JAKOBSEN, L., SPANGHOLM, D. J., PEDERSEN, K., JENSEN, L. B., EMBORG, H., AGERSO, Y., AARESTRUP, F. M., HAMMERUM, A. M. AND FRIMODT-MOLLER, N. (2010b) Broiler chickens, broiler chicken meat, pigs and pork as sources of ExPEC related virulence genes and resistance in *Escherichia coli* isolates from community-dwelling humans and UTI patients. *International Journal of Food Microbiology* 142(1-2): 264-272.

JAKOBSSON, B., BERG, U. AND SVENSSON, L. (1994) Renal scarring after acute pyelonephritis. *Archives of Disease in Childhood* 70(2): 111-115.

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(1): 560.

JENSEN, U. S., KNUDSEN, J. D., ØSTERGAARD, C., GRADEL, K. O., FRIMODT-MØLLER, N. AND SCHØNHEYDER, H. C. (2010) Recurrent bacteraemia: A 10-year regional population-based study of clinical and microbiological risk factors. *Journal of Infection* 60(3): 191-199.

JOHNSON, J. R. AND CLABOTS, C. (2006) Sharing of Virulent *Escherichia coli* Clones among Household Members of a Woman with Acute Cystitis. *Clinical Infectious Diseases* 43(10): e101-e108.

JOHNSON, J. R. AND O'BRYAN, T. T. (2000) Improved Repetitive-Element PCR Fingerprinting for Resolving Pathogenic and Nonpathogenic Phylogenetic Groups within *Escherichia coli*. *Clinical and Vaccine Immunology* 7(2): 265-273.

JOHNSON, J. R. AND RUSSO, T. A. (2002) Extraintestinal pathogenic *Escherichia coli* : "The other bad E coli". *Journal of Laboratory and Clinical Medicine* 139(3): 155-162.

JOHNSON, J. R. AND RUSSO, T. A. (2005) Molecular epidemiology of extraintestinal pathogenic (uropathogenic) *Escherichia coli*. *International Journal of Medical Microbiology* 295(6-7): 383-404.

JOHNSON, J. AND STELL, A. (2000) Extended Virulence Genotypes of *Escherichia coli* Strains from Patients with Urosepsis in Relation to Phylogeny and Host Compromise. *The Journal of Infectious Diseases* 181(1): 261-272.

JOHNSON, J., JOHNSTON, B., CLABOTS, C., KUSKOWSKI, M. AND CASTANHEIRA, M. (2010) *Escherichia coli* Sequence Type ST131 as the Major Cause of Serious Multidrug-Resistant *E.coli* Infections in the United States. *Clinical Infectious Diseases* 51(3): 286-294.

JOHNSON, J. R., JOHNSTON, B., CLABOTS, C. R., KUSKOWSKI, M. A., ROBERTS, E. AND DEBROY, C. (2008a) Virulence Genotypes and Phylogenetic Background of *Escherichia coli* Serogroup O6 Isolates from Humans, Dogs, and Cats. *Journal of Clinical Microbiology* 46(2): 417-422.

JOHNSON, J. R., MANGES, A. R., O'BRYAN, T. T. AND RILEY, L. W. (2002) A disseminated multidrug-resistant clonal group of uropathogenic *Escherichia coli* in pyelonephritis. *The Lancet* 359(9325): 2249-2251.

JOHNSON, J. R., MENARD, M., JOHNSTON, B., KUSKOWSKI, M. A., NICHOL, K. AND ZHANEL, G. G. (2009) Epidemic Clonal Groups of *Escherichia coli* as a Cause of Antimicrobial-Resistant Urinary Tract Infections in Canada, 2002 to 2004. *Antimicrobial Agents and Chemotherapy* 53(7): 2733-2739.

JOHNSON, J. R., MURRAY, A. C., KUSKOWSKI, M. A., SCHUBERT, S., PRÈRE, M., PICARD, B., COLODNER, R. AND RAZ, R. (2005) Distribution and Characteristics of *Escherichia coli* Clonal Group A1. *Emerg. Infect. Dis.* 11(1): 141-145.

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., DEBROY, C., ROBICSEK, A., HANSEN, G., URBAN, C., PLATELL, J., TROTT, D. J., ZHANEL, G., WEISSMAN, S. J., COOKSON, B. T., FANG, F. C., LIMAYE, A. P., SCHOLES, D., CHATTOPADHYAY, S., HOOPER, D. C. AND SOKURENKO, E. V. (2013) Abrupt Emergence of a Single Dominant Multidrug-Resistant Strain of *Escherichia coli*. *Journal of Infectious Diseases* 207(6): 919-928.

JOHNSON, T. J. AND NOLAN, L. K. (2009) Pathogenomics of the Virulence Plasmids of *Escherichia coli*. *Microbiology and Molecular Biology Reviews* 74(3): 477-478.

JOHNSON, T. J., KARIYAWASAM, S., WANNEMUEHLER, Y., MANGIAMELE, P., JOHNSON, S. J., DOETKOTT, C., SKYBERG, J. A., LYNNE, A. M., JOHNSON, J. R. AND NOLAN, L. K. (2007) The Genome Sequence of Avian Pathogenic *Escherichia coli* Strain

O1:K1:H7 Shares Strong Similarities with Human Extraintestinal Pathogenic *E.coli* Genomes. *Journal of Bacteriology* 189(8): 3228-3236.

JOHNSON, T. J., WANNEMUEHLER, Y., JOHNSON, S. J., STELL, A. L., DOETKOTT, C., JOHNSON, J. R., KIM, K. S., SPANJAARD, L. AND NOLAN, L. K. (2008b) Comparison of Extraintestinal Pathogenic *Escherichia coli* Strains from Human and Avian Sources Reveals a Mixed Subset Representing Potential Zoonotic Pathogens. *Applied and Environmental Microbiology* 74(22): 7043-7050.

KANG, C., SONG, J., CHUNG, D. R., PECK, K. R., KO, K. S., YEOM, J., KI, H. K., SON, J. S., LEE, S. S., KIM, Y., JUNG, S., KIM, S., CHANG, H., RYU, S. Y., KWON, K. T., LEE, H., MOON, C. AND SHIN, S. Y. (2010) Risk factors and treatment outcomes of community-onset bacteraemia caused by extended-spectrum β-lactamase-producing *Escherichia coli*. *International Journal of Antimicrobial Agents* 36(3): 284-287.

KAPER, J. B., NATARO, J. P. AND MOBLEY, H. L. T. (2004) Pathogenic *Escherichia coli*. *Nature Review Microbiology.* 2(2): 123-40.

KARIM, A., POIREL, L., NAGARAJAN, S. AND NORDMANN, P. (2001) Plasmid-mediated extended-spectrum Î²-lactamase (CTX-M-3 like) from India and gene association with insertion sequence ISEcp1. *FEMS Microbiology Letters* 201(2): 237-241.

KARISIK, E., ELLINGTON, M., PIKE, R., WARREN, R., LIVERMORE, D. AND WOODFORD, N. (2006) Molecular characterization of plasmids encoding CTX-M-15 β -lactamases from *Escherichia coli* strains in the United Kingdom. *Journal of Antimicrobial Chemotherapy* 58(3): 665-668.

KASHEF, N., DJAVID, G. E. AND SHAHBAZI, S. (2010) Antimicrobial susceptibility patterns of community-acquired uropathogens in Tehran, Iran. *J Infect Dev Ctries* 4(04).

KAU, A. L., HUNSTAD, D. A. AND HULTGREN, S. J. (2005) Interaction of uropathogenic *Escherichia coli* with host uroepithelium. *Current Opinion in Microbiology* 8(1): 54-59.

KAYA, O., AKCAM, F. Z., GONEN, I., UNAL, O. AND CEYLAN, T. (2013) Risk factors for bacteremia due to extended-spectrum beta-lactamase-producing *Escherichia coli* in a Turkish hospital. *J Infect Dev Ctries* 7(07).

KEYES, K., HUDSON, C., MAURER, J. J., THAYER, S., WHITE, D. G. AND LEE, M. D. (2000) Detection of Florfenicol Resistance Genes in *Escherichia coli* Isolated from Sick Chickens. *Antimicrobial Agents and Chemotherapy* 44(2): 421-424.

KLIEBE, C., NIES, B. A., MEYER, J. F., TOLXDORFF-NEUTZLING, R. M. AND WIEDEMANN, B. (1985) Evolution of plasmid-coded resistance to broad-spectrum cephalosporins. *Antimicrobial Agents and Chemotherapy* 28(2): 302-307.

KNOTHE, H., SHAH, P., KRCMERY, V., ANTAL, M. AND MITSUHASHI, S. (1983) Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. *Infection* 11(6): 315-317.

KOGA, V. L., TOMAZETTO, G., CYOIA, P. S., NEVES, M. S., VIDOTTO, M. C., NAKAZATO, G. AND KOBAYASHI, R. K. T. (2014) Molecular Screening of Virulence Genes in Extraintestinal Pathogenic *Escherichia coli* Isolated from Human Blood Culture in Brazil. *BioMed Research International* 2014: 1-9.

KÖHLER, C. AND DOBRINDT, U. (2011) What defines extraintestinal pathogenic *Escherichia coli*?. *International Journal of Medical Microbiology* 301(8): 642-647.

KOTLOFF, K. L., NATARO, J. P., BLACKWELDER, W. C., NASRIN, D., FARAG, T. H., PANCHALINGAM, S., WU, Y., SOW, S. O., SUR, D., BREIMAN, R. F., FARUQUE, A. S., ZAIDI, A. K., SAHA, D., ALONSO, P. L., TAMBOURA, B., SANOGO, D.,

ONWUCHEKWA, U., MANNA, B., RAMAMURTHY, T., KANUNGO, S., OCHIENG, J. B., OMORE, R., OUNDO, J. O., HOSSAIN, A., DAS, S. K., AHMED, S., QURESHI, S., QUADRI, F., ADEGBOLA, R. A., ANTONIO, M., HOSSAIN, M. J., AKINSOLA, A., MANDOMANDO, I., NHAMPOSSA, T., ACÁCIO, S., BISWAS, K., O'REILLY, C. E., MINTZ, E. D., BERKELEY, L. Y., MUHSEN, K., SOMMERFELT, H., ROBINS-BROWNE, R. M. AND LEVINE, M. M. (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet* 382(9888): 209-222.

LAESTADIUS, Å., RICHTER-DAHLFORS, A. AND APERIA, A. (2002) Dual effects of *Escherichia coli* α-hemolysin on rat renal proximal tubule cells. *Kidney International* 62(6): 2035-2042.

LANDRAUD, L., GAUTHIER, M., FOSSE, T. AND BOQUET, P. (2000) Frequency of *Escherichia coli* strains producing the cytotoxic necrotizing factor (CNF1) in nosocomial urinary tract infections. *Letters in Applied Microbiology* 30(3): 213-216.

LANZA, V. F., DE TORO, M., GARCILLÁN-BARCIA, M. P., MORA, A., BLANCO, J., COQUE, T. M. AND DE LA CRUZ, F. (2014) Plasmid Flux in Escherichia coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences. *PLoS Genetics* 10(12): e1004766.

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 Â -lactamase, all belong to the international O25:H4-ST131 clone. *Journal of Antimicrobial Chemotherapy* 62(6): 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. *Journal of Clinical Microbiology* 46(3): 1076-1080.

LAUPLAND, K., GREGSON, D., CHURCH, D., ROSS, T. AND PITOUT, J. (2008) Incidence, risk factors and outcomes of *Escherichia coli* bloodstream infections in a large Canadian region. *Clinical Microbiology and Infection* 14(11): 1041-1047.

LAWRENCE, J. G. AND OCHMAN, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences* 95(16): 9413-9417.

LEFFLER, H. AND EDEN, C. (1980) Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. *FEMS Microbiology Letters* 8(3): 127-134.

Lehman, R. (1981) *The Enzymes* (14th edition). Academic Press Inc., 198. Available at: http://lehman.stanford.edu/pdfs/117_endonucleases_specific.pdf (accessed 08/ 01/ 15).

LEVERSTEIN-VAN HALL, M., DIERIKX, C., STUART, J. C., VOETS, G., VAN DEN MUNCKHOF, M., VAN ESSEN-ZANDBERGEN, A., PLATTEEL, T., FLUIT, A., VAN DE SANDE-BRUINSMA, N., SCHARINGA, J., BONTEN, M. AND MEVIUS, D. (2011) Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clinical Microbiology and Infection* 17(6): 873-880.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. AND DURBIN, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079.

LIU, L., JOHNSON, H. L., COUSENS, S., PERIN, J., SCOTT, S., LAWN, J. E., RUDAN, I., CAMPBELL, H., CIBULSKIS, R., LI, M., MATHERS, C. AND BLACK, R. E. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet* 379(9832): 2151-2161.

LIVERMORE, D. M., CANTON, R., GNIADKOWSKI, M., NORDMANN, P., ROSSOLINI, G. M., ARLET, G., AYALA, J., COQUE, T. M., KERN-ZDANOWICZ, I., LUZZARO, F., POIREL, L. AND WOODFORD, N. (2007) CTX-M: changing the face of ESBLs in Europe. *Journal of Antimicrobial Chemotherapy* 59(2): 165-174.

LLOYD, A. L., RASKO, D. A. AND MOBLEY, H. L. T. (2007) Defining Genomic Islands and Uropathogen-Specific Genes in Uropathogenic *Escherichia coli*. *Journal of Bacteriology* 189(9): 3532-3546.

MAIDEN, M. C. J., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. AND SPRATT, B. G. (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences* 95(6): 3140-3145.

MANDAR, R. AND MIKELSAAR, M. (1996) Transmission of Mother’s Microflora to the Newborn at Birth. *Neonatology* 69(1): 30-35.

MANGES, A. R. AND JOHNSON, J. R. (2012) Food-Borne Origins of Escherichia coli Causing Extraintestinal Infections. *Clinical Infectious Diseases* 55(5): 712-719.

MANGES, A. R., JOHNSON, J. R., FOXMAN, B., O'BRYAN, T. T., FULLERTON, K. E. AND RILEY, L. W. (2001) Widespread Distribution of Urinary Tract Infections Caused by a Multidrug-Resistant *Escherichia coli* Clonal Group. *New England Journal of Medicine* 345(14): 1007-1013.

MANGES, A. R., TABOR, H., TELLIS, P., VINCENT, C. AND TELLIER, P. (2008) Endemic and Epidemic Lineages of *Escherichia coli* that Cause Urinary Tract Infections. *Emerg. Infect. Dis.* 14(10): 1575-1583.

MARSCHALL, J., AGNIEL, D., FRASER, V. J., DOHERTY, J. AND WARREN, D. K. (2008) Gram-negative bacteraemia in non-ICU patients: factors associated with inadequate antibiotic therapy and impact on outcomes. *Journal of Antimicrobial Chemotherapy* 61(6): 1376-1383.

MARTINEZ, J. J., MULVEY, M., SCHILLING, J., PINKNER, J. AND HULTGREN, S. (2000) Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *The EMBO Journal* 19(12): 2803-2812.

MARTINEZ-MEDINA, M., MORA, A., BLANCO, M., LOPEZ, C., ALONSO, M. P., BONACORSI, S., NICOLAS-CHANOINE, M., DARFEUILLE-MICHAUD, A., GARCIA-GIL, J. AND BLANCO, J. (2009) Similarity and Divergence among Adherent-Invasive *Escherichia coli* and Extraintestinal Pathogenic *E.coli* Strains. *Journal of Clinical Microbiology* 47(12): 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-β-lactamase-producing *Escherichia coli* in Japan. *Journal of Antimicrobial Chemotherapy* 67(11): 2612-2620.

MCBEAN, M. AND RAJAMANI, S. (2001) Increasing Rates of Hospitalization Due to Septicemia in the US Elderly Population, 1986â€“1997. *The Journal of Infectious Diseases* 183(4): 596-603.

MCNALLY, A., ALHASHASH, F., COLLINS, M., ALQASIM, A., PASZCKIEWICZ, K., WESTON, V. AND DIGGLE, M. (2013) Genomic analysis of extra-intestinal pathogenic *Escherichia coli* urosepsis. *Clinical Microbiology and Infection* 19(8): e328-e334.

MCNALLY, A., CHENG, L., HARRIS, S. R. AND CORANDER, J. (2013) The Evolutionary Path to Extraintestinal Pathogenic, Drug-Resistant *Escherichia coli* Is Marked by Drastic Reduction in Detectable Recombination within the Core Genome. *Genome Biology and Evolution* 5(4): 699-710.

MELICAN, K., SANDOVAL, R. M., KADER, A., JOSEFSSON, L., TANNER, G. A., MOLITORIS, B. A. AND RICHTER-DAHLFORS, A. (2011) Uropathogenic *Escherichia coli* P and Type 1 Fimbriae Act in Synergy in a Living Host to Facilitate Renal Colonization Leading to Nephron Obstruction. *PLoS Pathogens* 7(2): e1001298.

MELZER, M. AND PETERSEN, I. (2007) Mortality following bacteraemic infection caused by extended spectrum beta-lactamase (ESBL) producing *E.coli* compared to non-ESBL producing *E.coli*. *Journal of Infection* 55(3): 254-259.

METZKER, M. L. (2009) Sequencing technologies-the next generation. *Nat Rev Genet* 11(1): 31-46.

MIAJLOVIC, H., COOKE, N. M., MORAN, G. P., ROGERS, T. R. F. AND SMITH, S. G. (2014) Response of Extraintestinal Pathogenic *Escherichia coli* to Human Serum Reveals a Protective Role for Rcs-Regulated Exopolysaccharide Colanic Acid. *Infection and Immunity* 82(1): 298-305.

MILLS, M., MEYSICK, K. C. AND O'BRIEN, A. D. (2000) Cytotoxic Necrotizing Factor Type 1 of Uropathogenic *Escherichia coli* Kills Cultured Human Uroepithelial 5637 Cells by an Apoptotic Mechanism. *Infection and Immunity* 68(10): 5869-5880.

MLST DATABASES AT THE ERI, University College Cork. *Escherichia coli* MLST Databases.

http://mlst.ucc.ie/mlst/dbs/Ecoli (28 October 2012, date last accessed).

MOBLEY, H. L. T., GREEN, D. M., TRIFILLIS, A. L., JOHNSON, D. E., CHIPPENDALE, G. R., LOCKATELL, C. V., JONES, B. D. and WARREN, J. W. (1990) Pyelonephritogenic *Escherichia coli* and Killing of Cultured Human Renal Proximal Tubular Epithelial Cells: Role of Hemolysin in Some Strains. *INFECTION AND IMMUNITY* 58(5): 1281-1289.

MOLL, A., MANNING, P. A. and TIMMIS, K. N. (1980) Plasmid-Determined Resistance to Serum Bactericidal Activity: a Major Outer Membrane Protein, the traT Gene Product, Is Responsible for Plasmid-Specified Serum Resistance in Escherichia coli. *INFECTION AND IMMUNITY* 28(2): 359-367.

MONTENEGRO, M. A., BITTER-SUERMANN, D., TIMMIS, J. K., AGUERO, M. E., CABELLO, F. C., SANYAL, S. C. and TIMMIS, K. N. (1985) traT Gene Sequences, Serum Resistance and Pathogenicity-related Factors in Clinical Isolates of *Escherichia coli* and Other Gram-negative Bacteria. *Microbiology* 131(6): 1511-1521.

MORA, A., HERRERA, A., MAMANI, R., LOPEZ, C., ALONSO, M. P., BLANCO, J. E., BLANCO, M., DAHBI, G., GARCIA-GARROTE, F., PITA, J. M., COIRA, A., BERNARDEZ, M. I. AND BLANCO, J. (2010) Recent Emergence of Clonal Group O25b:K1:H4-B2-ST131 ibeA Strains among *Escherichia coli* Poultry Isolates, Including CTX-M-9-Producing Strains, and Comparison with Clinical Human Isolates. *Applied and Environmental Microbiology* 76(21): 6991-6997.

MORA, A., LÓPEZ, C., DABHI, G., BLANCO, M., BLANCO, J. E., ALONSO, M., 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(1): 132.

MORA, A., VISO, S., LÓPEZ, C., ALONSO, M. P., GARCÍA-GARROTE, F., DABHI, G., MAMANI, R., HERRERA, A., MARZOA, J., BLANCO, M., BLANCO, J. E., MOULIN-SCHOULEUR, M., SCHOULER, C. AND BLANCO, J. (2013) Poultry as reservoir for extraintestinal pathogenic *Escherichia coli* O45:K1:H7-B2-ST95 in humans. *Veterinary Microbiology* 167(3-4): 506-512.

MORENO, E., ANDREU, A., PIGRAU, C., KUSKOWSKI, M. A., JOHNSON, J. R. AND PRATS, G. (2008) Relationship between *Escherichia coli* Strains Causing Acute Cystitis in Women and the Fecal *E.coli* Population of the Host. *Journal of Clinical Microbiology* 46(8): 2529-2534.

MORENO, E., PLANELLS, I., PRATS, G., PLANES, A. M., MORENO, G. AND ANDREU, A. (2005) Comparative study of *Escherichia coli* virulence determinants in strains causing urinary tract bacteremia versus strains causing pyelonephritis and other sources of bacteremia. *Diagnostic Microbiology and Infectious Disease* 53(2): 93-99.

MOSSORO, C., GLAZIOU, P., YASSIBANDA, S., LAN, N. T. P., BEKONDI, C., MINSSART, P., BERNIER, C., LE BOUGUENEC, C. AND GERMANI, Y. (2002) Chronic Diarrhea, Hemorrhagic Colitis, and Hemolytic-Uremic Syndrome Associated with HEp-2 Adherent Escherichia coli in Adults Infected with Human Immunodeficiency Virus in Bangui, Central African Republic. *Journal of Clinical Microbiology* 40(8): 3086-3088.

MOULIN-SCHOULEUR, M., REPERANT, M., LAURENT, S., BREE, A., MIGNON-GRASTEAU, S., GERMON, P., RASSCHAERT, D. AND SCHOULER, C. (2007) Extraintestinal Pathogenic *Escherichia coli* Strains of Avian and Human Origin: Link between Phylogenetic Relationships and Common Virulence Patterns. *Journal of Clinical Microbiology* 45(10): 3366-3376.

MULVEY, M. A. (2002) Adhesion and entry of uropathogenic *Escherichia coli*. *Cell Microbiol* 4(5): 257-271.

NICOLAS-CHANOINE, M., BLANCO, J., LEFLON-GUIBOUT, V., DEMARTY, R., ALONSO, M. P., CANICA, M. M., PARK, Y., LAVIGNE, J., PITOUT, J. AND JOHNSON, J. R. (2008) Intercontinental emergence of *Escherichia coli* clone O25:H4-ST131 producing CTX-M-15. *Journal of Antimicrobial Chemotherapy* 61(2): 273-281.

NOVAIS, A., CANTON, R., MOREIRA, R., PEIXE, L., BAQUERO, F. AND COQUE, T. M. (2007) Emergence and Dissemination of Enterobacteriaceae Isolates Producing CTX-M-1-Like Enzymes in Spain Are Associated with IncFII (CTX-M-15) and Broad-Host-Range (CTX-M-1, -3, and -32) Plasmids. *Antimicrobial Agents and Chemotherapy* 51(2): 796-799.

NOVAIS, A., PIRES, J., FERREIRA, H., COSTA, L., MONTENEGRO, C., VUOTTO, C., DONELLI, G., COQUE, T. M. AND PEIXE, L. (2012) Characterization of Globally Spread *Escherichia coli* ST131 Isolates (1991 to 2010). *Antimicrobial Agents and Chemotherapy* 56(7): 3973-3976.

NOWICKI, B., SELVARANGAN, R. AND NOWICKI, S. (2001) Family of *Escherichia coli* Dr Adhesins: Decay-Accelerating Factor Receptor Recognition and Invasiveness. *The Journal of Infectious Diseases* 183(s1): S24-S27.

NOWROUZIAN, F., HESSELMAR, B., SAALMAN, R., STRANNEGÅRD, I., ÅBERG, N., WOLD, A. E. AND ADLERBERTH, I. (2003) *Escherichia coli* in Infants' Intestinal Microflora: Colonization Rate, Strain Turnover, and Virulence Gene Carriage. *Pediatr Res* 54(1): 8-14.

ODONKOR, S. T. AND AMPOFO, J. K. (2013) *Escherichia coli* as an indicator of bacteriological quality of water: an overview. *Microbiology Research* 4(1): 2.

OELSCHLAEGER, T., DOBRINDT, U. AND HACKER, J. (2002) Pathogenicity islands of uropathogenic *E.coli* and the evolution of virulence. *International Journal of Antimicrobial Agents* 19(6): 517-521.

OLESEN, B., HANSEN, D. S., NILSSON, F., FRIMODT-MOLLER, 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. *Journal of Clinical Microbiology* 51(6): 1779-1785.

OTEO, J., DIESTRA, K., JUAN, C., BAUTISTA, V., NOVAIS, Â., PÉREZ-VÁZQUEZ, M., MOYÁ, B., MIRÓ, E., COQUE, T. M., OLIVER, A., CANTÓN, R., NAVARRO, F. AND CAMPOS, J. (2009) Extended-spectrum β-lactamase-producing *Escherichia coli* in Spain belong to a large variety of multilocus sequence typing types, including ST10 complex/A, ST23 complex/A and ST131/B2. *International Journal of Antimicrobial Agents* 34(2): 173-176.

OTTO, B. R., VAN DOOREN, *. S. J. M., DOZOIS, C. M., LUIRINK, J. AND OUDEGA, B. (2002) *Escherichia coli* Hemoglobin Protease Autotransporter Contributes to Synergistic Abscess Formation and Heme-Dependent Growth of *Bacteroides fragilis*. *Infection and Immunity* 70(1): 5-10.

PEIRANO, G. AND PITOUT, J. D. (2010) Molecular epidemiology of *Escherichia coli* producing CTX-M β-lactamases: the worldwide emergence of clone ST131 O25:H4. *International Journal of Antimicrobial Agents* 35(4): 316-321.

PEIRANO, G., VAN DER BIJ, A. K., GREGSON, D. B. AND PITOUT, J. D. D. (2012) Molecular Epidemiology over an 11-Year Period (2000 to 2010) of Extended-Spectrum α-Lactamase-Producing *Escherichia coli* Causing Bacteremia in a Centralized Canadian Region. *Journal of Clinical Microbiology* 50(2): 294-299.

PETTY, N. K., BEN ZAKOUR, N. L., STANTON-COOK, M., SKIPPINGTON, E., TOTSIKA, M., FORDE, B. M., PHAN, M., GOMES MORIEL, D., PETERS, K. M., DAVIES, M., ROGERS, B. A., DOUGAN, G., RODRIGUEZ-BANO, J., PASCUAL, A., PITOUT, J. D. D., UPTON, M., PATERSON, D. L., WALSH, T. R., SCHEMBRI, M. A. AND BEATSON, S. A. (2014). Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proceedings of the National Academy of Sciences* 111(15): 5694-5699.

PHAN, M., PETERS, K. M., SARKAR, S., LUKOWSKI, S. W., ALLSOPP, L. P., MORIEL, D. G., ACHARD, M. E. S., TOTSIKA, M., MARSHALL, V. M., UPTON, M., BEATSON, S. A. AND SCHEMBRI, M. A. (2013). The Serum Resistome of a Globally Disseminated Multidrug Resistant Uropathogenic *Escherichia coli* Clone. *PLoS Genetics* 9(10): e1003834.

PICHON, C., HÉCHARD, C., DU MERLE, L., CHAUDRAY, C., BONNE, I., GUADAGNINI, S., VANDEWALLE, A. AND LE BOUGUÉNEC, C. (2009) Uropathogenic *Escherichia coli* AL511 requires flagellum to enter renal collecting duct cells. *Cellular Microbiology* 11(4): 616-628.

PITOUT, J. D. AND LAUPLAND, K. B. (2008) Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern. *The Lancet Infectious Diseases* 8(3): 159-166.

PITOUT, J. D. D., CHURCH, D. L., GREGSON, D. B., CHOW, B. L., MCCRACKEN, M., MULVEY, M. R. AND LAUPLAND, K. B. (2007) Molecular Epidemiology of CTX-M-Producing *Escherichia coli* in the Calgary Health Region: Emergence of CTX-M-15-Producing Isolates. *Antimicrobial Agents and Chemotherapy* 51(4): 1281-1286.

PITOUT, J. D. D., NORDMANN,, P., LAUPLAND, K. AND POIREL, L. (2005) Emergence of Enterobacteriaceae producing extended-spectrum α-lactamases (ESBLs) in the community. *Journal of Antimicrobial Chemotherapy* 56(1): 52-59.

PLATELL, J. L., COBBOLD, R. N., JOHNSON, J. R., HEISIG, A., HEISIG, P., CLABOTS, C., KUSKOWSKI, M. A. AND TROTT, D. J. (2011a) Commonality among Fluoroquinolone-Resistant Sequence Type ST131 Extraintestinal *Escherichia coli* Isolates from Humans and Companion Animals in Australia. *Antimicrobial Agents and Chemotherapy* 55(8): 3782-3787.

PLATELL, J. L., JOHNSON, J. R., COBBOLD, R. N. AND TROTT, D. J. (2011b) Multidrug-resistant extraintestinal pathogenic *Escherichia coli* of sequence type ST131 in animals and foods. *Veterinary Microbiology* 153(1-2): 99-108.

PLOWMAN, R., GRAVES, N., GRIFFIN, M., ROBERTS, J., SWAN, A., COOKSON, B. AND TAYLOR, L. (2001) The rate and cost of hospital-acquired infections occurring in patients admitted to selected specialties of a district general hospital in England and the national burden imposed. *Journal of Hospital Infection* 47(3): 198-209.

PODCHUN, R., FISCHER, A. AND ULLMANN, U. (2001) Characterisation of *Hafnia alvei* isolates from human clinical extra-intestinal specimens: haemagglutinins, serum resistance and siderophore synthesis. *J Med Microbiol* 50(3): 208-214.

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-α-Lactamase-Producing *Escherichia coli* ST131 Is Driven by a Single Highly Pathogenic Subclone, H30-Rx. *mBio* 4(6): e00377-13-e00377-13.

RAMCHANDANI, M., MANGES, A. R., DEBROY, C., SMITH, S. P., JOHNSON, J. R. AND RILEY, L. W. (2005) Possible Animal Origin of Human-Associated, Multidrug-Resistant, Uropathogenic *Escherichia coli*. *Clinical Infectious Diseases* 40(2): 251-257.

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(20): 6881-6893.

RICHTER, A. M., POVOLOTSKY, T. L., WIELER, L. H. AND HENGGE, R. (2014) Cyclic-di-GMP signalling and biofilm-related properties of the Shiga toxin-producing 2011 German outbreak *Escherichia coli* O104:H4. *EMBO Molecular Medicine* 6(12): 1622-1637.

RIJAVEC, M., MULLER-PREMRU, M., ZAKOTNIK, B. AND ZGUR-BERTOK, D. (2008) Virulence factors and biofilm production among *Escherichia coli* strains causing bacteraemia of urinary tract origin. *Journal of Medical Microbiology* 57(11): 1329-1334.

RILEY L., REMIS R., HELGERSON S., MCGEE H., WELLS J., DAVIS B., HEBERT R., OLCOTT E., JOHNSON L., HARGRETT N., BLAKE P. AND COHEN M. (1983). Hemorrhagic Colitis Associated with a Rare *Escherichia coli* Serotype. *New England Journal of Medicine* 308(12): 681-685.

RODRÍGUEZ, F., OLIVER, J., MARÍN, A. AND MEDINA, J. (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* 142(4): 485-501.

RODRIGUEZ-BANO, J., NAVARRO, M. D., ROMERO, L., MUNIAIN, M. A., CUETO, M. D., RIOS, M. J., HERNANDEZ, J. R. AND PASCUAL, A. (2006) Bacteremia Due to Extended-Spectrum α-Lactamase-Producing *Escherichia coli* in the CTX-M Era: A New Clinical Challenge. *Clinical Infectious Diseases* 43(11): 1407-1414.

RODRÍGUEZ-BAÑO, J., NAVARRO, M., ROMERO, L., MUNIAIN, M., DE CUETO, M., GÁLVEZ, J., PEREA, E. AND PASCUAL, A. (2008) Risk-factors for emerging bloodstream

infections caused by extended-spectrum β-lactamase-producing *Escherichia coli*. *Clinical Microbiology and Infection* 14(2): 180-183.

ROGERS, B. A., SIDJABAT, H. E. AND PATERSON, D. L. (2011) *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. *Journal of Antimicrobial Chemotherapy* 66(1): 1-14.

RON, E. Z. (2010) Distribution and evolution of virulence factors in septicemic *Escherichia coli*. *International Journal of Medical Microbiology* 300(6): 367-370.

ROSSOLINI, G., D'ANDREA, M. AND MUGNAIOLI, C. (2008) The spread of CTX-M-type extended-spectrum β-lactamases. *Clinical Microbiology and Infection* 14: 33-41.

RUSSO, T. AND JOHNSON, J. (2000) Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of *Escherichia coli*: ExPEC. *The Journal of Infectious Diseases* 181(5): 1753-1754.

RUSSO, T. A. AND JOHNSON, J. R. (2003) Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes and Infection* 5(5): 449-456.

SAHUQUILLO-ARCE, J. M., SELVA, M., PERPINAN, H., GOBERNADO, M., ARMERO, C., LOPEZ-QUILEZ, A., GONZALEZ, F. AND VANACLOCHA, H. (2011) Antimicrobial Resistance in More than 100,000 *Escherichia coli* Isolates According to Culture Site and Patient Age, Gender, and Location. *Antimicrobial Agents and Chemotherapy* 55(3): 1222-1228.

SANGER.AC.UK., 2014. SMALT - Wellcome Trust Sanger Institute [Online]. Available at: https://www.sanger.ac.uk/resources/software/smalt/#t_1 [Accessed: 2 October 2014].

SANTOS MENDONÇA, R. C., MORELLI, A. M. F., PEREIRA, J. A. M., DE CARVALHO, M. M. AND DE SOUZA, N. L. (2012) Prediction of *Escherichia coli* O157:H7 adhesion and potential to form biofilm under experimental conditions. *Food Control* 23(2): 389-396.

SARKAR, S., ULETT, G. C., TOTSIKA, M., PHAN, M. AND SCHEMBRI, M. A. (2014) Role of Capsule and O Antigen in the Virulence of Uropathogenic *Escherichia coli*. *PLoS ONE* 9(4): e94786.

SCHWAN, W. R. (2008) Flagella allow uropathogenic *Escherichia coli* ascension into murine kidneys. *International Journal of Medical Microbiology* 298(5-6): 441-447.

SEEMANN, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068-2069.

SERVIN, A. L. (2005) Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*. *Clinical Microbiology Reviews* 18(2): 264-292.

SHEPARD, S. M., DANZEISEN, J. L., ISAACSON, R. E., SEEMANN, T., ACHTMAN, M. AND JOHNSON, T. J. (2011) Genome Sequences and Phylogenetic Analysis of K88- and F18-Positive Porcine Enterotoxigenic *Escherichia coli*. *Journal of Bacteriology* 194(2): 395-405.

SIROT, D., SIROT, J., LABIA, R., MORAND, A., COURVALIN, P., DARFEUILLE-MICHAUD, A., PERROUX, R. AND CLUZEL, R. (1987) Transferable resistance to third-generation cephalosporins in clinical isolates of *Klebsiella pneumoniae* : identification of CTX-1, a novel Î²-lactamase. *Journal of Antimicrobial Chemotherapy* 20(3): 323-334.

SKJØT-RASMUSSEN, L., EJRNÆS, K., LUNDGREN, B., HAMMERUM, A. M. AND FRIMODT-MØLLER, N. (2012) Virulence factors and phylogenetic grouping of *Escherichia coli* isolates from patients with bacteraemia of urinary tract origin relate to sex and hospital-

vs. community-acquired origin. *International Journal of Medical Microbiology* 302(3): 129-134.

SKJØT-RASMUSSEN, L., OLSEN, S., JAKOBSEN, L., EJRNÆS, K., SCHEUTZ, F., LUNDGREN, B., FRIMODT-MØLLER, N. AND HAMMERUM, A. (2013) *Escherichia coli* clonal group A causing bacteraemia of urinary tract origin. *Clinical Microbiology and Infection* 19(7): 656-661.

SMITH, B. A., FAZIL, A. AND LAMMERDING, A. M. (2013) A risk assessment model for *Escherichia coli* O157:H7 in ground beef and beef cuts in Canada: Evaluating the effects of interventions. *Food Control* 29(2): 364-381.

SMITH, J. L., FRATAMICO, P. M. AND GUNTHER, N. W. (2007) Extraintestinal Pathogenic *Escherichia coli*. *Foodborne Pathogens and Disease* 4(2): 134-163.

SMITH, Y. C., GRANDE, K. K., RASMUSSEN, S. B. AND O'BRIEN, A. D. (2006) Novel Three-Dimensional Organoid Model for Evaluation of the Interaction of Uropathogenic *Escherichia coli* with Terminally Differentiated Human Urothelial Cells. *Infection and Immunity* 74(1): 750-757.

SOKURENKO, E. V., CHESNOKOVA, V., DYKHUIZEN, D. E., OFEK, I., WU, X., KROGFELT, K. A., STRUVE, C., SCHEMBRI, M. A. AND HASTY, D. L. (1998) Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proceedings of the National Academy of Sciences* 95(15): 8922-8926.

SOLBERG, O. D., AJIBOYE, R. M. AND RILEY, L. W. (2006) Origin of Class 1 and 2 Integrons and Gene Cassettes in a Population-Based Sample of Uropathogenic *Escherichia coli*. *Journal of Clinical Microbiology* 44(4): 1347-1351.

SOUGAKOFF, W., GOUSSARD, S. AND COURVALIN, P. (1988) The TEM-3 β-lactamase, which hydrolyzes broad-spectrum cephalosporins, is derived from the TEM-2 penicillinase by two amino acid substitutions. *FEMS Microbiology Letters* 56(3): 343-348.

SPURBECK, R. R., DINH, P. C., WALK, S. T., STAPLETON, A. E., HOOTON, T. M., NOLAN, L. K., KIM, K. S., JOHNSON, J. R. and MOBLEY, H. L. T. (2012) *Escherichia coli* Isolates That Carry vat, fyuA, chuA, and yfcV Efficiently Colonize the Urinary Tract. *Infection and Immunity* 80(12): 4115-4122.

STAMATAKIS, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688-2690.

STAMEY, T. A. AND TIMOTHY, M. M. (1976) STUDIES OF introital colonization in women with recurrent urinary infections. I. The role of vaginal pH. *Obstetrical & Gynecological Survey* 31(9): 696.

STÜRENBURG, E. AND MACK, D. (2003) Extended-spectrum β-lactamases: implications for the clinical microbiology laboratory, therapy, and infection control. *Journal of Infection* 47(4): 273-295.

SUZUKI, S., SHIBATA, N., YAMANE, K., WACHINO, J., ITO, K. AND ARAKAWA, Y. (2009) Change in the prevalence of extended-spectrum-α -lactamase-producing *Escherichia coli* in Japan by clonal spread. *Journal of Antimicrobial Chemotherapy* 63(1): 72-79.

SWAIN, M. T., TSAI, I. J., ASSEFA, S. A., NEWBOLD, C., BERRIMAN, M. AND OTTO, T. D. (2012) A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nature Protocols* 7(7): 1260-1284.

SYSTEMS.ILLUMINA.COM, 2014. HiSeq 2500 Ultra-High-Throughput Sequencing System Illumina [Online]. Available at: http://systems.illumina.com/systems/hiseq_2500_1500.ilmn [Accessed: 25 September 2014].

SZCZEPANOWSKI, R., BRAUN, S., RIEDEL, V., SCHNEIKER, S., KRAHN, I., PUHLER, A. AND SCHLUTER, A. (2005) The 120 592 bp IncF plasmid pRSB107 isolated from a sewage-treatment plant encodes nine different antibiotic-resistance determinants, two iron-acquisition systems and other putative virulence-associated functions. *Microbiology* 151(4): 1095-1111.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. AND KUMAR, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30(12): 2725-2729.

TARTOF, S. Y., SOLBERG, O. D., MANGES, A. R. AND RILEY, L. W. (2005) Analysis of a Uropathogenic *Escherichia coli* Clonal Group by Multilocus Sequence Typing. *Journal of Clinical Microbiology* 43(12): 5860-5864.

TENAILLON, O., SKURNIK, D., PICARD, B. AND DENAMUR, E. (2010) The population genetics of commensal *Escherichia coli*. *Nat Rev Micro* 8(3): 207-217.

TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., WARD, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., MARGARIT Y ROS, I., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C., MADUPU, R., BRINKAC, L. M., DODSON, R. J., ROSOVITZ, M. J., SULLIVAN, S. A., DAUGHERTY, S. C., HAFT, D. H., SELENGUT, J., GWINN, M. L., ZHOU, L., ZAFAR, N., KHOURI, H., RADUNE, D., DIMITROV, G., WATKINS, K., O'CONNOR, K. J. B., SMITH, S., UTTERBACK, T. R., WHITE, O., RUBENS, C. E., GRANDI, G., MADOFF, L. C., KASPER, D. L., TELFORD, J.

L., WESSELS, M. R., RAPPUOLI, R. AND FRASER, C. M. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences* 102(39): 13950-13955.

TETTELIN, H., RILEY, D., CATTUTO, C. AND MEDINI, D. (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology* 11(5): 472-477.

THE DEPARTMENT OF HEALTH. Extension of mandatory surveillance to *E. coli* bloodstream Infections - June 2011.

http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_126221.pdf (26 October 2012, date last accessed).

THUMBIKAT, P., BERRY, R. E., ZHOU, G., BILLIPS, B. K., YAGGIE, R. E., ZAICHUK, T., SUN, T., SCHAEFFER, A. J. AND KLUMPP, D. J. (2009) Bacteria-Induced Uroplakin Signaling Mediates Bladder Response to Infection. *PLoS Pathogens* 5(5): e1000415.

TIVENDALE, K. A., LOGUE, C. M., KARIYAWASAM, S., JORDAN, D., HUSSEIN, A., LI, G., WANNEMUEHLER, Y. AND NOLAN, L. K. (2010) Avian-Pathogenic Escherichia coli Strains Are Similar to Neonatal Meningitis *E.coli* Strains and Are Able To Cause Meningitis in the Rat Model of Human Disease. *Infection and Immunity* 78(8): 3412-3419.

TO, K. K., LO, W., CHAN, J. F., TSE, H., CHENG, V. C. AND HO, P. (2013) Clinical outcome of extended-spectrum beta-lactamase-producing *Escherichia coli* bacteremia in an area with high endemicity. *International Journal of Infectious Diseases* 17(2): e120-e124.

TOBE, T., HAYASHI, T., HAN, C., SCHOOLNIK, G., OHTSUBO, E. AND SASAKAWA, C. (1999) Complete DNA Sequence and Structural Analysis of the Enteropathogenic *Escherichia coli* Adherence Factor Plasmid. *Infection and Immunity* 67(10): 5455-5462.

TOTSIKA, M., BEATSON, S. A., SARKAR, S., PHAN, M., PETTY, N. K., BACHMANN, N., SZUBERT, M., SIDJABAT, H. E., PATERSON, D. L., UPTON, M. AND SCHEMBRI, M. A. (2011) Insights into a Multidrug Resistant *Escherichia coli* Pathogen of the Globally Disseminated ST131 Lineage: Genome Analysis and Virulence Mechanisms. *PLoS ONE* 6(10): e26578.

TREE.BIO.ED.AC.UK/SOFTWARE, 2014. FigTree Software [Online]. Available at: http://tree.bio.ed.ac.uk/software/ [Accessed: 2 October 2014].

UNDERWOOD, J., KLEIN, J. AND NEWSHOLME, W. (2011) *Escherichia coli* bacteraemia: how preventable is it?. *Journal of Hospital Infection* 79(4): 364-365.

VÄISÄNEN, V., TALLGREN, L., HELENA MÄKELÄ, P., KÄLLENIUS, G., HULTBERG, H., ELO, J., SIITONEN, A., SVANBORG-EDÉN, C., SVENSON, S. AND KORHONEN, T. (1981) Mannose-resistant haemagglutination and p antigen recognition are characteristic of *Escherichia coli* causing primary pyelonephritis. *The Lancet* 318(8260-8261): 1366-1369.

VAN ELSAS, J. D., SEMENOV, A. V., COSTA, R. AND TREVORS, J. T. (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *The ISME Journal* 5(2): 367-367.

VILLA, L., GARCIA-FERNANDEZ, A., FORTINI, D. AND CARATTOLI, A. (2010) Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *Journal of Antimicrobial Chemotherapy* 65(12): 2518-2529.

VINCENT, C., BOERLIN, P., DAIGNAULT, D., DOZOIS, C. M., DUTIL, L., GALANAKIS, C., REID-SMITH, R. J., TELLIER, P., TELLIS, P. A., ZIEBELL, K. AND MANGES, A. R. (2010) Food Reservoir for *Escherichia coli* Causing Urinary Tract Infections. *Emerg. Infect. Dis.* 16(1): 88-95.

Virkola, R. (1987) Binding characteristics of *Escherichia coli* type 1 fimbriae in the human kidney. *FEMS Microbiology Letters* 40(2-3): 257-262.

VON MENTZER, A., CONNOR, T. R., WIELER, L. H., SEMMLER, T., IGUCHI, A., THOMSON, N. R., RASKO, D. A., JOFFRE, E., CORANDER, J., PICKARD, D., WIKLUND, G., SVENNERHOLM, A., SJÖLING, Å. AND DOUGAN, G. (2014) Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* 46(12): 1321-1326.

WAGENLEHNER, F. M., WEIDNER, W. AND NABER, K. G. (2007) Optimal management of urosepsis from the urological perspective. *International Journal of Antimicrobial Agents* 30(5): 390-397.

WARREN, J. W., ABRUTYN, E., HEBEL, J. R., JOHNSON, J. R., SCHAEFFER, A. J. AND STAMM, W. E. (1999) Guidelines for Antimicrobial Treatment of Uncomplicated Acute Bacterial Cystitis and Acute Pyelonephritis in Women. *Clinical Infectious Diseases* 29(4): 745-758.

WASWA, J., IRUDAYARAJ, J. AND DEBROY, C. (2007) Direct detection of *E.coli* O157:H7 in selected food systems by a surface plasmon resonance biosensor. *LWT - Food Science and Technology* 40(2): 187-192.

WATT, S., LANOTTE, P., MEREGHETTI, L., MOULIN-SCHOULEUR, M., PICARD, B. AND QUENTIN, R. (2003) *Escherichia coli* Strains from Pregnant Women and Neonates: Intraspecies Genetic Distribution and Prevalence of Virulence Factors. *Journal of Clinical Microbiology* 41(5): 1929-1935.

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*. *Molecular Microbiology* 59(3): 975-988.

WELCH, R. A., BURLAND, V., PLUNKETT, G., REDFORD, P., ROESCH, P., RASKO, D., BUCKLES, E. L., LIOU, S. R., BOUTIN, A., HACKETT, J., STROUD, D., MAYHEW, G. F., ROSE, D. J., ZHOU, S., SCHWARTZ, D. C., PERNA, N. T., MOBLEY, H. L. T., DONNENBERG, M. S. AND BLATTNER, F. R. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences* 99(26): 17020-17024.

WELLCOME TRUST SANGER INSTITUTE, 2014a. SMALT - A New Mapper for DNA Sequencing Reads [Online]. Available at: http://cdn.f1000.com/posters/docs/327 [Accessed: 2 October 2014].

WELLCOME TRUST SANGER INSTITUTE, 2014b. SMALT - An efficient and accurate aligner for DNA sequencing reads [Online]. Available at: http://www.bdebate.org/sites/default/files/ficheros/articulos/smaap_zemin_ning_rgasp3_bdebate.pdf [Accessed: 2 October 2014].

WILES, T. J., KULESUS, R. R. AND MULVEY, M. A. (2008) Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Experimental and Molecular Pathology* 85(1): 11-19.

WILLEMS, R. J. L., TOP, J., VAN SCHAIK, W., LEAVIS, H., BONTEN, M., SIREN, J., HANAGE, W. P. AND CORANDER, J. (2012) Restricted Gene Flow among Hospital Subpopulations of *Enterococcus faecium*. *mBio* 3(4): e00151-12-e00151-12.

WILSON, J., COELLO, R., JEPSON, A., BRANNIGAN, E., RICHARDS, M., HASSALL, S. AND HOLMES, A. (2011) Increasing burden of *E.coli* bacteraemia and changing epidemiology. *BMC Proceedings* 5(Suppl 6): O3.

WIRTH, T., FALUSH, D., LAN, R., COLLES, F., MENSA, P., WIELER, L. H., KARCH, H., REEVES, P. R., MAIDEN, M. C. J., OCHMAN, H. AND ACHTMAN, M. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular Microbiology* 60(5): 1136-1151.

WOODFORD, N., WARD, M., KAUFMANN, M., TURTON, J., FAGAN, E., JAMES, D., JOHNSON, A., PIKE, R., WARNER, M., CHEASTY, T., PEARSON, A., HARRY, S., LEACH, J., A, L., LOWES, J., WARREN, R. AND LIVERMORE, D. (2004) Community and hospital spread of *Escherichia coli* producing CTX-M extended-spectrum Â -lactamases in the UK. *Journal of Antimicrobial Chemotherapy* 54(4): 735-743.

WRIGHT, K. J., SEED, P. C. AND HULTGREN, S. J. (2007) Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cellular Microbiology* 9(9): 2230-2241.

WU, U., WANG, J., CHEN, W., CHANG, S. AND CHEN, Y. (2011) Risk factors and outcomes of *Escherichia coli* bacteremia caused by strains that produce CTX-M or non-CTX-M extended-spectrum-beta-lactamases. *Eur J Clin Microbiol Infect Dis* 30(1): 33-39.

WULLT, B., BERGSTEN, G., CONNELL, H., RÖLLANO, P., GEBRETSADIK, N., HULL, R. AND SVANBORG, C. (2000) P fimbriae enhance the early establishment of *Escherichia coli* in the human urinary tract. *Molecular Microbiology* 38(3): 456-464.

WULLT, B., BERGSTEN, G., SAMUELSSON, M. AND SVANBORG, C. (2002) The role of P fimbriae for *Escherichia coli* establishment and mucosal inflammation in the human urinary tract. *International Journal of Antimicrobial Agents* 19(6): 522-538.

YAMAICHI, Y., CHAO, M. C., SASABE, J., CLARK, L., DAVIS, B. M., YAMAMOTO, N., MORI, H., KUROKAWA, K. AND WALDOR, M. K. (2014). High-resolution genetic analysis

of the requirements for horizontal transmission of the ESBL plasmid from *Escherichia coli* O104:H4. *Nucleic Acids Research* 43(1): 348-360.

YAMAMOTO, S. (2007). Molecular epidemiology of uropathogenic *Escherichia coli*. *Journal of Infection and Chemotherapy* 13(2): 68-73.

ZDZIARSKI, J., SVANBORG, C., WULLT, B., HACKER, J. AND DOBRINDT, U. (2008) Molecular Basis of Commensalism in the Urinary Tract: Low Virulence or Virulence Attenuation?. *Infection and Immunity* 76(2): 695-703.

ZERBINO, D. R. AND BIRNEY, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18(5): 821-829.

ZIENKIEWICZ, M., KERN-ZDANOWICZ, I., GOLEBIEWSKI, M., ZYLINSKA, J., MIECZKOWSKI, P., GNIADKOWSKI, M., BARDOWSKI, J. AND CEGLOWSKI, P. (2007) Mosaic Structure of p1658/97, a 125-Kilobase Plasmid Harboring an Active Amplicon with the Extended-Spectrum α-Lactamase Gene blaSHV-5. *Antimicrobial Agents and Chemotherapy* 51(4): 1164-1171.

ZONG, Z., YU, R., WANG, X. AND LU, X. (2010) blaCTX-M-65 is carried by a Tn1722-like element on an IncN conjugative plasmid of ST131 *Escherichia coli*. *Journal of Medical Microbiology* 60(4): 435-441.