Determining the fitness and competitive advantage of the epidemic ExPEC clone *E. coli* ST131

Abdulaziz Saad Alqasim

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

February 2015

Abstract

Extraintestinal pathogenic *E. coli* (ExPEC) is the major aetiological agent of urinary tract infections (UTIs) in humans. The emergence of the CTX-M producing *E. coli* ST131 clone represents a major challenge to public health worldwide because of its ability to cause a wide range of difficult-to-treat infections in the healthcare and community settings. The key aim of this study was to characterise the traits that give *E. coli* ST131 a competitive fitness advantage over other potential ExPEC clones. Comparative phenotypic characterisation of a collection of ExPEC strains showed that there was no difference between ST131 and non-ST131 strains in terms of their growth rates in different culture media, their capacity to associate with, invade and form intracellular bacterial communities within T24 human bladder epithelial cells and their ability to persist within U937 human macrophages. Afterwards, this study tested and compared the metabolic activity of a collection of ST131 and non-ST131 strains using two different testing methodologies: API strips and phenotypic microarray (PM) technology. Our API data showed that ST131 strains had a lower metabolic activity for 5 substrates. Further testing of the metabolic activity of *E. coli* using phenotypic microarray demonstrated the absence of a specific metabolic profile for ST131 strains suggesting that ST131 is not a metabolically distinct lineage of ExPEC and thus altered metabolism might not contribute to the fitness of this clone. The gene content of a group of *E. coli* including ST131 and non-ST131 strains was investigated to identify the presence of other loci that are uniquely associated with ST131 H30Rx clade, which involves ST131 isolates belonging to the *fimH30* lineage and

associated with fluoroquinolones resistance and CTX-M-15 production. Our data identified the presence of 150 loci unique to ST131 H30Rx strains, and the most striking finding at a genomic level was the identification of the secondary flagellar locus Flag 2 as a region uniquely associated with ST131 H30Rx strains. The ability of a collection of ST131 and non-ST131 strains to resist human serum was tested and compared. Our data showed that all ST131 and ST73 strains were associated with high serum resistance phenotype, and this might suggest serum resistance as an important factor in driving the current success of this ST131 as a major cause of bloodstream infections worldwide. Given many reports showing that polysaccharide capsules might be a major factor allowing *E. coli* to resist the human serum, and based on many studies demonstrating the genetic and biochemical diversity in the capsule region of ST131 strains, the capsule region of a collection of ExPEC belonging to ST131 H30Rx clade and non-ST131 was tested in more detail at a genomic and biochemical level. Our capsule genetics data showed a surprising level of diversity within the capsule locus of the H30Rx clade with a phylogenetic distribution highly suggestive of frequent recombination at the locus. Subsequent analysis demonstrated that this recombination had no obvious detectable effect on virulence-associated phenotypes *in-vitro*. Given the level of diversity observed at the capsule locus of ST131 H30Rx strains, it is tempting to speculate that there is significant selective pressure occurring at this site during the life cycle of the H30Rx clade, and that frequent recombination allows the clade to subvert that pressure and might provide a fitness advantage to ST131. This study provided detailed insights into the phenotypic, metabolic and genetic traits of ST131 and highlighted the factors that might drive its success.

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 and in the Microbiology Research Laboratory at the University of Surrey. Comparative metabolic studies were carried out in collaboration with Profs Roberto La Ragione and Richard Emes. The phenotypic microarray (PM) assay was carried out with the help of Dr Jane Newcombe, University of Surrey. Classical capsule typing assays were performed at the Statens Serum Institute, Denmark. Finally, writing the Post Assembly Genome Improvement Toolkit (PAGIT) script, pan-genome analysis using the Large-Scale Blast Score Ratio (LS-BSR) software and whole-genome phylogeny work were carried out by my director of studies Dr Alan McNally.

Abdulaziz Alqasim

Acknowledgments

I would like to thank Almighty Allah for giving me the strength and courage to lead this project to completion.

There are many people to whom I owe my gratitude for helping me with the completion of this PhD project. Firstly, my very special thanks to my director of studies Dr Alan McNally for giving me the opportunity to carry out this PhD research project and for his support and encouragement throughout this project. This work would have not been possible without his invaluable advice and guidance. I would like to thank our collaborators Prof. Roberto La Ragione and Dr. Richard Emes for their efforts in our published work on comparative phenotypic microarray analysis of ExPEC strains. My deep thanks are also extended to Mr Gordon Arnott for his assistance with using the confocal microscope.

I would like to extend my sincerest thanks and appreciation to my lovely father, mother, brother and sisters for their prayers and support throughout all my studies. Special thanks go to my beloved fiancée Mai, who has provided me with her never-ending emotional and practical support over the past year.

I also owe my thanks to all members in the Pathogen Research Group, Nottingham Trent University for their help and kindness in the past four years. I would like also to thank my colleagues Fahad, Dhahi, Alya and Miquette for providing a co-operative, comfortable and most enjoyable environment for work. The administrative and financial support of the Saudi Arabian Cultural Bureau is gratefully acknowledged.

Publications

1. Research articles

Alqasim, A., Emes, R., Clark, G., Newcombe, J., La Ragione, R. & McNally, A. (2014). Phenotypic microarrays suggest *Escherichia coli* ST131 is not a metabolically distinct lineage of extra-Intestinal pathogenic *E. coli*. *PloS One,* **9,** e88374.

Alqasim, A., Scheutz, F., Zong, Z. & 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**,** 830.

2. Poster presentations

Alqasim, A., Diggle, M., Weston, V., Cheetham P. & McNally, A. (2011). Investigation of clinical *E. coli* ST131 isolates. Society of General Microbiology Autumn Conference, University of York.

Alqasim, A., Diggle, M., Weston, V., Cheetham P. & McNally, A. (2012). Determining the fitness and competitive advantage of the epidemic ExPEC clone *E. coli* ST131. School of Science & Technology Research Conference, Nottingham Trent University.

Alqasim, A., Emes, R., Newcombe, J., La Ragione, R. & McNally, A. (2013). Phenotypic microarray analysis of ExPEC strains of different sequence types shows the absence of ST specific metabolism. Federation of Microbiological Societies in Europe Conference, Leipzig, Germany.

Table of Contents

List of Tables

List of Figures

Abbreviations

AIEC	Adherent invasive *E. coli*
APEC	Avian pathogenic *E. coli*
ATCC	American Type Culture Collection
BMM	bone marrow-derived macrophages
Bp	Base-pair
BSIs	Bloodstream infections
CDS	Coding sequences
CFU/ml	Colony-forming unit/ml
CNF1	Cytotoxic necrotising factor 1
CTX-M	Active on CefoTaXime, first isolated in Munich
DAEC	Diffusely adherent *E. coli*
DAPI	4', 6-diamidino-2-phenylindole
DNTPs	Deoxyribonucleotide triphosphate
E. coli	*Escherichia coli*
E. fergusonii	*Escherichia fergusonii*
EAEC	Enteroaggregative *E. coli*
EAHEC	Enteroaggregative haemorrhagic *E. coli*
ECOR	*E. coli* Reference Collection
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
LEE	Enterocyte effacement
EPEC	Enteropathogenic *E. coli*
ESBLs	Extended-spectrum β-lactamases
ESC	Extended-spectrum cephalosporins
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
FBS	Foetal bovine serum

G + C	Guanine and cytosine
G/L	Gram/Litre
GFP	Green florescent protein
HAIs	Healthcare-associated infections
IBCs	Intracellular bacterial communities
IPEC	Intestinal pathogenic *E. coli*
Kb	Kilo-base
LB	Luria-Bertani
LBB	Luria-Bertani broth
LPS	Lipopolysaccharides
LS-BSR	Large-Scale Blast Score Ratio
MDR	Multidrug-resistant
MGR	Maximal growth rate
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
M.O.I	Multiplicity of infection
MRSA	Methicillin-resistant *Staphylococcus aurous*
NCBI	National Centre of Biotechnology Information
NEAA	Non-essential amino acid
NGS	Next-generation sequencing
NMEC	Neonatal meningitis-associated *E. coli*
OD	Optical density
PAIs	Pathogenicity islands
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PHE	Public Health England
PM	Phenotypic microarray

PMA	Phoebe-myristate-acetate
P. mirabilis	*Proteus mirabilis*
SAT	Secreted autotransporter toxin
SEPEC	Sepsis-associated *E. coli*
SHV	Sulfhydryl variable
SNP	Single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
ST	Sequence type
S. saprophyticus	*Staphylococcus saprophyticus*
STEC	Shiga toxin-producing *E. coli*
S. typhimurium	*Salmonella typhimurium*
TEM	ESBL variant isolated from a patient named Temoneira
TraDIS	Transposon directed insertion-site sequencing
Two-way ANOVA	Two-way analysis of variance
UPEC	Uropathogenic *E. coli*
UPMGA	Unweighted Pair Group Method with Arithmetic Mean
UTI	Urinary tract infection
UV	Ultraviolet
V/V	Volume/volume
VF	Virulence factor
VAG	Virulence-associated gene
W/V	Weight/volume

Chapter one

Introduction

1. Introduction

1.1 The species *Escherichia coli*

Escherichia coli (*E. coli*) was first discovered, as "*Bacterium coli commune",* in the gut in 1885 by the German bacteriologist-paediatrician Theodore von Escherich (Escherich 1988). In 1897, Alphonse Lesage proposed that *E. coli* comprised harmless strains as well as other strains with distinct pathogenic potential (Lesage 1897). Over the past decades, *E. coli* has been the subject of an enormous number of studies and become one of the best-characterised prokaryotic organisms (Kunin 1962). However it remains a major cause of a wide range of intestinal and extraintestinal infections in humans (Pitout 2012a). *E. coli* is a member of the genus *Escherichia* which is one of the key genera of the *Enterobacteriaceae* family (Madigan *et al.* 2009). *E. coli* is a facultatively anaerobic, gram-negative, nonsporulating bacillus. It can either be motile or non-motile, depending on the possession of peritrichous flagella (Madigan *et al.* 2009). Additionally, *E. coli* grows in temperatures ranging from 7°C to 50°C and can tolerate acidic condition below pH 2.5 (Richard and Foster 2003, Kaper *et al.* 2004, Kramer *et al.* 2006, Kuhnert *et al.* 2006). *E. coli* is one of the most commonly isolated organisms from clinical specimens (Russo and Johnson 2003), and it can be recovered from these specimens using simple culture media incubated at 37°C either with or without oxygen (Weintraub 2007).

1.2 The classification of *E. coli* strains

Human-associated *E. coli* strains can be broadly classified, according to their genetic and clinical features, into three groups: commensal *E. coli* strains, enteric or intestinal pathogenic *E. coli* (IPEC) strains and

extraintestinal pathogenic *E. coli* (ExPEC) strains (Russo and Johnson 2000). Commensal *E. coli* strains constitute part of the normal intestinal flora in human and mammal hosts, and they are usually incapable of causing intestinal tract infections since they lack the specialized virulence factors that allow pathogenic *E. coli* strains to cause disease. However commensal *E. coli* may cause extraintestinal infections in some patient groups such as immunocompromised patients (Russo and Johnson 2003). The *E. coli* K-12 is an example of commensal *E. coli* strains (Kuhnert *et al.* 1995).

Intestinal pathogenic *E. coli* (IPEC) strains are associated with causing gastrointestinal infections such as enteritis and colitis (Russo and Johnson 2003). On the basis of their pathogenic features, IPEC strains are classically categorised into six major pathotypes: enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC) including the enterohaemorrhagic *E. coli* (EHEC) subset, enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Nataro and Kaper 1998). However the use of recent molecular approaches for the detection of strains belonging to some IPEC pathotypes has demonstrated that these strains can be further classified into subtypes within a particular pathotype (Croxen *et al.* 2013). For example, it has been found that EPEC is classified into typical (tEPEC) and atypical (aEPEC) subtypes based on the presence or absence of the *E. coli* adherence factor plasmid (pEAF) (Trabulsi *et al.* 2002). Additionally, recent advances in whole-genome sequencing have led to an increasing knowledge of the plasticity of *E. coli* genome, with the classification of some IPEC strains now becoming even more complicated (Croxen *et al.* 2013). For example,

previous studies published prior to 2011 showed that the Shiga toxin-producing EAEC was associated with causing a few cases of bloody diarrhoea and haemolytic uremic syndrome (HUS) (Morabito *et al.* 1998, Iyoda *et al.* 2000). However characterisation of EAEC O104:H4, the causative agent of the 2011 outbreak in Germany, identified major virulence traits of different IPEC pathotypes, such as an aggregative adhesive phenotype *in vitro*, lack of the enterocyte effacement (LEE) pathogenicity island, and expression of a Shiga toxin (Bielaszewska *et al.* 2011). Therefore, these isolates can be considered a hybrid of both EAEC and EHEC, suggesting that the Shiga toxin-producing EAEC strain O104:H4 associated with the 2011 German outbreak can be rather called enteroaggregative haemorrhagic *E. coli* (EAHEC) (Brzuszkiewicz *et al.* 2011).

Extraintestinal pathogenic *E. coli* (ExPEC) strains are part of the normal intestinal flora that asymptomatically colonise the gut of healthy individuals (Köhler and Dobrindt 2011). However when ExPEC strains get access to niches outside the gastrointestinal tract of the host, they have the capability to colonise these niches and subsequently cause a range of infections such as urinary tract infections (UTIs), neonatal meningitis, cellulitis, pneumonia, sepsis and bacteraemia (Russo and Johnson 2003, Köhler and Dobrindt 2011, Pitout 2012b). Despite their clinical importance, ExPEC strains have not been well categorised compared to IPEC strains (Johnson and Russo 2002a). However traditional designations for ExPEC strains can be used on the basis of to the type of infection they cause, and they include: uropathogenic *E. coli* (UPEC), sepsis-associated *E. coli* (SEPEC) and neonatal meningitis-associated *E. coli* (NMEC) (Johnson and Russo 2002a,

Johnson and Russo 2002b). Comparison of the pathogenic behaviours of the three main categories of *E. coli* is shown in Table 1.1.

Table 1.1: The pathogenic behaviours of the three main categories of *E. coli*, as obtained from Johnson and Russo (2002a).

	Clinical manifestation[1]		
***E. coli* category**	**Asymptomatic intestinal colonisation**	**Gastrointestinal infections**	**Extraintestinal infections**
Commensal	+++	-	+
IPEC	-	+++	-
ExPEC	+	-	+++

[1]Relative involvement of each *E. coli* category in the specific clinical manifestation is indicated on a semi-quantitative scale, from - (absent) to +++ (maximal).

1.3 The genetic structure and phylogenetic history of *E. coli*

It has been shown that the balance between recombination and mutation basically defines the population structure of bacteria (Tenaillon *et al.* 2010). Based on the level of recombination, bacterial population structure can either be clonal, when the recombination level is low, or panmitic, when the recombination level is high (Smith *et al.* 1993). Many previous studies showed that the population structure of *E. coli* is generally clonal. Initially, the use of serotyping to identify *E. coli* strains based on the polymorphism of their O (somatic), K (capsular) and H (flagellar) antigens showed the strong association of some serotypes with causing certain diseases, providing the first description of the clonal population structure of *E. coli* (Orskov and Orskov 1983, Orskov and Orskov 1992). Afterwards, the application of multilocus enzyme electrophoresis (MLEE), based on detecting the allelic variation in multiple chromosomal genes encoding bacterial enzymes by measuring the electrophoretic mobility patterns of

enzymes derived from different *E. coli* strains, demonstrated that there are only a few distinctive genotypes, despite the large global genetic diversity, and that clones isolated from geographically and temporally distinct hosts were identical (Selander and Levin 1980). These findings suggested that the rate of recombination in *E. coli* populations was low and therefore supported the previous description that the population structure of *E. coli* is clonal.

Early phylogenetic analysis of MLEE profiles of the *E. coli* Reference Collection (ECOR) showed that *E. coli* strains can be defined into six phylogenetic groups, designated A, B1, B2, C, D and E (Selander *et al.* 1987, Whittam *et al.* 1993). A subsequent MLEE analysis of similar data using 38 enzyme loci found that *E. coli* strains clustered into five phylogenetic groups: A, B1, B2, D and few unclassified lineages belonging to group E (Herzer *et al.* 1990). Commensal *E. coli* strains typically derive from phylogenetic groups A or B1 (Johnson *et al.* 2001a). Intestinal pathogenic *E. coli* strains derive from the phylogenetic groups A, B1, D or E (Pupo *et al.* 1997), while ExPEC strains derive from phylogenetic groups B2 or D (Boyd and Hartl 1998, Johnson and Kuskowski 2000, Johnson *et al.* 2001a).

Several subsequent studies demonstrated that the use of MLEE is not ideal for phylogenetic analysis since enzymes with little sequence similarity may show comparable electrophoretic mobility (Bisercić *et al.* 1991). Nucleotide and amino acid sequences can rather be used to provide a platform for more refined phylogenetic analysis because they are unlikely to suffer from convergence (Chaudhuri and Henderson 2012). In this regard, many

nucleotide sequence-based approaches were developed to resolve MLEE limitations. For example, Milkman and Crawford (1983) compared the nucleotide sequence of a 1648bp region of the *trp* operon, which controls the biosynthesis of tryptophan in the cell, from 12 *E. coli* K-12 strains. The authors found 11 of the 12 K-12 strains to be derived from the phylogenetic groups A and B1, with the other strain found to be more divergent since its *trp* sequence differed from that of *E. coli* K-12 at 44 positions and therefore placed it in group B2. This was interpreted as a possible occurrence of recombination events (Milkman and Crawford 1983).

Multilocus sequence typing (MLST) has been used to describe the clonal diversity and genetic structure of bacterial populations (Maiden *et al.* 1998). MLST data can also be used as an alternative to MLEE and other existing methods for determining the phylogenetic relationships among deep lineages, providing a complementary view of the population structure (Feil 2004). For example, Escobar-Páramo *et al.* (2004) investigated the phylogenetic distribution of a collection of 98 commensal and pathogenic *E. coli* strains from different pathotypes using an MLST scheme including 6 housekeeping genes (*trpA*, *trpB*, *pabB*, *putP*, *icd* and *polB*). The authors found the same phylogenetic groups described previously by MLEE analysis, however an additional group, designated group C (although unrelated to the group C from the original MLEE analysis), was identified between group D and groups A and B1. They also found some of the hyper-virulent *E. coli* pathotypes such as EHEC and ETEC to be restricted to specific phylogenetic groups, suggesting that a specific genetic background is essential for the expression and maintenance of genes associated with pathogenesis (Escobar-Páramo *et al.* 2004). However a different MLST scheme including

7 housekeeping genes: *adk*, *icd*, *fumC*, *recA*, *mdh*, *gyrB* and *purA* was used to analyse 462 *E. coli* isolates and found that the phylogenetic grouping of these isolates was different to that seen either in MLEE analysis (Herzer *et al.* 1990) or in the aforementioned MLST analysis (Escobar-Páramo *et al.* 2004). This finding suggested the presence of high levels of recombination in *E. coli*, and it was further supported by the identification of strains from the large MLST dataset that were found to be hybrids between groups A and B1 (AxB1), or with multiple sources of ancestry (ABD) (Wirth *et al.* 2006). The inconsistent findings obtained from using different MLST schemes questioned their suitability for phylogenetic studies. To overcome this problem, whole genome phylogenetic analysis (phylogenomics) of *E. coli* strains is now being used as an alternative of MLST, although its validity depends on the core genome disruption by homologous recombination between lineages (Chaudhuri and Henderson 2012). For example, Touchon *et al.* (2009) constructed a whole genome phylogeny for a number of *E. coli* genomes and found that they clustered into the same phylogenetic groups predicted by earlier MLEE (Herzer *et al.* 1990) and MLST studies (Escobar-Páramo *et al.* 2004). However using *Escherichia fergusonii* (*E. fergusonii*) to determine the root of the tree showed that it was placed within group D, splitting this group into two clades: D1 and D2 (Touchon *et al.* 2009). This was the first spilt in *E. coli* phylogenetic history, which might suggest the robustness of phylogenomics as tool for studying the phylogenetic structure of *E. coli*. The Phylogenetic relationships amongst selected *E. coli* strains, as determined by phylogenetic analysis of complete and draft genome sequences are shown in Figure 1.1.

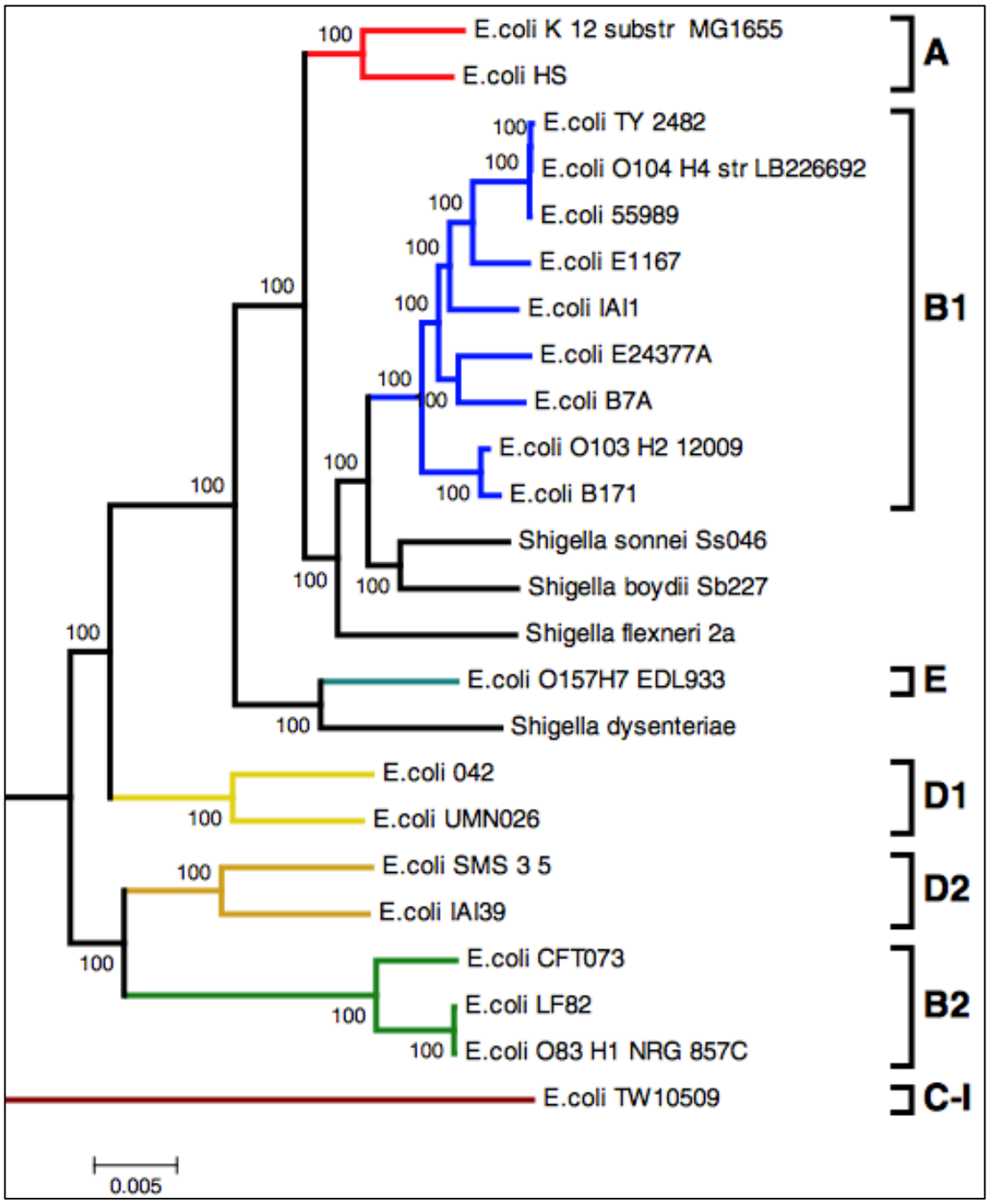


Figure 1.1: Phylogenetic relationships amongst selected *E. coli* strains, as determined by phylogenetic analysis of complete and draft genome sequences available in GenBank. The tree illustrates the major phylogenetic groups, and shows the close relationship between the German EHEC O104:H4 outbreak strains (TY2482, LB226692) and the EAEC strain 55989 [Taken from (Chaudhuri and Henderson 2012)].

1.4 Introduction to ExPEC

ExPEC strains are the major aetiological agents of a broad range of infections in the community and hospital settings (Rogers *et al.* 2011). They are also associated with causing community-wide outbreaks such as the

community-acquired UTIs and septicaemia outbreak caused by *E. coli* O15:K52:H1 in South London between 1986-1987 (Manges *et al.* 2008). ExPEC has now overtaken methicillin-resistant *Staphylococcus aurous* (MRSA) as the leading cause of healthcare-associated infections (HAIs) (HPA 2010). Therefore, infections due to ExPEC have a great impact not only on public health, but also on the global economy (Russo and Johnson 2003).

1.4.1 Overview of the medical and economic impact of major infections due to ExPEC

Urinary tract infections

Urinary tract infections (UTIs) are among the most common bacterial infectious diseases acquired in community and hospital settings (Foxman 2010). Worldwide approximately 150 million people are diagnosed with UTI each year, costing the global economy more than 6 billion US dollars (Akram *et al.* 2007). UTIs occur in all age groups and in both genders (McCormick *et al.* 1995, Litwin *et al.* 2005). However their incidence increases with age, and the annual incidence of UTI in the elderly population ranges from 10% in the community to as high as 30% of hospitalised patients (Cove Smith and Almond 2007). UTIs are more common in women than men (McCormick *et al.* 1995), with an estimated 33% of women suffering from a UTI by the age of 24 (Foxman 2002). UTIs can be clinically presented as one of three syndromes: cystitis, acute pyelonephritis or asymptomatic bacteriuria (Kumazawa and Matsumoto 1997). In addition, UTIs are divided into uncomplicated and complicated infections. Uncomplicated UTIs occur in otherwise healthy individuals who have normal

urinary tracts without any abnormalities or foreign bodies within the urinary tract (Hooton and Stamm 1997). However complicated UTIs occur in patients who have anatomical abnormalities in the urinary tract or are catheterised, which is frequently the case in the elderly population, and it has been shown that the presence of complicating factors in the infection is usually associated with high levels of antimicrobial resistance (Melekos and Naber 2000).

UPEC is the major causative agent of UTIs, accounting for as much as 90% of all UTIs among non-hospitalized patients and up to 50% of all nosocomial UTIs (Srinivasan *et al.* 2003, Tartof *et al.* 2005). Additionally, UPEC is associated with causing more than 85% of uncomplicated cystitis in premenopausal women (Stamm and Hooton 1993, Hooton and Stamm 1997), and the estimated number of cases of uncomplicated cystitis per year is more than 130 million worldwide costing the global economy more than 1 billion US dollars (Russo and Johnson 2003). UPEC also cause 90% of uncomplicated pyelonephritis in premenopausal women (Talan *et al.* 2000). In the US, the number of uncomplicated pyelonephritis cases per year is 250000 and the estimated annual healthcare costs associated with pyelonephritis due to UPEC are 175 million dollars (Russo and Johnson 2003).

Bloodstream infections

E. coli is the leading cause of a range of bloodstream infections (BSIs), such as bacteraemia, septicaemia and urosepsis, in nursing facilities (Mylotte *et al.* 2002), hospitalized individuals (Yardena *et al.* 2002) and children (Ron

2010). According to previous studies from all continents, *E. coli* comprises 17–37% of all clinically significant bacteria isolated from bloodstream infections (Russo and Johnson 2003). Recently, the number of reported bloodstream infections due to ExPEC has increased markedly (Rogers *et al.* 2011). Bloodstream infections usually result from principal extraintestinal infections, especially UTIs (Russo and Johnson 2003). Urosepsis, which is defined as a bacteraemia infection caused by an organism previously causing an infection in the urinary tract of a patient, represents a large number of the recently reported BSIs due to ExPEC (McNally *et al.* 2013a). Several studies have shown that bloodstream infections are major causes of mortality. For example, previous hospital-based reports demonstrated that the mortality rates in patients of bacteraemia due to ExPEC ranged from 5% to 30% (Laupland *et al.* 2008). Another study found that the mortality rates in elderly hospitalized patients from urosepsis can be as high as 33% (Tal *et al.* 2005). With respect to the impact of bloodstream infections on the economy, previous data showed that the estimated healthcare costs associated with sepsis are approximately 7.6 billion euros and 16 billion US dollars per year in Europe and US, respectively (Robson and Daniels 2008).

Neonatal Meningitis

ExPEC is one of the major causes of neonatal meningitis, accounting for 20-40% of the annual cases in the US (de Louvois 1994, Murphy 2000). However it has been reported that the excessive use of ampicillin to reduce the vertical transmission of group B *streptococcus*, the principal cause of neonatal meningitis, can lead to an increase in the prevalence of neonatal meningitis due to *E. coli* because of the increased resistance to ampicillin

among *E. coli* strains (Stoll *et al.* 2002). Meningitis due to *E. coli* can also be associated with high healthcare costs because of the prolonged hospitalization and the extensive rehabilitative care required for patients suffering from this infection (Russo and Johnson 2003).

1.4.2 Virulence factors of ExPEC

The virulence capacity of all pathogenic bacteria is generally determined by the possession of several specific virulence factors (VFs) (Dobrindt 2005). ExPEC is associated with the possession of a range of virulence factors including adhesins, toxins, lipopolysaccharides, polysaccharide capsules, proteases and iron-acquisition determinants. These VFs allow ExPEC to colonise key anatomical sites, evade and/or subvert host defences, increase acquisition of limiting nutrients such as iron, and stimulate a harmful host inflammatory response, cumulatively leading to infections outside the gastrointestinal tract (Rogers *et al.* 2011, Da Silva and Mendonça 2012). Many of these VFs are encoded on pathogenicity islands (PAIs) (Groisman and Ochman 1996). PAIs, which are defined as chromosomally inserted genetic elements with a distinct G + C content, harbour virulence-associated genes that are absent from commensal *E. coli* (Hacker et al. 1990). PAIs are typically acquired via horizontal gene transfer that might be a crucial step in the evolution of the pathogen (Hacker *et al.* 1990).

The expression of several VFs is more common among UPEC isolates compared to *E. coli* faecal isolates, and is more common among UPEC isolated from upper urinary tract (kidney and ureter) than UPEC isolated from lower urinary tract (bladder and urethra) (Johnson 1991). A recent study used RNA-sequencing (RNA-seq) and comparative transcriptional

analysis to identify novel fitness genes exclusively expressed by UPEC isolates during human infection, and found genes involved in ion transport, including copper efflux, nickel and potassium import systems, as key fitness factors of UPEC that can promote survival within the urinary tract of the host (Subashchandrabose *et al.* 2014). This suggests that the possession of these factors can be important for the ability of ExPEC to overcome the host defences and establish extraintestinal infections.

Phylogenetic analysis has shown that ExPEC strains belong to the phylogroups B2 and D (Boyd and Hartl 1998, Johnson and Kuskowski 2000, Johnson *et al.* 2001a), and it is suggested that group B2 *E. coli* strains are associated with the possession of more virulence factors than *E. coli* strains belonging to other phylogroups (Russo and Johnson 2000). Bacterial genomics has become a powerful tool for defining the genes encoding for factors that contribute to ExPEC fitness and virulence (Dobrindt 2005). In this regard, two previous studies demonstrated that a large fraction of genetic information of ExPEC, which has so far been thought to be virulence-associated, could also be found in commensal *E. coli* strains (Grozdanov *et al.* 2004, Hejnova *et al.* 2005). Therefore, some of these features, such as bacteriocins, proteases, iron-acquisition systems, fimbriae and other adhesins, can contribute to fitness and thus increasing adaptability, competitiveness and the effective colonisation of the host, rather than being specific virulence factors directly involved in infection (Wold *et al.* 1992, Herias *et al.* 1995, Herias *et al.* 1997). This demonstrates the thin line between virulence and fitness factors as well as the importance of having a definition of ExPEC virulence factors that directly contribute to its pathogenesis (Dobrindt 2005).

Adhesins

Adhesion to host structures is a crucial virulence factor of many bacterial pathogens (Klemm and Schembri 2000). Bacteria attach to host tissues to avoid being removed by the normal flow of body fluids such as blood and urine (Slavchev *et al.* 2013). Bacterial attachment is an essential first step in the colonisation of host mucosal surfaces and a precedent to invasive infection (Johnson 1991). In *E. coli*, attachment to uroepithelial cells is a characteristic by which uropathogenic *E. coli* strains are differentiated from faecal *E. coli* strains (Edén 1986). UPEC attachment to different host tissues is mediated by a number of fimbrial and afimbrial adhesins, each of which has its specific role in the pathogenesis (Slavchev *et al.* 2013). Among all adhesins expressed by UPEC, the type 1 fimbriae and P fimbriae are the most common (Johnson 1991).

Type 1 fimbriae were first recognised by their ability to mediate mannose-sensitive agglutination of some cell types such as erythrocytes, macrophages and yeast cells (Stentebjerg-Olesen *et al.* 1999). Type 1 fimbriae are found in more than 90% of all *E. coli*, including commensal and pathogenic strains (Arthur *et al.* 1989, Slavchev *et al.* 2013). A type 1 fimbria is a thin rod-shape surface organelle and comprises four adherence-associated fimbrial subunits: the major subunit FimA, and the minor components FimF, FimG, and FimH (Stentebjerg-Olesen *et al.* 1999). The export system of type 1 fimbriae involves the chaperone FimC and the outer membrane usher FimD (Krogfelt *et al.* 1990), while the expression of type 1 fimbriae is regulated by the FimB and FimE recombinases (Stentebjerg-Olesen *et al.* 1999). The structural components and the biosynthesis

machinery of type 1 fimbriae are encoded by the chromosomally located *fim* gene cluster (Klemm *et al.* 1985). Type 1 fimbriae mediate the interaction between UPEC and lower urinary tract of the host via the adhesive tip protein FimH. This protein binds to mannosylated uroplakin lining the bladder lumen leading to invasion of the superficial umbrella cells of bladder epithelium (Mulvey *et al.* 1998). Many studies have shown the important role of type I fimbriae in enhancing UPEC adherence, colonisation and invasion of host bladder epithelial cells. For example, a previous study on the ExPEC strain UTI89 found that type 1 fimbriae-mediated adherence to superficial bladder facet cells leads to invasion, rapid bacterial multiplication in the cytoplasm and formation of intracellular bacterial communities (IBCs) (Hannan *et al.* 2012). Additionally a transcriptome analysis of the UPEC strain CFT073 has suggested that the expression of type 1 fimbriae is required for the effective colonisation of the urinary tract of the host (Snyder *et al.* 2004). Another study has demonstrated that type 1 fimbriae may contribute to bacterial persistence in patients with UTI due to *E. coli* associated with the use of an indwelling bladder catheter (Mobley *et al.* 1987). Type I fimbriae can also mediate the inactivation of host innate immune pathways in the murine UTI model, and promote biofilm formation and host cell invasion (Nielubowicz and Mobley 2010).

P-fimbriae are the second most common adhesin in UPEC (Johnson 1991). They were originally identified by the ability to agglutinate human type O erythrocytes without inhibition by mannose, distinguishing it from type 1 fimbriae (Whitt and Salyers 2002). P-fimbriae are commonly associated with UTI and pyelonephritis (Katouli 2010), and they can be found in 36% of cystitis isolates and in more than 70% of pyelonephritis isolates (Johnson

1991). A P-fimbria is composed of many adherence-associated fimbrial subunits including the major subunit PapA and five other minor components PapC, PapH, PapE, PapF and PapG (Mu and Bullitt 2006). There are also some accessory Pap proteins that are not directly involved in adhesion such as the fimbrial assembly protein PapD (Whitt and Salyers 2002). The *pap* gene cluster encodes the structural and accessory components of P-fimbriae (Johnson 1991). P-fimbriae bind specifically to the α-D-Galp-(1-4)-β-D-Galp carbohydrate portion of P-blood group antigen-specific glycosphingolipids receptors, that are found in uroepithelial and renal epithelial cells, via the adhesive tip protein PapG (Korhonen *et al.* 1982). This binding can therefore mediate UPEC colonisation and infection of upper urinary tract of the host (Katouli 2010), and can also promote the activation of immune cell response leading to the development of severe local inflammation in the urinary tract (Johnson 1991).

ExPEC strains can also possess some other fimbrial adhesins such as S and F1C fimbriae as well as afimbrial adhesins such as the Dr-binding adhesin Afa/draBC (Bien *et al.* 2012). Various ExPEC-associated adhesins and their proposed role in pathogenesis are shown in Table 1.2.

Table 1.2: Summary of ExPEC-associated adhesins and their role in pathogenesis.

Adhesin	Activity/role in pathogenesis
S/F1C fimbriae (*sfa*/*foc*)	Bind to epithelial and endothelial cells of urinary tract and kidney, and associated with causing ascending UTIs, sepsis and meningitis (1)
Dr-adhesin (*afa*/*draBC*)	Bind to type IV collagen and decay-accelerating factor (DAF) in the kidney, and associated with ExPEC strains causing gestational pyelonephritis and recurring cystitis (2).
M-agglutinin (*bma*)	Agglutinates human erythrocytes carrying the M blood group antigen and associated with few pyelonephritogenic ExPEC strains (3).
N-acetyl-d glucosamine-specific fimbriae (*gaf*)	Responsible for the glucosamine-sensitive hemagglutination of endo-β-galactosidase-treated human erythrocytes, and it is associated with few pyelonephritogenic ExPEC strains (4).
Enterobactin receptor/adhesin (*iha*)	Contributes to the establishment of UTIs in mouse model by enhancing the adherence to uroepithelial cells (5).

Summarised from: (1): (Mulvey 2002, Bien *et al.* 2012); (2): (Bien *et al.* 2012); (3): (Rhen *et al.* 1986); (4): (Saarela *et al.* 1995) and (5): (Johnson *et al.* 2000, Johnson *et al.* 2005).

Toxins

ExPEC strains are associated with the secretion of several bacterial toxins, such as α-haemolysin (HlyA), the cytotoxic necrotising factor 1 (CNF1) and the secreted autotransporter toxin (SAT). These toxins can play a major role in the virulence of ExPEC. The HlyA is a pore-forming toxin that has a major role in the pathogenesis of ExPEC (Ulett *et al.* 2013). This toxin is encoded by the chromosomally located *hlyA* gene (Welch 1991), and found

to be mainly possessed by ExPEC strains associated with causing upper UTIs (Johnson 1991). A previous study showed that 50% of *E. coli* strains that are associated with causing pyelonephritis produce HlyA toxin (Bien *et al.* 2012). Another study found that HlyA mediates the lysis of host erythrocytes and renal epithelial cells, promoting the release of nutrients such as iron that can be utilised by UPEC for growth and/or survival (Ulett *et al.* 2013). Additionally, HlyA can also induce apoptosis of target host cells such as neutrophils and renal cells, and can stimulate the exfoliation of bladder epithelial cells (Russo *et al.* 2005, Chen *et al.* 2006a, Smith *et al.* 2006).

The CNF1 is a virulence factor produced by approximately 30% of *E. coli* strains that cause pyelonephritis, and it may play a major role in the UPEC invasion of renal cells (Bien *et al.* 2012). CNF1 toxin is chromosomally encoded by *cnf1* (Johnson 1991). A previous study showed that CNF1 stimulates actin stress fibres formation and membrane ruffle formation in a Rho GTPase-dependent manner, which can therefore facilitate the entry of *E. coli* into the cells (Bien *et al.* 2012). Another study found that CNF1 toxin facilitates the dissemination of UPEC by promoting apoptosis of bladder epithelial cells and inhibiting the phagocytic and chemotactic activities of neutrophils (Mills *et al.* 2000, Davis *et al.* 2006).

Secreted autotransporter toxin (SAT) is a toxin produced mainly by pyelonephrogenic *E. coli* strains (Bien *et al.* 2012). Many studies have shown that this toxin has a proposed role in pathogenesis of UTIs (Guyer *et al.* 2000, Guyer *et al.* 2002). For instance, a previous study found that

SAT could induce cytopathic effects on bladder and kidney cells and elicit glomerular damage in a mouse UTI model (Guyer *et al.* 2002).

Iron-acquisition determinants

Iron is an essential nutrient for most microorganisms and is required for many biological processes such as energy generation, oxygen transport and DNA replication (Skaar 2010). In the mammalian hosts, the availability of free extracellular iron is limited because it is tightly bound to proteins such as transferrin in the serum and lactoferrin on mucosal surfaces (Lewis 2010). Iron is primarily intracellular and can be found as heme and heme proteins such as haemoglobin, iron-sulfur proteins or the iron-storage protein ferritin (Katouli 2010, Lewis 2010). In order to acquire iron from the host, ExPEC strains use a variety of iron uptake mechanisms such as the synthesis and transport of low-molecular-weight iron chelators, called siderophores (Neilands 1976, Crosa 1989). Siderophores are important not only for scavenging iron, but also for competing with host defences to release iron from transferrin and lactoferrin (Litwin and Calderwood 1993). ExPEC strains possess several iron uptake systems such as the yersiniabactin siderophore system (irp-fyuA), the salmochelin siderophore system (iroN) and the aerobactin siderophore system (iuc-iutA) (Johnson *et al.* 2007). A previous study demonstrated that yersiniabactin is involved in the establishment of UTI as it plays a major role in iron uptake *in vivo* and promotes biofilm formation in human urine (Hancock *et al.* 2008, Garcia *et al.* 2011). Additionally, it has been shown that salmochelin is important in virulence of UPEC strain CP9 and NMEC strain S88 (Russo *et al.* 2002, Nègre *et al.* 2004, Feldmann *et al.* 2007, Peigne *et al.* 2009). It has also been found that the *iroN* gene is overexpressed in intracellular bacterial

communities (IBCs) found in bladder epithelium cells infected with UPEC strain UTI89 in a mouse UTI model (Reigstad *et al.* 2007). The aerobactin siderophore system contributes to iron uptake *in vivo* in a UTI model and promotes bacterial growth in conditions with low iron concentrations (Garcia *et al.* 2011).

Flagella

Flagella are filamentous organelles that are responsible for bacterial motility (Terashima *et al.* 2008). Additionally, flagella can be involved in biofilm formation, protein export and adhesion to host cells (Haiko and Westerlund-Wikström 2013). The flagellum comprises three main components: the filament, the hook, and the basal structure (Terashima *et al.* 2008). In *E. coli*, more than 50 genes are involved in the flagellum formation and regulation (Liu and Ochman 2007). These genes are clustered at three regions of the chromosome and grouped into three different classes: the single class I master operon that consists of *flhDC* genes responsible for transcriptional regulation of flagellar genes, the class II genes responsible for the synthesis of the flagellar basal body and hook (*flgMNBCDEFGHIJKL*, *flhBAE*, and *fliAZYDSTEFGHIJKLMNOPQR*), and the class III genes involved in flagellin synthesis (*fliC*) and motility (*motAB*, *cheAW*, and *cheRBYZ*) (Parthasarathy *et al.* 2007). Although flagella are common to commensal *E. coli* strains such as the K-12 MG1655 strain (Gauger *et al.* 2007), several studies have demonstrated the important role of flagella in the virulence of ExPEC. It has been shown that flagella can allow *E. coli* to ascend from the lower urinary tract to cause pyelonephritis, and that flagellated UPEC can cause 70-90% of all UTIs (Bien *et al.* 2012). Another study on pyelonephritogenic *E. coli* strains has demonstrated that flagellin can

mediate their invasion of renal collecting duct (Pichon *et al.* 2009). Additionally, flagella can be an important factor in the pathogenesis of meningitis caused by *E. coli* O18:K1:H7 strain (Parthasarathy *et al.* 2007).

Surface polysaccharide coatings

Bacterial surface polysaccharides such as lipopolysaccharides (LPS), polysaccharide capsules, and exopolysaccharide colanic acid are important virulence factors of ExPEC strains (Kusecek *et al.* 1984, Russo *et al.* 1996, Miajlovic and Smith 2014). Several reports have shown the role of these surface polysaccharides in ExPEC pathogenesis. For example a previous study found that LPS is important in the stimulation of a proinflammatory response in uncomplicated UTIs, however its role in mediating acute ascending UTIs is unclear (Bien *et al.* 2012). Additionally it has been shown that mutations affecting LPS synthesis might affect *E. coli* ability to adhere to abiotic surfaces, suggesting a role of LPS in adhesion processes (Genevaux *et al.* 1999). It has also been proposed that polysaccharide capsules provide protection against phagocytosis (Howard and Glynn 1971, Weiss *et al.* 1982, Sarkar *et al.* 2014), and that some ExPEC capsule types may be protective against the bactericidal effect of human serum (Leying *et al.* 1990, Buckles *et al.* 2009). Another study found a high overexpression of genes responsible for the capsular polysaccharide and LPS biosynthesis in the UPEC strain CFT073 during experimental UTI, which suggests the important role of these surface coatings in ExPEC pathogenesis (Snyder *et al.* 2004). Additionally, several studies have recently shown that exopolysaccharide colanic acid can contribute to *E. coli* survival in human serum (Phan *et al.* 2013, Miajlovic *et al.* 2014), and can also be implicated in fitness of an ExPEC strain in a mouse bacteraemia model

(Subashchandrabose *et al.* 2013).

1.4.3 Antimicrobial resistance of ExPEC

Over the past decade, the antimicrobial resistance of ExPEC to many front-line antibiotics, such as cephalosporins, fluoroquinolones, and trimethoprim–sulfamethoxazole, has increasingly been reported (Gupta *et al.* 1999). Several studies during the 2000s across Europe, North and South America demonstrated that resistance to ciprofloxacin and trimethoprim–sulfamethoxazole was observed in as many as 20-45% of ExPEC isolates (Foxman 2010). More recently the spectrum of antimicrobial resistance of ExPEC has also increased to include additional antimicrobial agents due to the emergence of extended-spectrum β-lactamases (ESBLs) producing *E. coli* (Johnson *et al.* 2010b). ESBLs are a group of enzymes that have the ability to hydrolyse the oxyimino-cephalosporins such as cefotaxime and the oxyimino-monobactams such as aztreonam and therefore render them inactive (Peirano and Pitout 2010). ESBLs were initially detected in 1985 and they comprise many plasmid-mediated derivatives such as TEM-1, TEM-2 and SHV-1 (Nicolas-Chanoine *et al.* 2008). Since 2000, a new group of ESBLs, called CTX-M (i.e. 'active on Cefo<u>T</u>a<u>X</u>ime, first isolated in <u>M</u>unich'), has emerged (Peirano and Pitout 2010). This group of ESBLs have shown an extensive pattern of antimicrobial resistance to many antibiotics, including β-lactam agents such as penicillin, cephalosporins (expect cephamycins) and monobactams (Rogers *et al.* 2011). In addition, CTX-M-producing *E. coli* strains are often associated with co-resistance to other antimicrobial agents such as aminoglycosides and fluoroquinolones (Rogers *et al.* 2011). CTX-M ESBLs are classified according to amino acid sequence similarity, into five subgroups (1-5). The CTX-M-15 enzyme, which belongs

to CTX-M subgroup 1, is the most common ESBL type globally and it is commonly found in *E. coli* strains (Dhanji *et al.* 2011). CTX-M-producing *E. coli* clones have recently emerged as important pathogens with a broad dissemination, posing a threat to public health worldwide (Rogers *et al.* 2011). Consequently the increasing prevalence of antimicrobial resistance of *E. coli* has limited the therapeutic options prescribed frequently by health care centres (Leflon-Guibout *et al.* 2008). Nowadays, few antimicrobial agents such as nitrofurantoin, pivmecillinam and carbapenems can be used to treat infections due to ExPEC (Vigil *et al.* 2010, Nicolas-Chanoine *et al.* 2014).

1.5 Microbial typing methods for the identification of pathogenic *E. coli* clones

Microbial typing is used for differentiating epidemiologically related from unrelated isolates within a bacterial species (Hallin *et al.* 2012). It can also be used for the identification of emerging bacterial strains or clones belonging to a same species, for determining bacterial population structure and genetic diversity, and for elucidating the source of transmission of bacteria associated with causing outbreaks (Van Belkum *et al.* 2007). Microbial typing methods may be classified into either phenotypic or genotypic methods. These methods vary greatly in terms of their discriminatory power, reproducibility, cost and time consumption (Van Belkum *et al.* 2001, Johnson and Russo 2005, Soll *et al.* 2006).

1.5.1 Phenotypic methods

Phenotypic methods, such as biotyping, antibiogram-based typing, phage typing and serotyping, are used to qualitatively and quantitatively assess the products of gene expression in order to differentiate bacterial strains to

a species or even to a subspecies level (Van Belkum *et al.* 2001, Van Belkum *et al.* 2007). Although the implementation of these methods is relatively easy and inexpensive, most of them are known of their poor discriminatory power and therefore they are considered unreliable for epidemiological studies (Hallin *et al.* 2012).

1.5.2 Genotypic methods

Genotypic methods, such as pulsed-field gel electrophoresis (PFGE) (Goering 2010), multilocus sequence typing (MLST) (Maiden *et al.* 1998) and ribotyping, are based on identifying the genetic variation between strains (Van Belkum *et al.* 2007). They include techniques that are used to detect the genetic composition (presence or absence of certain genes or plasmids), overall structure (restriction endonuclease profiles) and nucleotide-based sequence similarity of a number of housekeeping genes (Van Belkum *et al.* 2007). Several molecular typing methods can be used for describing pathogenic bacterial clones, which are defined as a group of bacterial isolates showing a high degree of similarity, based on near-identity of multilocus enzyme profiles and multilocus sequence types (Rogers *et al.* 2011). Among these methods, MLST is the "gold standard" for identification, and it requires the sequencing of pre-specified internal regions of 6-8, typically 7, highly conserved housekeeping genes, allowing comparison of the nucleotide sequences within publicly accessible databases (Maiden *et al.* 1998, Rogers *et al.* 2011). In 2008, MLST was used for the identification of globally disseminated ESBL-producing ExPEC clones, and was a useful tool in describing the extraintestinal pathogenic CTX-M producing *E. coli* sequence type 131 (*E. coli* ST131) clone (Woodford *et al.* 2009).

1.6 *E. coli* ST131

Although *E. coli* ST131 was first identified in 2008 using MLST (Woodford *et al.* 2009), analysis of the *E. coli* MLST database demonstrates that *E. coli* ST131 isolates have been found in circulation since 1992 (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/GetTableInfo_html). However it is only recently that this pathogen has emerged as a major rapidly-disseminated pandemic *E. coli* clone, responsible for widespread community-onset infections (Rogers *et al.* 2011).

1.6.1 Bacterial characteristics of *E. coli* ST131

1.6.1.1 Serotyping and phylogenetic group

Previous studies revealed that *E. coli* ST131 belongs to the serotype O25:H4, with a specific O25 type called O25b (Nicolas-Chanoine *et al.* 2014). However many studies from different countries have identified some *E. coli* ST131 isolates of serotype O16:H5 (Platell *et al.* 2011, Matsumura *et al.* 2012, Dahbi *et al.* 2013, Kudinha *et al.* 2013, Olesen *et al.* 2013, Blanc *et al.* 2014, Habeeb *et al.* 2014), while others demonstrated that some *E. coli* ST131 strains were nontypeable for O or H antigens (Suzuki *et al.* 2009, Dahbi *et al.* 2013). Phylogenetic analysis of *E. coli* ST131 has shown that it belongs to the hyper-virulent phylogenetic group B2 (Leflon-Guibout *et al.* 2008, Nicolas-Chanoine *et al.* 2008).

1.6.1.2 *fimH* subtyping of *E. coli* ST131

Like most *E. coli*, all *E. coli* ST131 strains carry the *fimH* gene (Johnson *et al.* 2001b) that shows high level of allelic diversity (Weissman *et al.* 2006, Tartof *et al.* 2007, Díaz *et al.* 2010). The *fimH*-typing region harbours a highly diverse set of alleles that can be considered phylogenetically

restricted (Weissman *et al.* 2012). A previous study investigated the clonal structure of 352 historical and recent ST131 isolates (1967-2009), and identified that these isolates belonged to seven *fimH*-based clonal lineages (*fimH15*, *fimH22*, *fimH27*, *fimH30*, *fimH35*, *fimH41*, *fimH94*) (Johnson *et al.* 2013). It also found that *fimH30* was the most common among all *fimH* types (67%), followed by *fimH22* (21%) (Johnson *et al.* 2013).

1.6.1.3 Antimicrobial resistance of *E. coli* ST131

E. coli ST131 is commonly associated with carrying the $bla_{CTX\text{-}M\text{-}15}$ gene, encoding the CTX-M-15 ESBL enzyme, on the IncFII plasmids (Coque *et al.* 2008). These plasmids are also capable of harbouring a variety of other antimicrobial resistance determinants such as OXA-1, TEM-1, SHV-12, CMY-2 β-lactamases and the aminoglycosides/fluoroquinolone acetyltransferase AAC (6')-Ib-cr (Woodford *et al.* 2009, Rogers *et al.* 2011). However the $bla_{CTX\text{-}M\text{-}15}$ gene in *E. coli* ST131 can be carried on other plasmid types such as FIA, FIB and FII (Rogers *et al.* 2011). The antimicrobial resistance of *E. coli* ST131 was first recognised amongst CTX-M-15 producing *E. coli* strains (Nicolas-Chanoine *et al.* 2008). Further studies on this resistance revealed that *E. coli* ST131 is highly prevalent amongst fluoroquinolone-resistant, non-ESBL-producing *E. coli* strains (Cagnacci *et al.* 2008, Jones *et al.* 2008, Johnson *et al.* 2009a, Johnson *et al.* 2010c, Uchida *et al.* 2010). This suggests that CTX-M enzymes might have been acquired by *E. coli* ST131 strains that were already resistant to fluoroquinolones. Moreover, previous studies demonstrated that *E. coli* ST131 were more commonly resistant to amikacin, amoxicillin-clavulanic acid, and piperacillin-tazobactam than non-ST131 (Nicolas-Chanoine *et al.*

2014). However not all ST131 isolates are associated with antimicrobial resistance and some studies demonstrated the existence of antimicrobial-susceptible ST131 strains. For example, a previous study found that *E. coli* ST131 isolates were prevalent amongst extended-spectrum cephalosporins (ESC) and fluoroquinolone-susceptible urine isolates (Golding *et al.* 2012).

The increasing spectrum of antimicrobial resistance of *E. coli* ST131 might be attributed to its ability to exchange diverse antimicrobial resistance genes and plasmids encoding these genes (Johnson *et al.* 2010b). Recent studies have discussed the evolution of *E. coli* ST131 and the role played either by its clonal spread or transferable plasmids and other resistance determinants, such as ESBL and fluoroquinolone resistance genes, in the worldwide dissemination of *E. coli* ST131 (Rogers *et al.* 2011). Lee *et al.* (2010) found no correlation between the acquisition of transferable resistance determinants and the global dissemination of *E. coli* ST131. However the spread of fluoroquinolone-resistant isolates, which then acquire a CTX-M gene, or probably the simultaneous spread of clonal organisms and genes, can provide potential explanations for the accelerated dissemination of *E. coli* ST131 (Oteo *et al.* 2009, Uchida *et al.* 2010).

1.6.1.4 Phylogeny of *E. coli* ST131 clinical isolates

A previous study investigated the population structure of fluoroquinolone resistance in 352 ST131 isolates including historical (1967-2009) and recent (2010-2011) ST131 isolates, using PFGE profiling and the sequencing of *fimH*, *gyrA* and *parC* genes (Johnson *et al.* 2013). Comparing the events affecting the *gyrA* and *parC* genes with those of 737 non-ST131

isolates showed the presence of 185 unique pulsotypes among the 352 ST131 isolates. The analysis also identified seven *fimH*-based clonal lineages within ST131 isolates (*H15*, *H22*, *H27*, *H30*, *H35*, *H41* and *H94*), and found *H30* to be the most common *fimH* allele (Johnson *et al.* 2013). Determining the association between ST131 *H* subclones and fluoroquinolone resistance showed that only fluoroquinolone-susceptible ST131 *H* subclones, particularly *H22*, were identified at the start of study period, 1967-1999. Fluoroquinolone-resistant ST131 isolates were first detected between 2000-2005, and nearly all of these isolates belonged to the *H30* subclone. Additionally, it was found that the *H30* subclone was strongly associated with fluoroquinolone resistance in the historical ST131 isolates, and also with the ESBL $bla_{CTX\text{-}M\text{-}15}$ production (Johnson *et al.* 2013). The history of single nucleotide polymorphism (SNP) and mutations of *gyrA* and *parC* genes leading to amino acid substitutions in ST131 and non-ST131 isolates showed an association between the *H30* subclone and a characteristic *gyrA* and *parC* allele combination arising from a single strain as little as 10 years ago (Johnson *et al.* 2013). Recently, a whole genome SNP-based study used 105 ST131 isolates to determine the phylogeny of ST131, and identified four large recombinant regions acquired by horizontal gene transfer (Price *et al.* 2013). The exclusion of SNPs from these four regions uncovered the presence of different clusters of O-type strains harbouring the *fimH30* allele and *gyrA* and *parC* alleles. Strains carrying the *fimH* allele clustered as a single low-diversity clade, designated *H30*. This clade contained 95% of all fluoroquinolone-resistant ST131 isolates. Additionally, all the CTX-M-15 ESBL producing ST131 isolates belonged to a single distinct subclade within the *H30* clade. To further resolve the

evolutionary history of the *H30* subclone, a new phylogenetic tree was constructed based on genomic analyses of *H30* strains and the ST131 NA114 reference genome. The analysis found that the *fimH30* allele was acquired before fluoroquinolone resistance, and that this acquisition was followed by the extensive clonal expansion of fluoroquinolone-resistant *H30* strains. The fluoroquinolone-resistant subclone within the *H30* lineage was designated *H30-R*. Additionally, the analysis demonstrated that 91% of the CTX-M-15 producing ST131 isolates constituted a distinct single ancestor subclone within *H30-R*, and it was therefore designated *H30-Rx* (Price *et al.* 2013). Taken together, these findings support the hypothesis that the spread of CTX-M-15 and fluoroquinolone resistance in ST131 depends mainly on clonal expansion. The phylogeny of ST131 including the three nested ST131 subclones: *H*30, *H*30-R, *H*30-Rx is shown in Figure 1.2.

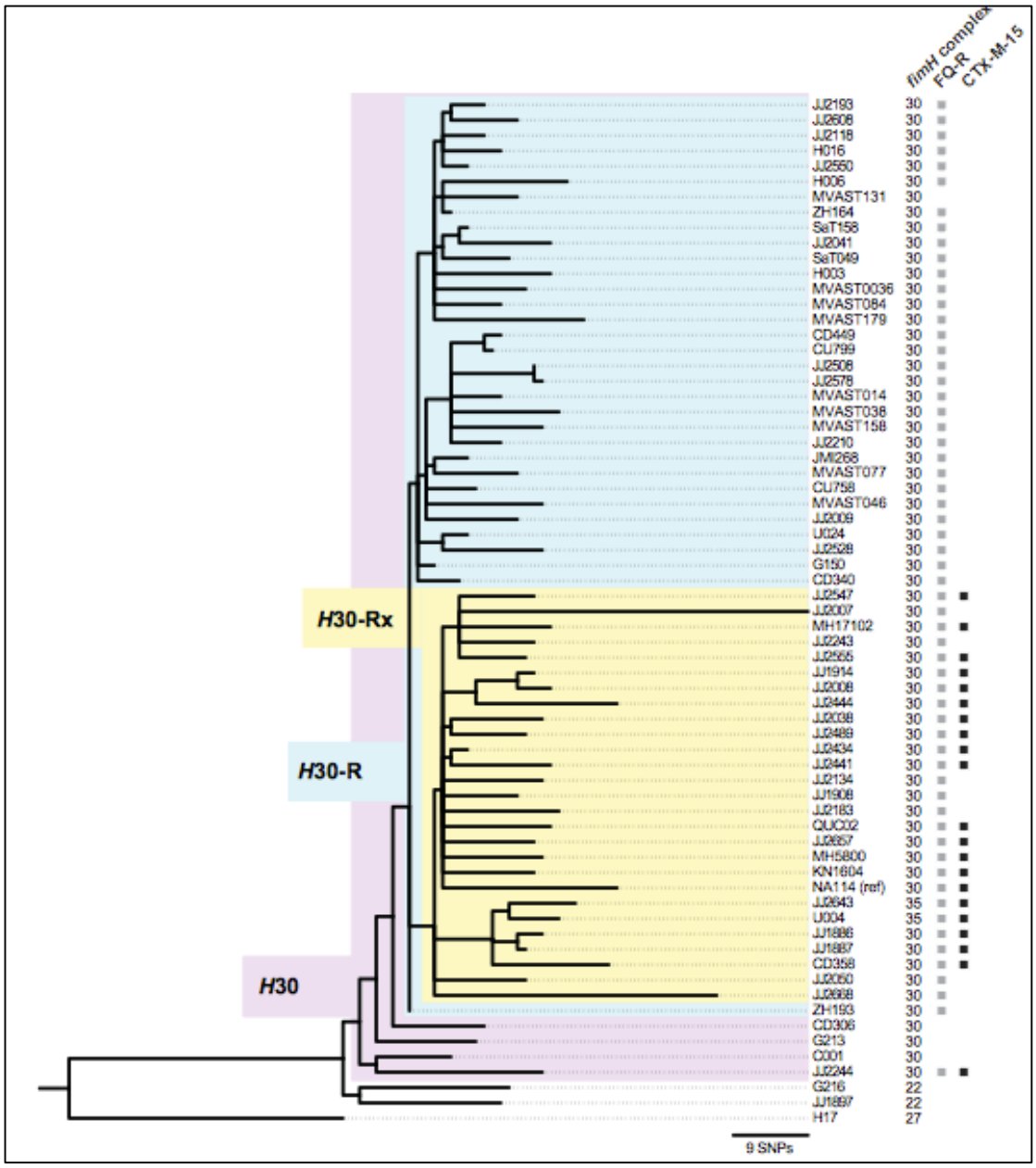


Figure 1.2: High-resolution phylogenetic analysis of the emergence of fluoroquinolone-resistance and CTX-M-15 ESBL production. Approximately 51.8% of the reference genome was shared among all isolates and sequenced at ≥10 coverage. Analysis of these shared genomic regions revealed 72 parsimony-informative SNPs and 771 total SNPs from the core genome (excluding horizontally acquired regions) that were used to construct the phylogeny presented here. Homoplasy index (HI) = 0.000. The coloured blocks highlight the three nested ST131 subclones, *H*30 (purple), *H*30-R (blue), *H*30-Rx (yellow) [Taken from (Price *et al.* 2013)].

1.6.2 Pathogenic characteristics of *E. coli* ST131

1.6.2.1 Scale of infection

Many studies have shown that *E. coli* ST131 strains are a major cause of community- and hospital-acquired UTIs (cystitis and pyelonephritis) and bacteraemia worldwide (Adams-Sapper *et al.* 2012, Gibreel *et al.* 2012b, Williamson *et al.* 2013). Additionally, *E. coli* ST131 strains have been associated with causing a range of infections such as intra-abdominal and soft tissue infections (López-Cerero *et al.* 2013), meningitis (Assimacopoulos *et al.* 2012), osteoarticular infection (Johnson *et al.* 2010a), myositis (Vigil *et al.* 2010) and septic shock (Ender *et al.* 2009). This scale of infection in addition to the multidrug resistance of *E. coli* ST131 has increased the concern about this clone globally.

1.6.2.2 Transmissibility

The transmission of *E. coli* ST131 has been reported between members of the same household and also between family members and pets (particularly dogs and cats) (Johnson *et al.* 2009b). A previous study documented the transmission of a CTX-M-15 producing ST131 isolate, which was resistant to gentamicin, trimethoprim-sulfamethoxazole and fluoroquinolones, between a father and his daughter (Ender *et al.* 2009). Another study reported a fatal case of urosepsis including community-associated intrafamilial spread of CTX-M-15 producing, fluoroquinolone-resistant ST131 strain (Owens *et al.* 2011). Previous molecular epidemiological analysis found a multidrug-resistant (MDR) *E. coli* strain from a human-associated pulsotype within ST131 colonising three of five dogs and cats within a household suggesting that companion animals might

represent a reservoir of ST131 (Johnson *et al.* 2009b). The transmission of ST131 between human and animal members, probably by direct host-to-host transmission, might considerably contribute to the community-wide dissemination of ST131 (Nicolas-Chanoine *et al.* 2014).

1.6.2.3 Virulence potential of *E. coli* ST131

As mentioned earlier, it has been shown that the ability of ExPEC to colonise and cause infection in the urinary tract results from the cumulative action of specific virulence factors which mediate adhesion, aid subversion of host immune defences, stimulate harmful host inflammatory responses and increase iron acquisition (Rogers *et al.* 2011). It has been widely thought that bacteria exhibiting high levels of antimicrobial resistance do so at the expense of a fitness advantage, which results in decreased pathogenesis (Foxman 2010). However initial studies on the virulence potential of *E. coli* ST131 strains showed high levels of virulence-associated gene (VAG) carriage (Coelho *et al.* 2010), and that *E. coli* ST131 strains have been responsible for causing large-scale disease outbreaks (Lau *et al.* 2008a, Coelho *et al.* 2010, Johnson *et al.* 2010b). Coelho *et al.* (2010) concluded that ST131 isolates were significantly associated with eleven VAGs (*papG III*, *afaFM955459*, *cnf1*, *sat*, *hlyA*, *kpsM II-K2*, *kpsM II-K5*, *traT*, *ibeA*, *malX* and *usp*), and they demonstrated that the worldwide dissemination of ST131 clone might be attributed to this high VAG carriage of ST131. Recently, although many comparative studies on the virulence capacity of strains belonging to different ExPEC STs found a number of VFs that are commonly possessed by ST131 strains (Table 1.3), these studies demonstrated that strains within the ST131 complex did not seem to possess any specific set of VAGs compared to other ExPEC STs. For

example, a previous study by our research group found that ST131 isolates did not exhibit higher VAG carriage than other ExPEC STs such as ST73 and ST127 (Croxall *et al.* 2011a). It also showed that some VAGs such as, *papG III*, *sfa/focDE*, *focG* and *cnf1*, were less prevalent in ST131 isolates compared to other ExPEC STs (Croxall *et al.* 2011a). Another study found that ST131 had moderate VAG profiles compared to other successful ExPEC STs such as ST69 and ST127 (Gibreel *et al.* 2012b). These studies suggest that the success of ST131 cannot be ascribed to the possession of any unique VAGs or the possession of any VAGs to a greater extent than in other ExPEC strains.

Table 1.3: List of VAGs that are commonly associated with ST131.

VAG	**Function**
sat	Secreted autotransporter toxin
fimH	Type 1 fimbriae adhesin
fyuA	Yersiniabactin siderophore receptor
kpsM II	Group 2 capsule synthesis
usp	Uropathogen specific protein
malX	Pathogenicity island marker
iha	Enterobactin receptor/adhesion
ompT	Outer membrane protein
iucD	Aerobactin
iutA	Aerobactin receptor
traT	Serum-resistance associated

Summarised from: (Karisik *et al.* 2008, Nicolas-Chanoine *et al.* 2008, Johnson *et al.* 2010b, Mora *et al.* 2011, Johnson *et al.* 2012, Lavigne *et al.* 2012, Van der Bij *et al.* 2012).

A recent whole genome-based study examined the genetic architecture of isolates belonging to the *E. coli* ST131 H30Rx clade (Petty *et al.* 2014), paying particular attention to virulence associated genes of ExPEC and to mobile genetic elements not found in non-ST131 ExPEC. This study suggested the absence of an ST131 specific virulence gene repertoire, though did highlight the unique nature of the secondary flagellar cluster (Flag-2) that had been previously identified in *E. coli* ST131 genomes (Clark *et al.* 2012). Additionally, the analysis also demonstrated the role of intra-ST131 recombination in shaping the lineage. It also identified a recombinant fragment common across ST131 within the group II capsule synthesis locus (Petty *et al.* 2014), with several variant region 2 gene clusters observed between region 1 (*kpsFEDUCS*) and region 3 (*kpsTM*) of ST131 H30Rx genomes. This variability in the capsule locus of ST131 H30Rx strains was consistent with the detectable diversity in K-antigen serotype of members of ST131 H30Rx strains (Olesen *et al.* 2013).

1.6.2.4 Metabolic potential of ST131

It has recently been suggested that high metabolic capacity enhances bacterial fitness and therefore contributes to pathogenesis (Le Bouguénec and Schouler 2011). With regard to the association between metabolic potential and increased *E. coli* ST131 fitness, a recent study tested the metabolic activity of a collection of 300 UPEC isolates representing the major STs using 47 biochemical tests (Gibreel *et al.* 2012a). The authors found that *E. coli* ST131 isolates have higher metabolic potential and biochemical profiles compared to other ExPEC ST isolates, which may contribute to the fitness of *E. coli* ST131. They also found a correlation

between metabolic activity and antibiotic susceptibility profiles, with resistant isolates showing the highest metabolic potential (Gibreel *et al.* 2012a). Another study tested the ability of one *E. coli* ST131 strain to colonise the intestinal tract by measuring its maximal growth rate (MGR) using three different culture conditions: a rich medium and two minimal media containing either gluconate or glucose (Vimont *et al.* 2012). By comparing the metabolic activity of the ST131 strain with three other *E. coli* strains: the commensal *E. coli* strain K-12 MG1655 and the ExPEC CFT073 and HT7 strains, the study found that the ST131 strain had a significantly higher MGR than the non-ST131 strains (Vimont *et al.* 2012).

1.6.2.5 Adhesion and invasion capacity of ST131

A previous study examined the similarity and divergence of a collection of adherent invasive *E. coli* (AIEC) and ExPEC strains, in terms of their genetic and phenotypic features, and identified two intestinal AIEC strains and one extraintestinal AIEC strain belonging to the O25:H4 ST131 clone (Martinez-Medina *et al.* 2009). It also found that the extraintestinal AIEC ST131 strain had a similar AIEC adhesion, invasion of intestinal epithelial cells and intra-macrophage replication phenotype, and that this strain showed distinct virulence and macrorestriction profiles compared to the other two intestinal AIEC ST131 strains (Martinez-Medina *et al.* 2009). Another study compared the capacity of four ExPEC, two ST131 and two non-ST131 strains, to adhere to human intestinal cell lines (Peirano *et al.* 2013), and found that the tested ST131 strains had lower adhesion capacity compared to the non-ST131 strains. However carrying out these experiments in the presence of mannose, which inhibits adhesion mediated by type 1 fimbriae, could be the reason behind the weak adhesion capacity of ST131 strains (Peirano *et*

al. 2013). These conflicting data highlight the importance of testing and comparing the adherence and invasion capacity between ST131 and non-ST131 strains.

1.7 Introduction to the project

Since the emergence of *E. coli* ST131 and in an attempt to identify the key factors that might contribute to its success, this clone has been the subject of numerous studies that have explored its antimicrobial resistance, virulence potential, metabolic capacity and the molecular epidemiological aspects of this clone. Early reports found an association between the rapid emergence and dissemination of *E. coli* ST131 and its increased antimicrobial resistance, however this resistance alone cannot provide an explanation for the ST131 dominance amongst other MDR uropathogens circulating worldwide in healthcare and community settings. Subsequently many comparative studies investigated the prevalence and virulence potential of ST131 and other successful ExPEC STs such as ST73, ST69, ST95 and ST127. Previous work by our research group showed that *E. coli* ST131 was the dominant strain type within a collection of elderly urine samples, that the *E. coli* ST131 was responsible for the high levels of antimicrobial resistance reported in the collection, and that there was variation in VAG profiles between strains, with no specific VAG profile associated with ST131 strains (Croxall *et al.* 2011a). The variable carriage of key virulence factors of *E. coli* ST131 suggests the presence of additional unidentified factors that can contribute to its fitness and therefore provide an explanation for what makes *E. coli* ST131 such a successful ExPEC clone. On the basis of our previous genome sequence data that identified a deletion of P fimbriae and a transposon insertion mutation in the *fimB* gene

of ten clinical *E. coli* ST131 isolates (Clark *et al.* 2012), it is important to examine the biological relevance and the effect of these mutations on the true virulence of ST131 isolates. The recent work by Gibreel *et al.* (2012a) highlighted the need to conduct comprehensive studies on the metabolic potential of *E. coli* ST131 and other ExPEC STs using different testing approaches to determine any link between metabolic potential and increased pathogenesis of *E. coli* ST131. Given that the comparative genomics performed to date on *E. coli* ST131 have mainly focussed on virulence associated genes (Petty *et al.* 2014), comparing the whole gene content of a collection of *E. coli* strains including ST131 and non-ST131 is important to identify the presence of other loci uniquely associated with the H30Rx clade of *E. coli* ST131. Finally, based on recent findings demonstrating the role of intra-ST131 recombination in shaping the lineage as well as diversity in the capsule region of ST131 H30Rx strains at both genetic (Petty *et al.* 2014) and biochemical (Olesen *et al.* 2013) levels, further investigation of the effect of this recombination and the reported diversity is essential to assess of their effects on the virulence of ST131 H30Rx strains.

1.8 Aims of the project

This PhD project mainly aimed to characterise the traits that give *E. coli* ST131 a competitive fitness advantage over other potential ExPEC clones. This aim was achieved through carrying out a range of phenotypic and genotypic assays including:

- Performing comparative bacterial growth assays in LB medium and in McCoy's 5A modified tissue culture medium.

- Testing the ability of *E. coli* ST131 strains to express functional type 1 fimbriae in presence of the aforementioned insertion mutation in the *fimB* gene.
- Examining the effect of the *fimB* gene insertion mutation on the capacity of ST131 strains to associate with and to invade the T24 human bladder epithelial cells by performing comparative association and invasion assays on ten green florescent protein (GFP)-tagged *E. coli* strains using T24 human bladder epithelial cells.
- Investigating the metabolic potential of *E. coli* ST131 through carrying out comparative metabolic profiling assays on a collection of *E. coli* ST131 and non-ST131 strains.
- Identifying the presence of other loci uniquely associated with the H30Rx clade of *E. coli* ST131 through comparing the gene content of a collection of *E. coli* including ST131 and non-ST131 strains.
- Investigating the genetic architecture for the diversity observed in the capsule locus of *E. coli* ST131 strains belonging to the H30Rx clade, and the diversity in the biochemical profile of K antigens of these strains.

Chapter two

General Materials and Methods

2. General Materials and Methods

2.1 Bacterial strains

A total of ninety-four *E. coli* strains were used in this project. Of these strains, forty-eight were isolated from urine samples of elderly patients and collected between October 2008 and June 2009 from Nottingham University Hospitals (NUH), Nottingham. Additionally, forty-one strains were collected between March 2011 and June 2011 from urine and blood samples of patients from Nottingham University Hospitals (NUH), Nottingham. The MLST of the strains and identification of virulence factors and antibiotic susceptibility profiles have been previously described (Croxall *et al.* 2011a, Croxall *et al.* 2011b, Alhashash *et al.* 2013). Three additional whole-genome sequenced *E. coli* non-ST131 strains were included in the strain collection and they were obtained from different sources (Welch *et al.* 2002, Chen *et al.* 2006b, McNally *et al.* 2013a). The invasive ExPEC strain *E. coli* CFT073 and the *E. coli* K-12 MG1655 strain were used as positive and negative control strains. Additionally, *Citrobacter koseri* (*C. koseri*) SMT319 strain (Townsend *et al.* 2003) was used as a positive control in U937 cell culture assays. Table 2.1 shows details on the strains used in this project.

Table 2.1: The characteristics of strains used in this project.

Strain ID	ST	Strain history	Sample source	CTX-M	Reference
UTI18	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI24	ST131	Monomicrobial UTI	Urine	-	A
UTI32	ST131	Monomicrobial UTI	Urine	CTX-M-15	A
UTI39	ST131	Polymicrobial UTI	Urine	-	A
UTI62	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI144	ST131	Polymicrobial UTI	Urine	-	A
UTI149	ST131	Polymicrobial UTI	Urine	-	A
UTI182	ST131	Polymicrobial UTI	Urine	-	A
UTI188	ST131	Monomicrobial UTI	Urine	-	A
UTI190	ST131	Polymicrobial UTI	Urine	-	A
UTI226	ST131	Polymicrobial UTI	Urine	-	A
UTI233	ST131	Polymicrobial UTI	Urine	-	A
UTI263	ST131	Polymicrobial UTI	Urine	-	A
UTI275	ST131	Polymicrobial UTI	Urine	-	A
UTI306	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI383	ST131	Polymicrobial UTI	Urine	-	A
UTI423	ST131	Polymicrobial UTI	Urine	-	A
UTI445	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI453	ST131	Polymicrobial UTI	Urine	-	A
UTI514	ST131	Polymicrobial UTI	Urine	-	A
UTI524	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI531	ST131	Polymicrobial UTI	Urine	-	A
UTI555	ST131	Polymicrobial UTI	Urine	-	A
UTI570	ST131	Polymicrobial UTI	Urine	-	A

Strain ID	ST	Strain history	Sample source	CTX-M	Reference
UTI587	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI4	ST95	Monomicrobial UTI	Urine	+[3]	A
UTI14	ST10	Polymicrobial UTI	Urine	-	A
UTI35	ST141	Polymicrobial UTI	Urine	+	A
UTI36	ST1262	Polymicrobial UTI	Urine	-	A
UTI48	ST12	Polymicrobial UTI	Urine	-	A
UTI50	ST73	Polymicrobial UTI	Urine	+	A
UTI56	ST69	Polymicrobial UTI	Urine	-	A
UTI83	ST48	Polymicrobial UTI	Urine	-	A
UTI147	ST73	Polymicrobial UTI	Urine	-	A
UTI187	ST144	Monomicrobial UTI	Urine	-	A
UTI220	ST73	Polymicrobial UTI	Urine	-	A
UTI229	ST38	Polymicrobial UTI	Urine	-	A
UTI239	ST538	Polymicrobial UTI	Urine	-	A
UTI243	ST69	Polymicrobial UTI	Urine	-	A
UTI284	ST73	Monomicrobial UTI	Urine	-	A
UTI326	ST12	Polymicrobial UTI	Urine	-	A
UTI337	ST95	Monomicrobial UTI	Urine	+	A
UTI396	ST393	Polymicrobial UTI	Urine	-	A
UTI398	ST12	Polymicrobial UTI	Urine	-	A
UTI427	ST421	Polymicrobial UTI	Urine	-	A
UTI450	ST14	Polymicrobial UTI	Urine	-	A
UTI501	ST69	Polymicrobial UTI	Urine	-	A
UTI565	ST69	Polymicrobial UTI	Urine	-	A
UTI583	ST127	Polymicrobial UTI	Urine	+	A

Strain ID	ST	Strain history	Sample source	CTX-M	Reference
UTI89	ST95	Uncomplicated cystitis	Urine	-	B
CFT073	ST73	Acute pyelonephritis	Urine	-	C
P5B	ST10	Bacteraemia	Blood	-	D
B3	ST131	Bacteraemia	Blood	-	E
B5	ST131	Bacteraemia	Blood	CTX-M-15	E
B16	ST131	Bacteraemia	Blood	CTX-M-15	E
B22	ST131	Bacteraemia	Blood	-	E
B37	ST131	Bacteraemia	Blood	-	E
B44	ST131	Bacteraemia	Blood	-	E
B64	ST131	Bacteraemia	Blood	CTX-M-15	E
B65	ST131	Bacteraemia	Blood	CTX-M-15	E
B75	ST131	Bacteraemia	Blood	CTX-M-15	E
B89	ST131	Bacteraemia	Blood	CTX-M-15	E
U2	ST131	UTI	Urine	CTX-M-15	E
U5	ST131	UTI	Urine	-	E
U12	ST131	UTI	Urine	CTX-M-15	E
U79	ST131	UTI	Urine	CTX-M-15	E
B14	ST73	Bacteraemia	Blood	-	E
B12	ST73	Bacteraemia	Blood	-	E
U1	ST73	UTI	Urine	-	E
U7	ST73	UTI	Urine	-	E
B31	ST69	Bacteraemia	Blood	-	E
B33	ST69	Bacteraemia	Blood	-	E
U64	ST69	UTI	Urine	-	E
U67	ST69	UTI	Urine	-	E
B34	ST95	Bacteraemia	Blood	-	E

Strain ID	ST	Strain history	Sample source	CTX-M	Reference
B38	ST95	Bacteraemia	Blood	-	E
U22	ST95	UTI	Urine	-	E
U60	ST95	UTI	Urine	-	E
B9	ST10	Bacteraemia	Blood	-	E
B20	ST10	Bacteraemia	Blood	-	E
U19	ST10	UTI	Urine	-	E
U95	ST10	UTI	Urine	-	E
U18	ST3451	UTI	Urine	-	E
U58	ST91	UTI	Urine	-	E
U104	ST3452	UTI	Urine	-	E
MG1655	ST10	na[1]	Urine	na	A
U9	ST69	UTI	Urine	-	E
U16	ST69	UTI	Urine	-	E
U23	ST69	UTI	Urine	-	E
U21	ST73	UTI	Urine	-	E
U24	ST73	UTI	Urine	-	E
U30	ST73	UTI	Urine	-	E
U74	ST95	UTI	Urine	-	E
U75	ST95	UTI	Urine	-	E
C. koseri	na	Meningitis	CSF[2]	-	F

[1]na: not available; [2]CSF: cerebrospinal fluid; [3]Positive CTX-M carriage by PCR but the CTX-M type is unavailable; A: (Croxall *et al.* 2011b); B:(Chen *et al.* 2006b); C: (Welch *et al.* 2002); D: (McNally *et al.* 2013a); E: (Alhashash *et al.* 2013) and F: (Townsend *et al.* 2003).

2.2 Antibiotics

Upon requirement, antibiotics were added to culture media or agar plates in the following concentrations: hygromycin B 100µg/ml, gentamicin 125µg/ml and 50µg/ml (Sigma Aldrich, UK).

2.3 Culture media

2.3.1 Luria-Bertani (LB) medium

LB agar is composed of 10g/L tryptone, 5g/L yeast extract, 10g/L sodium chloride (NaCL) and 15g/L granulated agar (Fisher Scientific, UK). It was prepared by dissolving 40g of LB agar in 1L of dH_2O followed by autoclaving at 121°C for 15 minutes.

2.3.2 LB broth (LBB)

LB broth is composed of 10g/L tryptone, 5g/L yeast extract and 5g/L sodium chloride (NaCL) (Sigma Aldrich, UK). It was prepared by dissolving 20g of LB broth in 1L of dH_2O followed by autoclaving at 121°C for 15 minutes.

2.3.3 Cysteine Lactose Electrolyte Deficient (CLED) agar

CLED agar is composed of 4g/L Peptone, 3g/L Lab-Lemco powder, 4g/L tryptone, 10g/L Lactose, 0.128g/L L-cysteine and 0.02g/L bromothymol blue (Oxoid Limited, UK). It was prepared by dissolving 36.2g in 1L of dH_2O followed by autoclaving at 121°C for 15 minutes.

2.3.4 McCoy's 5A modified medium

McCoy's 5A modified medium was used as a minimal defined medium in bacterial growth studies, and also as a growth and infection medium in cell culture assays. It is ready-to-use and contains 2.2g/L sodium bicarbonate (Sigma Aldrich, UK).

2.3.5 SOC medium

SOC medium is a rich medium used primarily in the recovery step of *E. coli* competent cell transformation. It is ready to use and is composed of 20g/L tryptone, 5g/L yeast extract, 4.8g/L $MgSO_4$, 3.603g/L dextrose, 0.5g/L NaCl and 0.186 g/L KCl (Sigma Aldrich, UK).

2.4 Bacterial culture maintenance and growth conditions

All *E. coli* stocks were maintained in -80°C freezer as 1ml aliquots in 20% glycerol LBB. The purity of *E. coli* stocks was routinely checked using CLED agar. Bacterial strains were grown in LB agar and incubated aerobically for 18h at 37°C. Cultured strains were stored at 4°C for no more than 3 days.

2.5 General media, buffers and reagents

2.5.1 1X Tris-acetate EDTA (TAE) buffer

1xTAE buffer is made up of 20ml of 50XTAE buffer diluted in 980ml of dH_2O followed by autoclaving at 121°C for 15 minutes. 50XTAE buffer is composed of 242g Tris base, 57.1ml Glacial acetic acid and 100ml 0.5M Na2EDTA (Sigma Aldrich, UK).

2.5.2 Dulbecco's phosphate buffered saline (PBS)

Dulbecco's phosphate buffered saline (PBS) was used to wash cells during tissue culture work, and it is composed of sodium and magnesium chloride solution (Sigma Aldrich, UK).

2.5.3 Saline solution

Saline solution was used as a diluent of bacterial cultures. One saline tablet was added to 500ml dH_2O followed by Autoclaving at 121°C for 15 minutes (Oxoid Limited, UK).

2.5.4 80% glycerol solution

80% glycerol solution (Fisher Scientific, UK) was used to make bacterial stocks and it was prepared by adding 80ml of glycerol into 20ml dH_2O followed by autoclaving at 121°C for 15 minutes.

2.5.5 10% glycerol solution

10% glycerol solution (Fisher Scientific, UK) was used to resuspend bacterial pellet during the preparation of electro-competent cells. It was made by adding 10ml of glycerol into 90ml dH_2O followed by autoclaving at 121°C for 15 minutes.

2.5.6 3% α-D-mannose solution

3% α-D-mannose (Sigma Aldrich, UK) solution was used to test the agglutination inhibition during the yeast cell agglutination assay. It was prepared by dissolving 3g of α-D-mannose in 100ml of dH_2O (w/v).

2.5.7 4% paraformaldehyde solution

4% paraformaldehyde (Sigma Aldrich, UK) was used as a fixative in cell culture assays. It was prepared by dissolving 4g of paraformaldehyde in 100ml PBS (w/v).

2.5.8 1% triton X-100

Triton X-100 (Sigma Aldrich, UK) was used during cell culture studies. 1% triton X-100 was prepared by adding 100µl of Triton X-100 to 10ml of dH_2O (v/v), followed by autoclaving at 121°C for 15 minutes.

2.5.9 Vectashield mounting medium with 4', 6-diamidino-2-phenylindole (DAPI)

DAPI (Vector Laboratories, UK) was used to mount the chamber slides during cell culture studies prior to confocal microscopy visualisation, and it

contains the blue-fluorescent DAPI nucleic acid stain that specifically stains the cell nuclei.

Chapter three

Comparative phenotypic characterisation of strains belonging to different ExPEC STs

3.1 Introduction

3.1.1 The role of studying bacterial growth in pathogenesis

Bacterial growth is an important factor for the establishment and development of bacterial infection. During the infection process, growth can directly or indirectly lead to tissue damage and disordered physiology in the host, and it has been shown that failure to control bacterial growth can be associated with disease severity (Barer 2001). Studying bacterial growth is essential for the development of modern microbial physiology as well as for the discovery of the regulation of gene expression (Klumpp and Hwa 2014). Cell growth is also important for studying systems biology (Neidhardt 1999, Schaechter 2010), since it is a system-level phenomenon that depends on the coordinated functions of many cellular components (Edwards *et al.* 2001). With regard to the importance of studying bacterial growth in determining *E. coli* fitness, a comparative proteomics analysis investigated the expression of UPEC cytoplasmic proteins during growth in LB medium and in the urinary tract environment (Alteri *et al.* 2009). It found that UPEC grown in human urine expresses 84 cytoplasmic proteins, including those responsible for the import of short peptides and enzymes required for the transport and catabolism of sialic acid, gluconate and the pentose sugars such as xylose and arabinose, compared to that grown in LB medium (Alteri *et al.* 2009). This suggests that these proteins can act as main carbon sources for *E. coli* during infection of the urinary tract and therefore might be of importance in UPEC fitness *in vivo* (Alteri *et al.* 2009).

From a practical perspective, determining bacterial growth rates *in vitro* can be important for the identification of bacterial species associated with causing disease, since it has been suggested that rapidly growing bacteria

might be associated with causing acute diseases (Barer 2001). Many methods can be used to measure microbial growth such as turbidimetric measurement which is based on determining total cell mass, and the plate count method which measures the total viable cell number (Barer 2001, Madigan *et al.* 2009).

3.1.2 UPEC attachment, invasion and intracellular survival in host bladder epithelial cells

It has been shown that UPEC adherence to host uroepithelium is a crucial step for the colonisation of the urinary tract and is also precedent to invasive infection (Johnson 1991). UPEC strains possess many adherence factors, such as type 1 fimbriae and P fimbriae, which mediate UPEC attachment to the uroepithelium of the host through specific binding to different glycoconjugate receptors (Sakarya *et al.* 2003). Previous studies demonstrated the role of type 1 fimbriae in mediating UPEC adherence to host uroepithelium (Wu *et al.* 1996), invasion of superficial bladder epithelial cells (Martinez *et al.* 2000) and the formation of intracellular bacterial communities (IBCs) (Anderson *et al.* 2003). P fimbriae have also been implicated in the pathogenesis of UPEC, and previous studies showed its importance in mediating UPEC colonisation and infection of the upper urinary tract of the host (Katouli 2010).

A previous study by our research group investigated and compared the virulence-associated gene (VAG) carriage of a collection of ST131 and non-ST131 strains isolated from urine samples of elderly patients using multiplex PCR assay (Croxall *et al.* 2011a). It found that VAG carriage was variable between strains belonging to different ExPEC STs, with no specific VAG profile associated with ST131 strains. More importantly, although the

study found the type 1 fimbriae gene *fimH* to be highly prevalent in almost all ST131 and non-ST131 strains, it showed a low association between ST131 strains and some VAGs, particularly the *pap* gene cluster encoding the structural and accessory components of P-fimbriae (Johnson 1991). Consequently, to examine the impact of the presence or absence of a specific set of virulence factors on the invasive capacity of strains, the ability of *E. coli* strains to associate with and to invade T24 human bladder epithelial cells was determined using the gentamicin protection assay (Croxall *et al.* 2011a). Comparing the capacity of ST131 strains to invade T24 cells to that of the invasive *E. coli* reference strain CFT073 showed the presence of a hyper-invasive phenotype in 15 out of 27 (56%) of the tested *E. coli* ST131 isolates. However when compared with the rest of the population, ST131 strains did not invade the T24 cells to significantly higher numbers (Croxall *et al.* 2011a).

Subsequently, a recent whole-genome based study by our research group identified the absence of the *pap* gene cluster and a transposon insertion mutation in the *fimB* gene, encoding the FimB recombinase that switches on the expression of type 1 fimbriae, of ten clinical ST131 strains (Clark *et al.* 2012). These findings highlight the importance of studying the biological relevance and the effect of these mutations on the true virulence nature of ST131 strains.

3.1.3 UPEC intracellular survival in host macrophages

It has been known that innate immunity plays an important role in host defences against UPEC (Svanborg *et al.* 2001). A previous study found that $Gr1^{high}$ $CCR2^{high}$ inflammatory monocytes were recruited in the bladder by

14h post-UPEC infection (Engel *et al.* 2008), suggesting that macrophages might contribute to the innate immune response against UPEC. Macrophages recognise many invasive bacterial pathogens by a set of pattern recognition receptors (PRRs), including the Toll-like receptors (TLRs), the Nod-like receptors and the C-type lectin receptors (Bokil *et al.* 2011). It has been demonstrated that the TLR4, which recognises LPS of Gram-negative bacteria, is crucial for innate resistance to UPEC (Schilling *et al.* 2003). Additionally, the FimH adhesin of type I fimbriae can also be recognised by TLR4 (Ashkar *et al.* 2008, Mossman *et al.* 2008). Previous studies showed the ability of UPEC to subvert TLR-initiated pro-inflammatory responses (Hilbert *et al.* 2008), and to secrete TIR domain-containing proteins (Tcps) that bind to the TLR adaptor protein MyD88 to provoke TLR responses (Cirl *et al.* 2008). Several studies have suggested that the intracellular survival of UPEC can contribute to pathogenesis. For example, Rosen *et al.* (2007) demonstrated the existence of IBCs in human patients with UTI. This might enable bacterial evasion of the host immune response, permit reinfection and contribute to antibiotic resistance (Blango and Mulvey 2010).

3.1.4 Aims of the study

Since this PhD project aimed to determine the factors that give ST131 a competitive fitness advantage over other potential ExPEC clones, and given many reports showing the importance of bacterial growth in the establishment and development of bacterial infection, we aimed to determine and compare the growth rates of a collection of ExPEC including ST131 and non-ST131 to identify the presence of a distinct growth rate in ST131 strains. Additionally, based on a previous finding by our research

group identifying the absence of the *pap* gene cluster and a transposon insertion mutation in the *fimB* gene of ten clinical ST131 isolates (Clark *et al.* 2012), we sought to test the presence of the *fimB* insertion mutation in all our ST131 strains. Afterwards we investigate the biological relevance and the effect of this mutation on the virulence potential of ST131 strains. This was achieved by determining the ability of ST131 strains to express functional type 1 fimbriae in the presence of the *fimB* gene insertion mutation. On the basis of our previous study that tested the invasion capacity of a collection of ST131 and non-ST131 strains to invade T24 cells using gentamicin protection assay, and showed the presence of a hyper-invasive phenotype in 15 out of 27 (56%) of the tested *E. coli* ST131 isolates (Croxall *et al.* 2011a), we aimed to examine the dynamics of bacterial-cell interaction processes such as the bacterial survival and proliferation inside infected cells. This was achieved by performing comparative association and invasion assays on ten green florescent protein (GFP)-tagged *E. coli* strains using T24 human bladder epithelial cells, followed by visualisation of the attached or internalised bacteria using confocal microscopy. It has been shown that gentamicin protection assay is widely used for cell culture assays (Elsinghorst 1994), and it can only give an estimation of the number of attached or invaded bacteria and it is not suitable for studying the cellular outcome of bacterial infection (Saini *et al.* 1999). Therefore, this study used confocal microscopy to visualise the attached or internalised bacteria providing more information on the dynamics of bacterial-cell interaction processes such as the bacterial survival and proliferation inside infected cells (Sheppard *et al.* 2003). Finally, given that macrophages are utilised by several intracellular

pathogens as a host niche, and that these cells are recruited to the bladder in response to UPEC infection (Bokil *et al.* 2011), we sought to examine and compare the ability of a collection of ST131 and non-ST131 strains to persist and replicate in macrophages using U937 human macrophage cells.

3.2 Materials and Methods

3.2.1 Bacterial strains

A total of fifty-three *E. coli* strains were included in this study. Firstly, forty-nine *E. coli* strains, twenty-four ST131, twenty-four non-ST131, and the control *E. coli* strain CFT073 were used to determine the bacterial growth using turbidity measurement assay. Then, a subset including ten randomly selected *E. coli* strains, five ST131 and five non-ST131, and the control *E. coli* strain CFT073 was used to perform viable cell count assay. Afterwards, twenty-five *E. coli* ST131 strains were used to test the presence of the transposon insertion mutation in the *fimB* gene using conventional PCR and to investigate their ability to express functional type 1 fimbriae using yeast cell agglutination assay. The strain collection also included the control strains, *E. coli* CFT073 and the *E. coli* ST1461 strain UTI258, since they were previously tested for the presence of the *fimB* insertion giving negative and positive results, respectively. Additionally, ten *E. coli* strains, five ST131, five non-ST131 and the negative control strain *E. coli* MG1655, were used to carry out comparative cell culture assays using the T24 human bladder epithelial cells. Finally, the ability of *E. coli* strains to persist and replicate within U937 human macrophages was tested on a strain set including ten *E. coli* strains, five ST131 and five non-ST131 strains. The *E. coli* MG1655 and the *C. koseri* SMT319 were used as negative and positive controls, respectively.

Table 3.1: List of strains used to perform the experimental work in chapter 3.

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
CFT073	ST73	UTI	Urine	-	A
UTI18	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI24	ST131	Monomicrobial UTI	Urine	-	A
UTI32	ST131	Monomicrobial UTI	Urine	CTX-M-15	A
UTI39	ST131	Polymicrobial UTI	Urine	-	A
UTI62	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI144	ST131	Polymicrobial UTI	Urine	-	A
UTI149	ST131	Polymicrobial UTI	Urine	-	A
UTI182	ST131	Polymicrobial UTI	Urine	-	A
UTI188	ST131	Monomicrobial UTI	Urine	-	A
UTI190	ST131	Polymicrobial UTI	Urine	-	A
UTI445	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI453	ST131	Polymicrobial UTI	Urine	-	A
UTI514	ST131	Polymicrobial UTI	Urine	-	A
UTI524	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI531	ST131	Polymicrobial UTI	Urine	-	A
UTI555	ST131	Polymicrobial UTI	Urine	-	A
UTI570	ST131	Polymicrobial UTI	Urine	-	A
UTI587	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI4	ST95	Monomicrobial UTI	Urine	+	A
UTI14	ST10	Polymicrobial UTI	Urine	-	A
UTI35	ST141	Polymicrobial UTI	Urine	+	A
UTI36	ST1262	Polymicrobial UTI	Urine	-	A

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
UTI48	ST12	Polymicrobial UTI	Urine	-	A
UTI50	ST73	Polymicrobial UTI	Urine	+	A
UTI56	ST69	Polymicrobial UTI	Urine	-	A
UTI83	ST48	Polymicrobial UTI	Urine	-	A
UTI147	ST73	Polymicrobial UTI	Urine	-	A
UTI187	ST144	Monomicrobial UTI	Urine	-	A
UTI220	ST73	Polymicrobial UTI	Urine	-	A
UTI229	ST38	Polymicrobial UTI	Urine	-	A
UTI239	ST538	Polymicrobial UTI	Urine	-	A
UTI243	ST69	Polymicrobial UTI	Urine	-	A
UTI284	ST73	Monomicrobial UTI	Urine	-	A
UTI326	ST12	Polymicrobial UTI	Urine	-	A
UTI337	ST95	Monomicrobial UTI	Urine	+	A
UTI396	ST393	Polymicrobial UTI	Urine	-	A
UTI398	ST12	Polymicrobial UTI	Urine	-	A
UTI427	ST421	Polymicrobial UTI	Urine	-	A
UTI450	ST14	Polymicrobial UTI	Urine	-	A
UTI501	ST69	Polymicrobial UTI	Urine	-	A
UTI565	ST69	Polymicrobial UTI	Urine	-	A
UTI583	ST127	Polymicrobial UTI	Urine	+	A
UTI89	ST95	Uncomplicated cystitis	Urine	-	B
P5B	ST10	Bacteraemia	Blood	-	C
MG1655	ST10	na[1]	Urine	na	A
C. koseri	na	Meningitis	CSF[2]	-	D

[1]na: not available; [2]CSF: cerebrospinal fluid; A: (Croxall *et al.* 2011b); B:(Chen *et al.* 2006b); C: (McNally *et al.* 2013a); D: (Townsend *et al.* 2003).

3.2.2 Plasmid and primer sets

Details on the plasmids and primers used to carry out the molecular biology techniques in this study are given in Table 3.2.

Table 3.2: List of plasmids and primers used in the molecular biology techniques.

Plasmids	**Details**	**Source**
pMN402	Hyg^r, CoIE1/PAL5000 *ori*, *gfp*$^+$	A
Primers	**Sequence (5' ➔ 3')**	**Source**
fimB-IS_F	TCCTGACCCATAGTGAAATCG	B
fimB-IS_R	GCTCTATCCCAGATGCCGTA	
GFP5	GATATCGATGAAGGAGATATACATATGGCTAG	C
GFP3	GCTCGAATTCATTATTTGTAGAG	

A: Dr. Michael Niederweis, University of Nurnberg; B: (Totsika *et al.* 2011); C: (McNally *et al.* 2006).

3.2.3 Comparative bacterial growth assays

Two different assays were used to determine and compare the growth of a collection of *E. coli* strains. Firstly, the growth curves of *E. coli* strains were determined using turbidity measurement assay. Afterwards, viable cell count or plate count assay was performed to accurately estimate and then compare the number of viable cells in bacterial samples.

3.2.3.1 Turbidity measurement assay

Bacterial strains were grown in LB agar plates and incubated for 18h at 37°C. 5ml of LB broth and McCoy's 5A modified medium was inoculated with a single bacterial colony from overnight agar plates and then incubated overnight under aerobic conditions at two different temperatures, 37°C (human body temperature) and 25°C (room temperature), with shaking at 200 rpm. The optical density (OD) value of each bacterial culture was

measured and suitable dilutions were made after which 1ml of bacterial culture was transferred from the first set of tubes containing overnight bacterial cultures to the second set of tubes containing fresh 49ml LB broth and McCoy's 5A modified medium (1:50 dilution). This was performed to maintain the same initial OD for all *E. coli* strains that corresponded to the same amount of cells. The bacterial cultures were then incubated aerobically at 37°C and 25°C with shaking at 200 rpm. Using a spectrophotometer, the OD values of strains were obtained after 1h of incubation (T0) and then in every 30 minutes for 6 hours at a wavelength of 600nm. Lastly, the OD values were used to generate growth curves for the tested strains. All bacterial strains were tested in triplicate in two independent occasions.

3.2.3.2 Viable cell count assay

The preparation of bacterial cultures was performed as for the turbidity measurement assay. A volume of 1ml of each sample was carefully removed from the flasks containing bacterial cultures after 1h of incubation (T0) and in the following 4 hours to determine the viable cell counts of the tested strains at different time points. This step was achieved by diluting 20µl of bacterial cultures in 180µl of sterile saline and serial tenfold dilution was made up to 10^{-8}. Three 20µl drops of each dilution were then placed on the surface of LB agar plate and incubated overnight for 18h at 37°C. Counting the average number of colonies which developed in the area of the drops allowed obtaining the viable cell count per 20µl of the dilution (Miles *et al.* 1938). The colony-forming unit per ml (CFU/ml) was calculated for each sample at different time points using this formula:

Viable cell count (CFU/ml) = average number of colonies for a dilution x 50 x dilution factor.

3.2.4 Genomic DNA extraction

Bacterial strains were grown on LB agar plates for 18 hours at 37°C. A single colony from the plate was placed in 5ml of LB broth and then incubated overnight at 37°C with shaking. The instructions were then followed from the: GenElute™ Bacterial Genomic DNA Kits (Sigma-Aldrich, UK; Catalogue Number: NA2120-1KT). The quantity and quality of the extracted DNA were tested using the NanoDrop (Thermo Scientific; model reference: NanoDrop 2000 spectrophotometer). Acceptable DNA extraction was indicated by a clean peak on the absorbance graph, nucleic acid quantity of ≥50ng/µl and absorbance ratio reading values of ~1.8 and 1.8-2.2, for ratios 260/280 and 260/230, respectively.

3.2.5 Polymerase chain reaction (PCR) screening of the *fimB* insertion

The presence of *fimB* insertion in *E. coli* ST131 strains was screened using the conventional PCR. The PCR mastermix was prepared as following: 5µl of 6X PCR buffer (Promega, UK), 5µl of 1.5mM $MgCl_2$ (Promega, UK), 36µl sterile distilled water, 2µl of 10mM DNTPs mix, 0.5µl of fimB-IS_F primer (10pmol), 0.5µl of fimB-IS_R primer (10pmol), 0.3µl of 5U/µM taq polymerase (Promega, UK) and 1µl of DNA template. PCR condition was carried out as follows: initial denaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 52°C for 30 sec, and extension at 72°C for 30 sec, with final extension at 72°C for 2 min and 30 sec. PCR products were finally held at 4°C. Amplification of DNA was

performed using thermal cycler (Manufacturer: TECHNE; Model Reference: TC-4000).

3.2.6 DNA analysis by agarose gel electrophoresis

PCR products were visualised on a 2% agarose gel by adding of 2g molecular agarose (Fisher Scientific, UK) to 100ml of 1XTAE. 10µL (0.01%) of SYBR® SafeTM gel stain (Invitrogen, UK) was then added to the mixture. Afterwards, gel was poured into a gel tray with comb (Gene flow, UK) and let to set. The gel was then placed in an electrophoresis tank filled with 1X TAE. 10µl of PCR product was then loaded onto the gel along with a 1kb DNA ladder (Promega, UK). The gel was run for 45 minutes at 90V. The DNA band on the gel was then visualised under ultraviolet (UV) light using InGenius® gel documentation system (Syngene, UK).

3.2.7 Preparation of *Saccharomyces cerevisiae* (*S. cerevisiae*) yeast suspension

S. cerevisiae suspension was provided by Michael Brice, Nottingham Trent University Teaching Laboratory. It was prepared by inoculating a potato dextrose agar (PDA) slope with *S. cerevisiae* followed by incubation for 24h at 25°C. 10ml of sterile saline solution was then added to the slope and yeast cells were gently extracted using a sterile swab. The yeast suspension was transferred to a sterile universal tube and stored at 4°C.

3.2.8 Yeast cell agglutination assay

E. coli ST131 strains were grown in LB agar and incubated overnight at 37°C. A single colony was then inoculated into 3ml of LB broth. Bacterial cultures were incubated overnight with shaking and statically at 37°C. Static cultures were then passaged by adding 30µl of a bacterial culture into 2970µl of LB broth followed by two further rounds of static incubation.

Agglutination assay was carried out by mixing equal volumes of bacterial suspension with *Saccharomyces cerevisiae* yeast cells diluted to 5% (v/v) in PBS on a glass slide (Totsika *et al.* 2011). Mannose inhibition of agglutination was confirmed by adding 3% α-D-mannose to the yeast suspension. *E. coli* ST131 strains that failed to produce visible bacterial–yeast cell agglutination after 3 minutes were scored negative. The *E. coli* CFT073 strain and a volume of PBS were used as positive and negative controls. All bacterial strains were tested in triplicate on three independent occasions.

3.2.9 Mini prep Plasmid purification

GFP-tagged bacteria were grown overnight in 5ml LB with 100µg/ml hygromycin B antibiotic at 37°C at 200 rpm. For medium to high copy numbers of plasmids, 1.5ml of culture was used to purify the plasmid with QIAGEN plasmid miniprep kit (Qiagen, UK) according to the manufactures instructions. All plasmid preps were then reconstituted in a final volume of 50µl sterile dH_2O.

3.2.10 Preparation of electro-competent cells

Overnight bacterial culture was prepared by inoculating a single colony of each *E. coli* strain in 5ml LB broth. 1ml of overnight bacterial cultures was transferred to fresh LB broth in 1:50 dilution followed by incubation at 37°C with shaking at 200 rpm until they reached the mid log phase, where the OD_{600} was between 0.6 and 1. Bacterial suspension was then transferred to a 50ml chilled falcon tube and left on ice for 15 minutes. Bacterial cells were harvested by centrifugation at 4,000 rpm for 10 minutes at 4°C. Supernatant was discarded and pellet was then gently resuspended in 25ml

of ice cold 10% glycerol. Bacteria were then harvested again and resuspended in 12.5ml of ice cold 10% glycerol. This step was followed by incubation for 1 hour on ice. Bacteria were harvested again and resuspended in 2ml of 10% glycerol. 50µl of bacterial suspension was then aliquoted into 1.5ml microcentrifuge tubes and stored in -80°C.

3.2.11 Transformation of electro-competent cells by electroporation

1µl of purified plasmid DNA was added into 50µl aliquot of cells, mixed by gentle pipetting and incubated on ice for 15 min. The mixture was transferred to a 2ml chilled electroporation cuvette (Flowgen, UK). The electroporator was set at 2.5kV, 2.5µF and 200Ω. Following the electroporation process, 950µl of SOC medium was immediately added to the suspension to increase the efficiency of transformation. Then, the mixture was transferred to a 1.5ml microcentrifuge tube and was incubated for 2 hours at 37°C with shaking at 200 rpm. A volume of 400µl was plated onto pre-dried LB agar plate with 100µg/ml hygromycin B antibiotic followed by overnight incubation at 37°C. Lastly, it is important to point out that an aliquot of electro-competent cells containing no plasmid DNA was used as a negative control. The entire aliquot was plated onto LB agar plate containing 100µg/ml hygromycin B antibiotic to check if the electro-competent cells were contaminated with an antibiotic resistant strain of bacteria during carrying out this experiment.

3.2.12 Confirmation of pMN402 plasmid transformation by PCR

To confirm the transformation of pMN402 plasmid carrying *gfp* gene into *E. coli* electro-competent cells, the plasmid DNA was used as template to carry out conventional PCR assay using the GFP primer set (Table 3.2). The PCR

mastermix was prepared as following: 5µl of 6X PCR buffer (Promega, UK), 5µl of 1.5mM $MgCl_2$ (Promega, UK), 36µl sterile distilled water, 2µl of 10mM DNTPs mix, 0.5µl of GFP-5 primer (10pmol), 0.5µl of GFP-3 primer (10pmol), 0.3µl of 5U/µM taq polymerase (Promega, UK) and 1µl of plasmid DNA template. PCR condition was carried out as follows: initial denaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 95°C for 30 sec, annealing at 58°C for 30 sec, and extension at 72°C for 30 sec, with final extension at 72°C for 2 min and 30 sec. PCR products were finally held at 4°C. The PCR products were visualised using the same method described previously for the *fimB* insertion PCR with an exception of using 100bp ladder rather than 1kb ladder.

3.2.13 Cell culture methods

3.2.13.1 Cell lines

Two different cell lines were used in this study (Table 3.3), and they were preserved in liquid nitrogen at -160°C until they were required.

Table 3.3: Information on the cell lines used in this study.

Cell line	Passages	Reference no.	Source
T24 human bladder epithelial cell line	3-17	85061107	Public Health England (PHE)
U937 human monocytic cell line	6-15	CRL-1593.2	American Type Culture Collection (ATCC)

3.2.13.2 Cell culture media

Information on the type and composition of the cell culture media used in this study is given in Table 3.4. All media were stored at 4°C and pre-warmed in a water bath prior to use.

Table 3.4: Cell culture media used in this study.

Cell line	Growth medium	Infection medium
T24	McCoy's 5A modified medium supplemented with 10% foetal bovine serum (FBS), 1% 2mM L-glutamine, and 1% penicillin-streptomycin (Sigma Aldrich, UK).	McCoy's 5A modified medium supplemented with 1% 2mM L-glutamine (Sigma Aldrich, UK).
U937	RPMI medium containing 10% FBS, 1% 2mM L-glutamine, 1% non-essential amino acid (NEAA), and 1% penicillin-streptomycin (Sigma Aldrich, UK).	RPMI medium containing 1% 2mM L-glutamine and 1% NEAA (Sigma Aldrich, UK).

3.2.13.3 Cell line growth and maintenance

To grow a cell line, a vial of the appropriate cell line was taken from liquid nitrogen and kept on ice. A volume of 20ml of pre-warmed cell culture growth medium (Table 3.4) was added to 75-cm^3 tissue culture flasks. The cell line vial was then thawed quickly and pipetted into the culture flask followed by overnight incubation at 37°C. Afterwards, the growth medium was replaced by 20ml of fresh medium to remove the residues of cell line preservatives. The flask was incubated for 48-72h to achieve confluent monolayer of cells. For cell line sub-culturing, cells were detached using 5ml of TrypLe™ express (Life Technologies, UK) and centrifuged for 3 minutes at 1500 rpm using CENTAUR2 centrifuge (MSE, UK). The supernatant was discarded and the pellet was resuspended in 4ml of growth medium from which 1ml was transferred into a new flask contained fresh growth medium. The cell lines of use were sub-cultured two times a week. T24 cells were seeded in triplicate into 8-well BD Falcon™ culture slides (SLS, UK) two days prior to infection assays at approximately 5 x 10^4 cells/ml. U937 cells

were seeded in triplicate into 24-well plate three days prior to infection assays at approximately 2 x 10^5 cells/ml and treated with 10nmol phorbol-myristate-acetate (PMA) (Sigma Aldrich, UK) in order to trigger U937 cell differentiation into active macrophages.

3.2.13.4 Preparation of bacterial inocula

Bacterial strains were grown in LB agar and incubated for 18h at 37°C. Bacterial broth cultures were prepared by inoculating a single colony of each strain into 5ml LB broth, and incubated aerobically for 18h at 37°C with shaking. To grow Green Florescent Protein (GFP)-tagged *E. coli* strains, 100µg/ml hygromycin B antibiotic was added to LB agar and broth before making the bacterial cultures. The preparation bacterial inocula depended on the multiplicity of infection (MOI), which is commonly defined as the ratio of bacteria to eukaryotic cells (Gobert *et al.* 2002), required to infect the cell line of choice. Since this study used two different cell lines T24 and U937, two different inocula were prepared for *E. coli* strains. For T24 cell culture assays, bacteria were harvested by centrifugation and resuspended in 1ml of supplemented cell infection medium (Table 3.4) to ~ 2 x 10^6 CFU/ml to achieve a MOI: 100. For U937 cell culture assays, bacteria were harvested by centrifugation and resuspended in 1ml of supplemented cell infection medium (Table 3.4) to ~ 2 x 10^6 CFU/ml to achieve a MOI: 10.

3.2.13.5 Comparative T24 cell infection assays

A collection of eleven GFP-tagged *E. coli*, including five ST131, five non-ST131 and the *E. coli* MG1655 strains, was used to carry out the association and invasion assays on T24 cell line. For bacterial association assay, 1ml of bacterial suspension was added to pre-seeded 8-well chamber slides at

MOI: 100. The infected cells were incubated with 5% CO_2 for 3h at 37°C. Then, the slides were washed 3 times with 500µl physiologically balanced PBS and fixed with 4% paraformaldehyde. For the 2h bacterial invasion assay, T24 cells were also infected with 1ml of bacterial suspension and incubated with 5% CO_2 for 3h at 37°C. Then, cells were washed 3 times with 500µl physiologically balanced PBS, and fresh supplemented infection medium containing 125µg/ml gentamicin was added. This step was followed by 2h incubation with 5% CO_2 at 37°C and paraformaldehyde fixation. For the 12h bacterial invasion assay, cells were infected and washed as for the 2h invasion assay, and fresh supplemented infection medium containing 50µg/ml gentamicin was added. Then, cells were incubated for 12h with 5% CO_2 at 37°C and fixed with 4% paraformaldehyde. Finally, the chamber slide wells were mounted using Vectashield mounting medium with 4', 6-diamidino-2-phenylindole (DAPI) and stored in the dark at room temperature. The association and invasion of all bacterial strains were assayed in triplicate on three independent occasions.

3.2.13.6 Confocal fluorescent microscopy

Confocal fluorescent microscopy was performed on a Leica TCS SP5 microscope with an excitation wavelength of 488 nm and an emission filter of 507 nm.

3.2.13.7 Statistical analysis for T24 cell culture data

The significance of difference between the number of associated and invasive bacterial cells for each tested strain and that of the negative control strain *E. coli* MG1655 strain was determined using the one-way ANOVA test

available in Prism software (version 6.0e), and the threshold for statistical significance was a *P* value of ≤0.05.

3.2.13.8 Persistence of *E. coli* strains within U937 cell line

A collection of ten *E. coli*, including five ST131 and five non-ST131 strains, was used to carry out this experiment. The *E. coli* MG1655 strain and the *C. koseri* SMT319 were used as negative and positive strains. Differentiated U937 cells were infected with 1ml of bacterial suspension in triplicate at a M.O.I: 10, and incubated for 1h with 5% CO_2 at 37°C. Cells were washed 3 times with 500µl physiologically balanced PBS and fresh supplemented infection medium containing 125µg/ml gentamicin was added. This was followed by incubation for 5h with 5% CO_2 at 37°C. For further U937 infection time points, cells were washed 3 times with 500µl physiologically balanced PBS and fresh supplemented infection medium containing 50µg/ml gentamicin was added with a further incubation for 24h and 48h with 5% CO_2 at 37°C. Infected U937 cells were washed twice in PBS and lysed by adding 200µl of 1% Triton X-100 solution for 10 min. Serial dilutions of the lysed infected cells were plated onto LB agar in triplicate and incubated at 37°C aerobically for 16h, and colonies were counted to determine the total number of intracellular bacteria, given in log (CFU/ml), at various time points. All bacterial strains were assayed in triplicate on three independent occasions.

3.2.13.9 Statistical analysis for U937 cell culture data

The significance of difference in the levels of U937 persistence between ST131 and non-ST131 strains, at different times points: T0, T5, T24 and T48, was determined using the unpaired t-test available in Prism software

(version 6.0e), and the threshold for statistical significance was a *P* value of ≤0.05.

3.3 Results

3.3.1 Growth curves of *E. coli* strains determined by OD measurement

In order to compare the growth of forty-nine *E. coli* strains, turbidity measurement assay was carried out in the enriched medium LB broth, and in the minimal defined McCoy's 5A modified medium, at two incubation temperatures, 37°C and 25°C. Firstly, The growth rates of *E. coli* strains grown in LB at 37°C were determined by measuring the OD values at different time points. The first OD value, taken at T0, was obtained after 1h of shaking incubation, and subsequently the OD values for strains at different time points were measured every 30 minutes for 6 hours. Figure 3.1 showed that there was no detected lag time for all tested strains, and they were able to grow exponentially from the starting time point (T0). During the exponential growth phase, the curves showed that the increase in the growth rates of *E. coli* strains was very rapid, and this phase lasted for approximately 270 minutes. Then, all strains reached the stationary phase where there was no net increase or decrease in the growth rate. The growth curves also showed little to no difference in the growth rates between the *E. coli* CFT073 control strain and the tested strains. Additionally, *E. coli* ST131 strains did not show increased growth rates compared to *E. coli* non-ST131 strains (Figure 3.1).

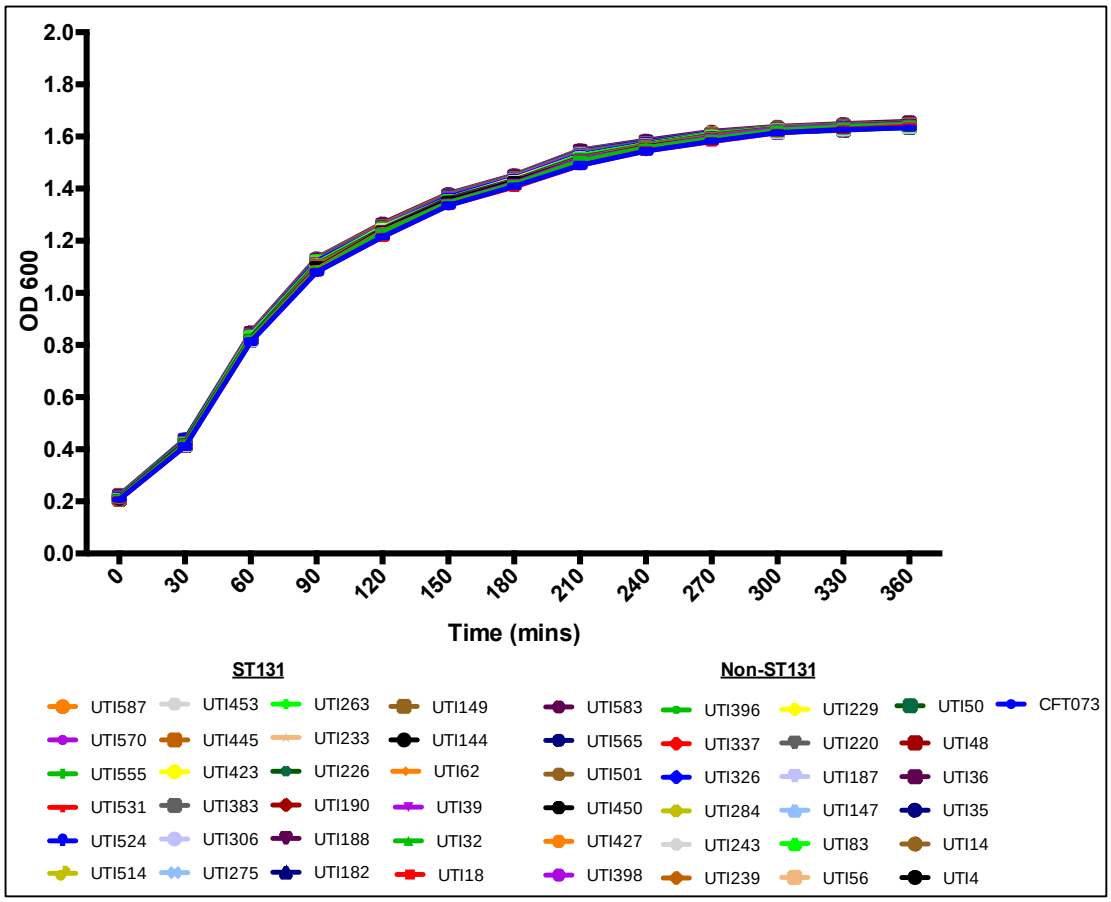


Figure 3.1: Growth curves for 49 *E. coli* strains obtained from performing turbidity measurement assay on LB medium at 37°C. The curves were plotted using averaged OD values for the tested strains.

Figure 3.2 shows the growth curves of *E. coli* strains grown in LB at 25°C. Similarly, the lag growth phase was undetectable and strains started to grow exponentially from T0. Generally, the growth rates of *E. coli* strains at 25°C was lower than that of strains grown at 37°C and this was shown by the slightly longer exponential growth phase at 25°C compared to that at 37°C. The exponential growth phase lasted for approximately 330 minutes after which *E. coli* strains entered the stationary growth phase. The growth of *E. coli* strains was almost similar to that of the *E. coli* CFT073 control strain, and there was no detectable difference between the growth rates of ST131 and non-ST131 strains when they were grown under this condition.

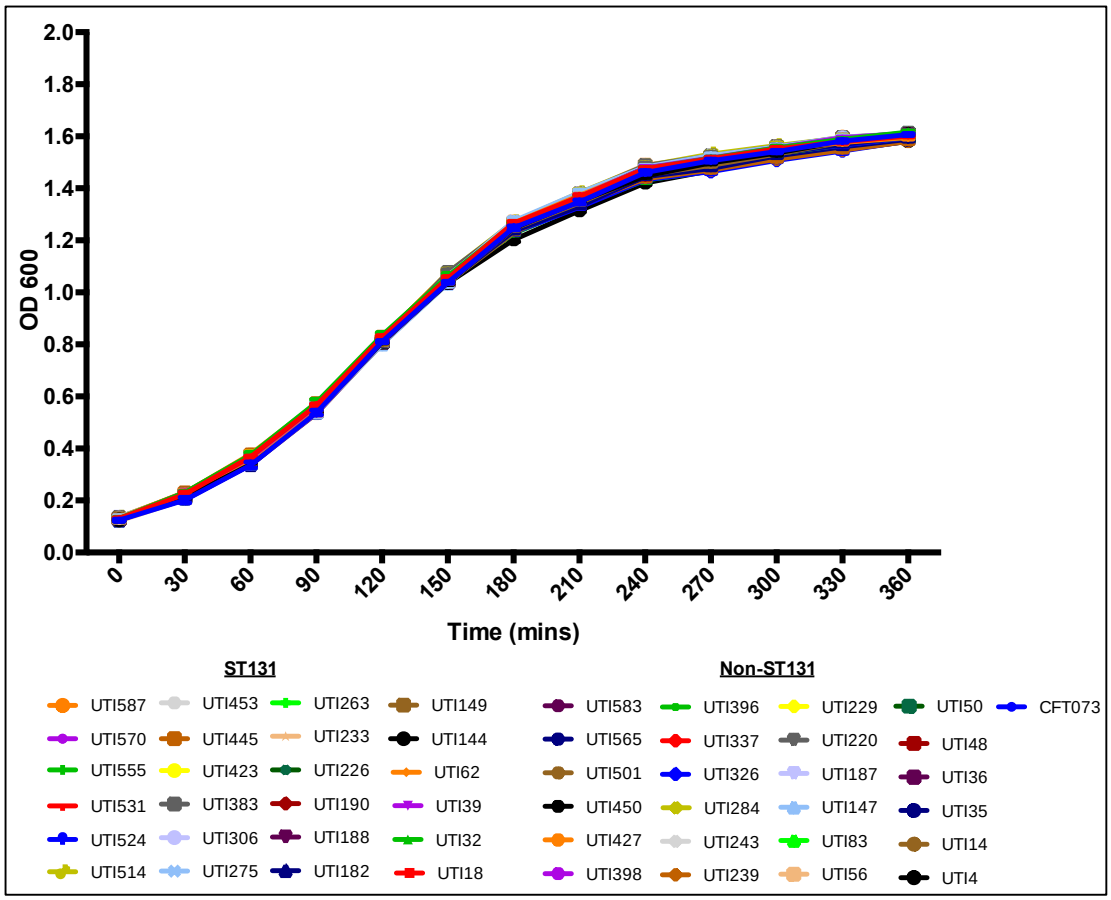


Figure 3.2: Growth curves for 49 *E. coli* strains obtained from performing turbidity measurement assay on LB medium at 25°C. The curves were plotted using averaged OD values for the tested strains.

Figure 3.3 shows the growth curves of *E. coli* strains grown in McCoy's 5A modified medium at 37°C. Although the growth curves of all strains showed the absence of lag growth phase, there was a reduction in the growth rates of *E. coli* strains compared to those determined using LB at different incubation temperatures. All strains were able to grow exponentially from T0 until approximately 330 minutes. Afterwards, strains reached the stationary growth phase. The growth of *E. coli* strains was almost comparable to that of the *E. coli* CFT073 control strain, and there was no difference between the growth rates of ST131 and non-ST131 strains.

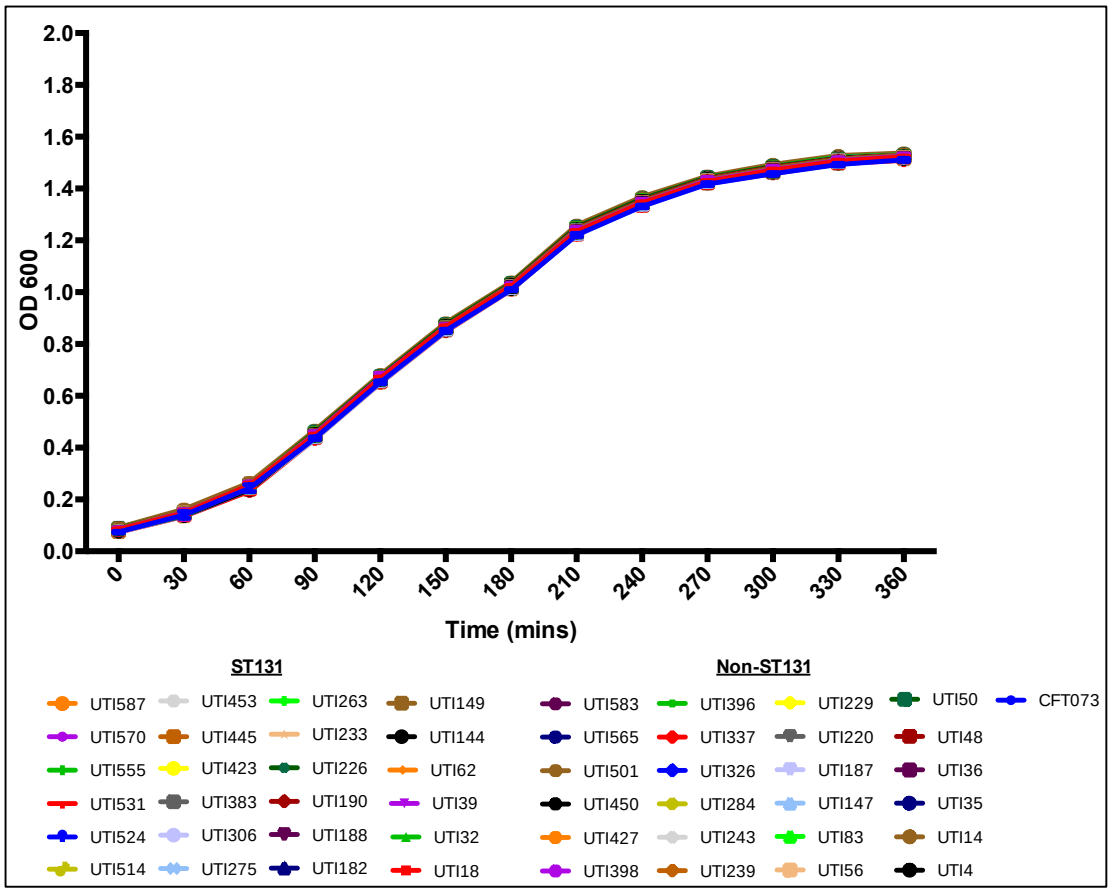


Figure 3.3: Growth curves for 49 *E. coli* strains obtained from performing turbidity measurement assay on McCoy's 5A modified medium at 37°C. The curves were plotted using averaged OD values for the tested strains.

Figure 3.4 shows the growth curves of *E. coli* strains grown in McCoy's 5A modified medium at 25°C. As for all previous condition, there was no lag growth phase for all strains. However *E. coli* strains had lowest growth rates in this condition compared to other aforementioned growth conditions used in this study. With regard to the growth rate of *E. coli* strains, the curves showed that there was no clear difference between *E. coli* ST131 and *E. coli* non-ST131 strains.

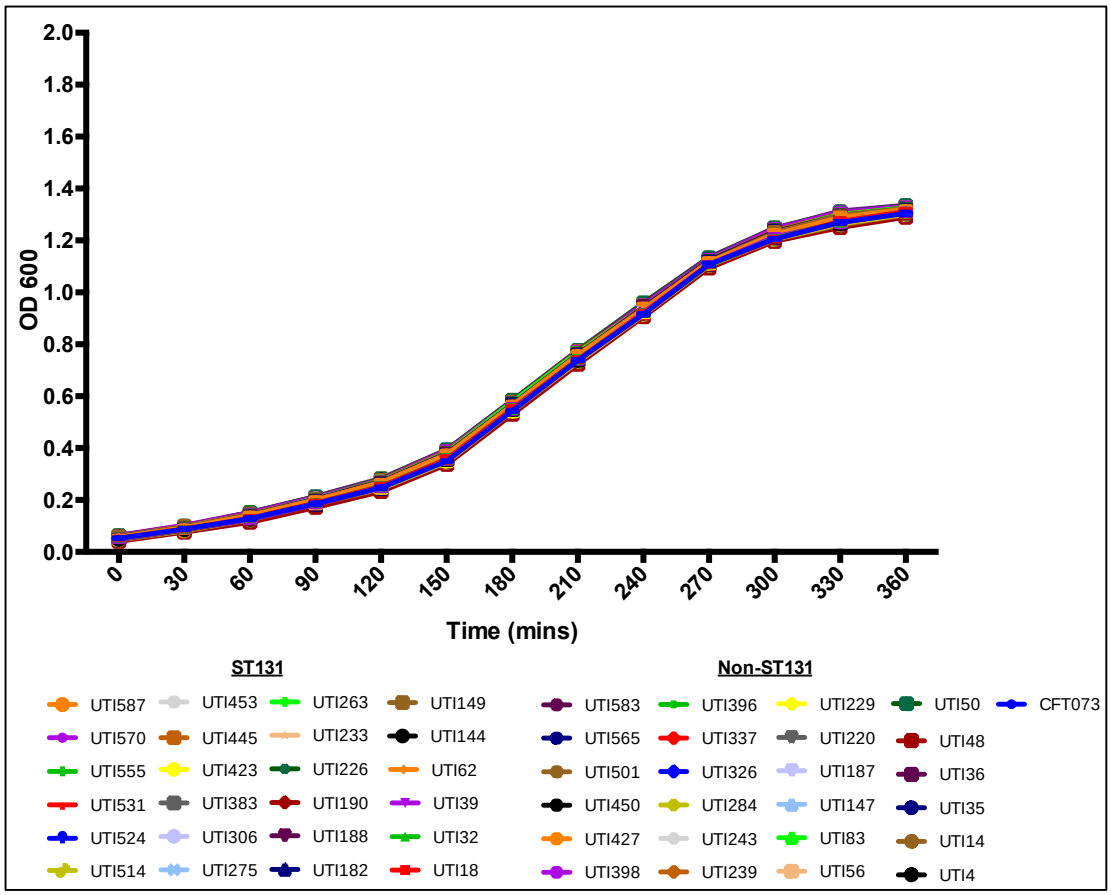


Figure 3.4: Growth curves for 49 *E. coli* strains obtained from performing turbidity measurement assay on McCoy's 5A modified medium at 25°C. The curves were plotted using averaged OD values for the tested strains.

3.3.2 Growth curves of *E. coli* strains determined by viable cell count measurement

Using the same growth conditions, viable cell counts were also determined and compared between ten randomly selected *E. coli* strains, five *E. coli* ST131, five non-ST131 and the positive control *E. coli* strain CFT073, representing the whole strain collection. Figure 3.5 shows the viable cell count values, given in log (CFU/ml), of *E. coli* strains grown in LB medium at 37°C. The log (CFU/ml) values were obtained at five different time points: T0, T60, T120, T180 and T240. Looking at the growth curves of *E. coli* strains, our data showed that all strains were able to grow exponentially from T0 with no detected lag time. Additionally, it showed that the growth curve progression of the *E. coli* strains was very rapid during the

exponential growth phase, which is consistent with the growth rates determined by OD measurement. Measuring the viable cell count values at different time points showed that the increase in the viable cell count was directly proportional to the incubation period. In general, the data showed no difference in the viable cell counts between the tested *E. coli* strains and the control strain, and between the *E. coli* ST131 and *E. coli* non-ST131 strains.

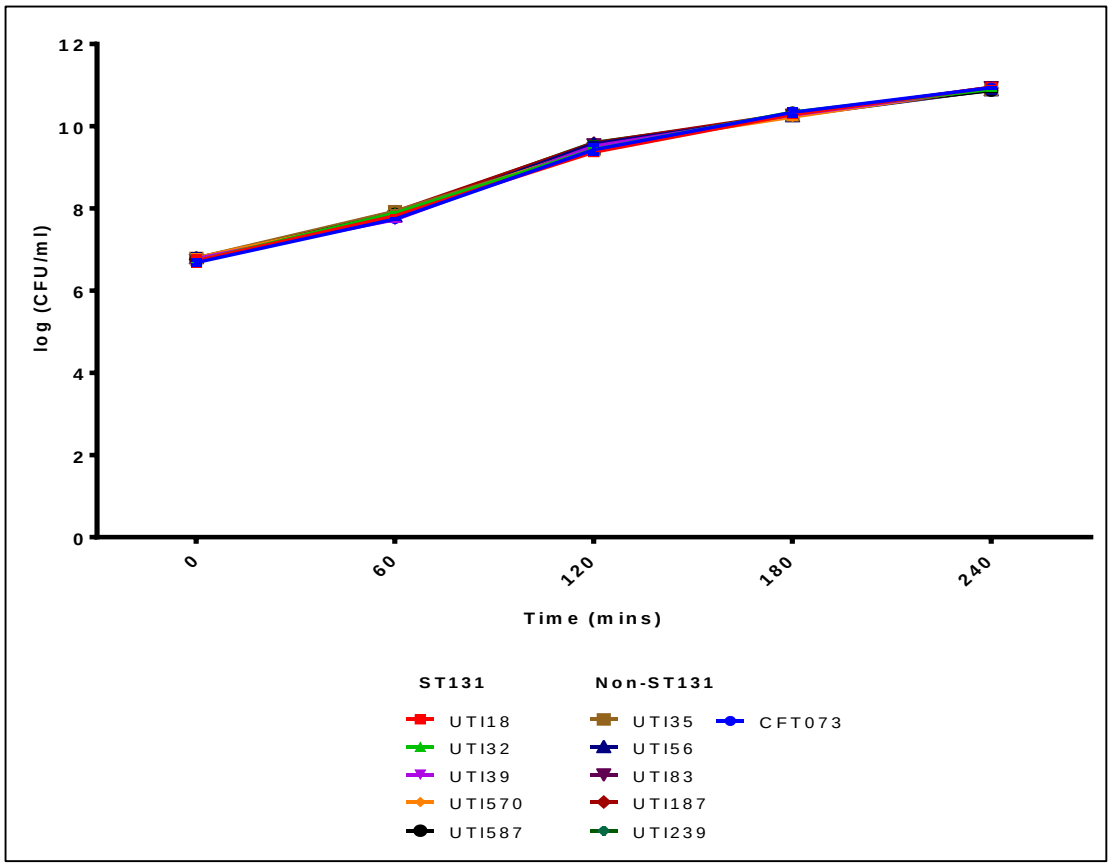


Figure 3.5: Growth curves for 11 *E. coli* strains obtained from performing viable cell count assay on LB medium at 37°C. The curves were plotted using averaged log (CFU/ml) values for the tested strains.

Figure 3.6 shows the viable cell count values, given in log (CFU/ml), of *E. coli* strains grown in LB medium at 25°C. As for the previous growth condition, the growth curves showed that all strains were able to grow exponentially from T0 with no detected lag time. However they showed that the growth of strains during the exponential growth phase was lower than

that of the previous growth condition. Our data evident undetectable difference in terms of the viable cell count values between the tested strains and the control strain, and between the *E. coli* ST131 and non-ST131 strains.

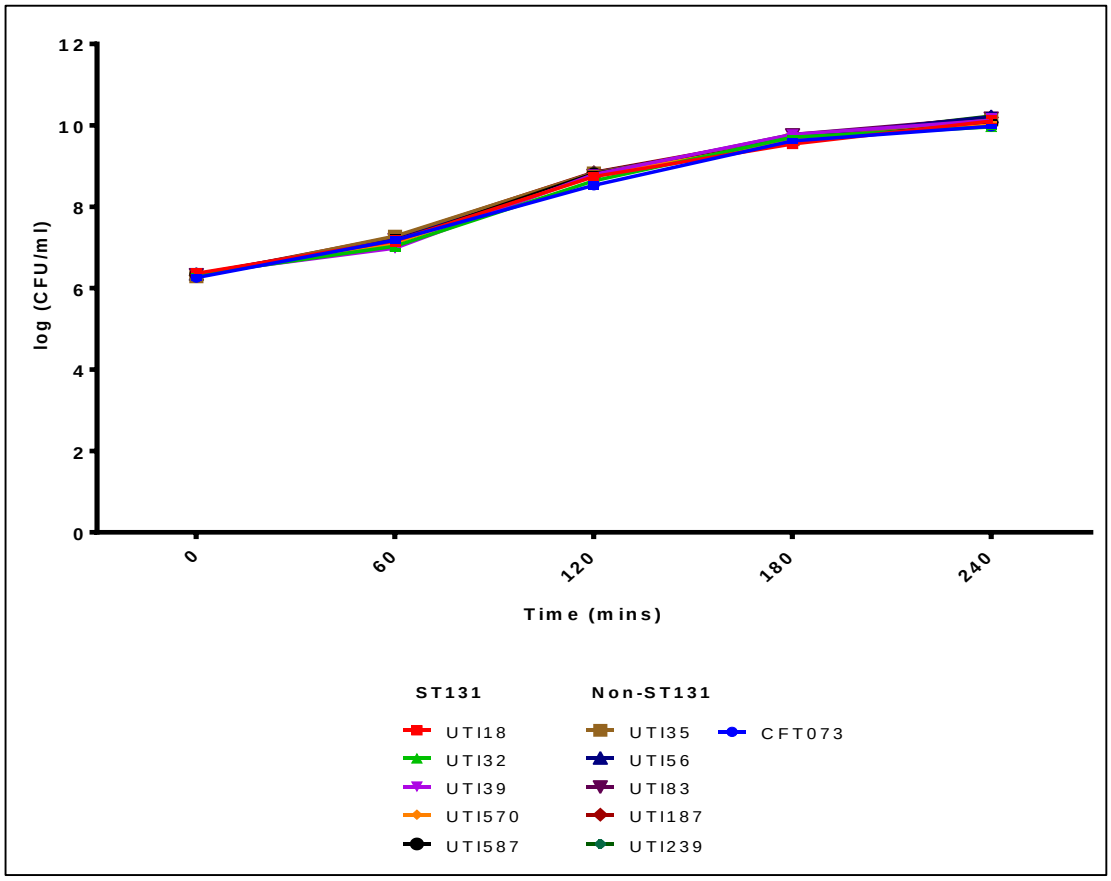


Figure 3.6: Growth curves for 11 *E. coli* strains obtained from performing viable cell count assay on LB medium at 25°C. The curves were plotted using averaged log (CFU/ml) values for the tested strains.

Figure 3.7 shows the viable cell count values, given in log (CFU/ml), of *E. coli* strains grown in McCoy's 5A modified medium at 37°C. Although the growth curves showed the absence of lag phase in all strains, there was a reduction in the growth rates of *E. coli* strains compared to those determined using LB at different incubation temperatures, 37°C and 25°C. The growth of *E. coli* strains was almost comparable to that of the control strain, and there was no difference between the growth rates of ST131 and non-ST131 strains.

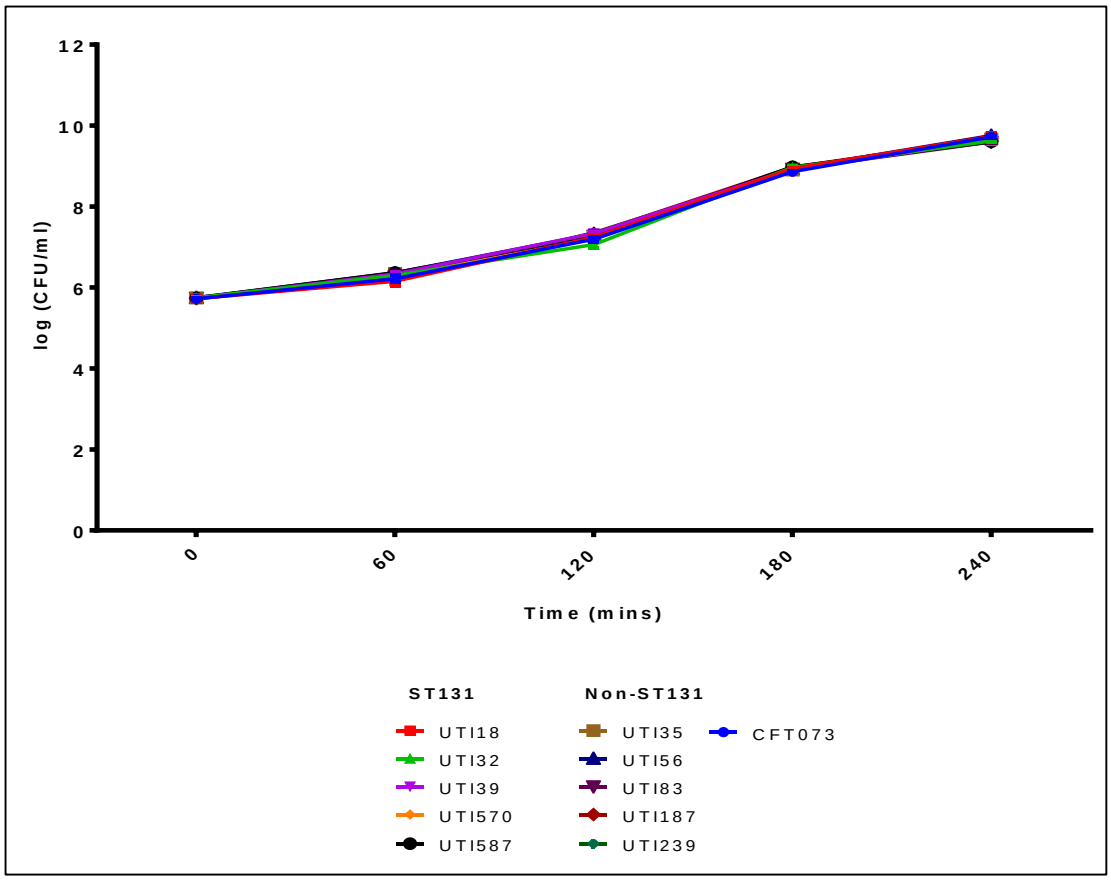


Figure 3.7: Growth curves for 11 *E. coli* strains obtained from performing viable cell count assay on McCoy's 5A modified medium at 37°C. The curves were plotted using averaged log (CFU/ml) values for the tested strains.

Figure 3.8 shows the viable cell count values, given in log (CFU/ml), of *E. coli* strains grown in McCoy's 5A modified medium at 25°C. In this growth condition, the tested *E. coli* strains had the lowest growth rates compared to the three aforementioned growth conditions. The growth rates of *E. coli* strains were almost similar to that of the *E. coli* CFT073 control strain, and there was no detectable difference between the growth rates of ST131 and non-ST131 strains.

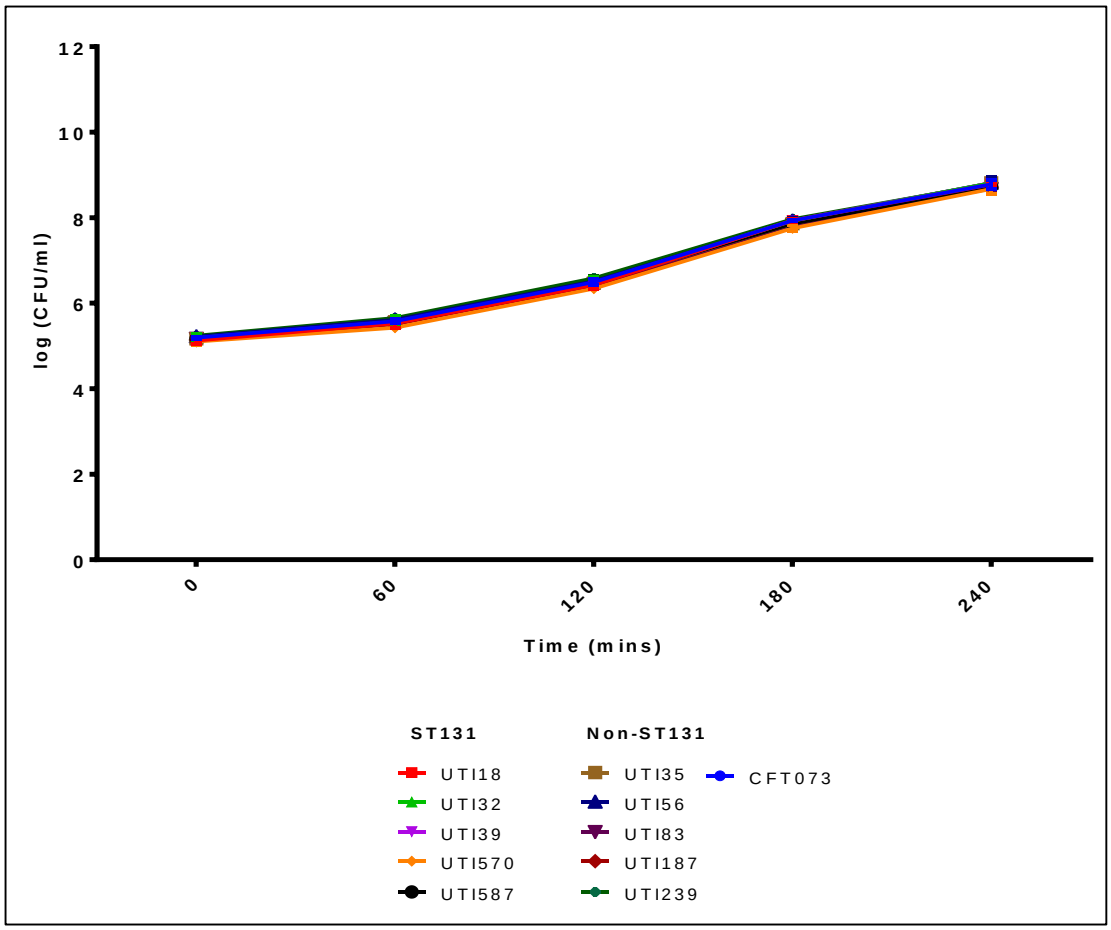


Figure 3.8: Growth curves for 11 *E. coli* strains obtained from performing viable cell count assay on McCoy's 5A modified medium at 25°C. The curves were plotted using averaged log (CFU/ml) values for the tested strains.

3.3.3 Type 1 fimbriae expression results

3.3.3.1 Screening of the *fimB* transposon insertion in *E. coli* ST131 strains

On the basis of previous genome analysis by our research group identifying the presence of a transposon insertion mutation in the *fimB* gene of ten whole-genome sequenced *E. coli* ST131 strains (Clark *et al.* 2012), twenty-five *E. coli* ST131 strains were tested for the presence of this insertion mutation using conventional PCR. The strain collection also included two control strains: *E. coli* CFT073 and UTI258. These two strains were previously tested for the presence of the *fimB* insertion, thus they were used in this assay as negative and positive controls. Figure 3.9 shows the

fimB gene PCR amplification products, and our data demonstrated that *E. coli* CFT073 had a normal *fimB* gene, i.e. without insertion, and this was showed by the presence of a 500bp *fimB* band. However a 1.895kb *fimB* transposon insertion was detected in the UTI258 strain and in 21 of the 25 ST131 strains (83.33%). Interestingly, the *E. coli* ST131 strain, UTI453, had a *fimB* insertion of lower size, which was about 1.3kb, compared to that detected in other *E. coli* ST131 strains.

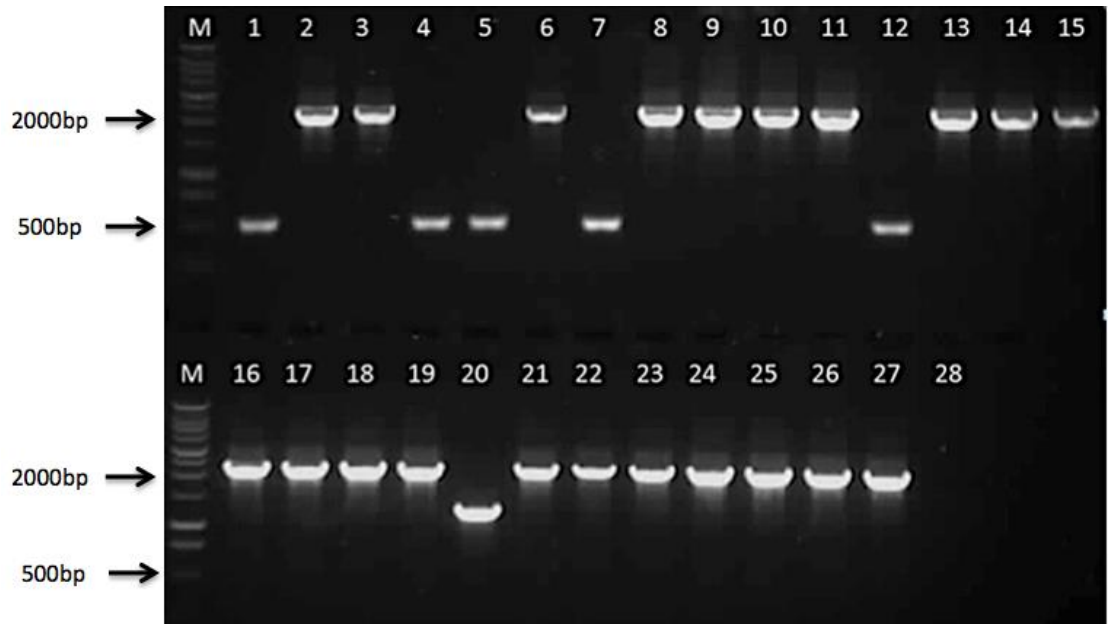


Figure 3.9: Gel electrophoresis of the *fimB* PCR products. *E. coli* strains with normal *fimB* gene had a band size of about 500bp. *E. coli* strains with detected transposon insertion were found to had a band size of about 1.895kb. The strains tested in each well were as follows: M: 1kb ladder, 1: *E. coli* CFT073, 2: UTI18, 3: UTI24, 4: UTI32, 5: UTI39, 6: UTI62, 7: UTI144, 8: UTI149, 9: UTI182, 10: UTI188, 11: UTI190, 12: UTI226, 13: UTI233, 14: UTI263, 15: UTI275, 16: UTI306, 17: UTI383, 18: UTI423, 19: UTI445, 20: UTI453, 21: UTI514, 22: UTI524, 23: UTI531, 24: UTI555, 25: UTI570, 26: UTI587, 27: UTI258, 28: PCR negative control.

3.3.3.2 Testing the ability of *E. coli* ST131 to express functional type 1 fimbriae

The presence of a transposon insertion in 21 *E. coli* ST131 strains called into question the role of this insertion in inactivating the FimB recombinase that switches on the expression of type 1 fimbriae. Yeast cell agglutination

assay was carried out to investigate the ability of all *E. coli* ST131 strains, either with or without *fimB* transposon insertion, to produce functional type 1 fimbriae. Table 3.5 shows the type 1 fimbriae expression profiles following shaking and static incubation conditions. For overnight shaking cultures, the control strain *E. coli* CFT073, which had no *fimB* insertion, was able to agglutinate the *S. cerevisiae* yeast suspension and thus produce type 1 fimbriae. However all *E. coli* ST131 strains, either with or without *fimB* insertion, and the UTI258 strain were found to be negative for type 1 fimbriae expression and this was shown by the inability of these strains to agglutinate the yeast cell suspension. When the overnight cultures of these strains were subjected to further two rounds of static incubation, all strains were able to agglutinate the yeast cell suspension and therefore express functional type 1 fimbriae. For all strains, there was variability in the yeast cell agglutination scores following the static growth, which reflected the difference in their ability to express functional type 1 fimbriae (Table 3.5). Generally, our data suggests no effect of the *fimB* insertion on the ability of ST131 strains to express functional type 1 fimbriae.

Table 3.5. The *fimB* transposon insertion and yeast cell agglutination profiles of *E. coli* strains.

Strain ID	***E. coli* ST**	***fimB* insertion**	**[a]YA (Shaking)**	**[b]YA (Static)**
E. coli CFT073	ST73	-	++	+++
UTI18	ST131	+	-	+++
UTI24	ST131	+	-	+++
UTI32	ST131	-	-	++
UTI39	ST131	-	-	+++
UTI62	ST131	+	-	++
UTI144	ST131	-	-	+++
UTI149	ST131	+	-	+++
UTI182	ST131	+	-	+++
UTI188	ST131	+	-	++
UTI190	ST131	+	-	+++
UTI226	ST131	-	-	+++
UTI233	ST131	+	-	++
UTI1263	ST131	+	-	++
UTI275	ST131	+	-	+
UTI306	ST131	+	-	++
UTI383	ST131	+	-	+++
UTI423	ST131	+	-	+++
UTI445	ST131	+	-	++
UTI453	ST131	+	-	++
UTI514	ST131	+	-	+++
UTI524	ST131	+	-	+++
UTI531	ST131	+	-	+++
UTI555	ST131	+	-	+++
UTI570	ST131	+	-	++

Strain ID	*E. coli* ST	*fimB* insertion	[a]YA (Shaking)	[b]YA (Static)
UTI587	ST131	+	-	++
UTI258	ST1461	+	-	++

[a] Yeast agglutination (YA) following shaking growth.
[b] Yeast agglutination (YA) following two rounds of static growth.

3.3.4 Confirmation of plasmid transformation by PCR

To confirm the transformation of the pMN402 plasmid carrying the *gfp* gene into the eleven *E. coli* strains included in this study, a PCR assay was carried out, using the GFP primer set designed previously by our research group (McNally *et al.* 2006), to detect the presence of the *gfp* gene in all strains. Figure 3.10 shows the *gfp* gene PCR amplification products of strains. Our data identified the absence of *gfp* band in the negative control, which was an aliquot of electro-competent cells containing no plasmid DNA. However it showed the presence of *gfp* band of the expected product size, 700bp, in all transformed *E. coli* strains. This therefore confirmed the successful transformation of the pMN402 plasmid carrying the *gfp* gene into all *E. coli* strains.

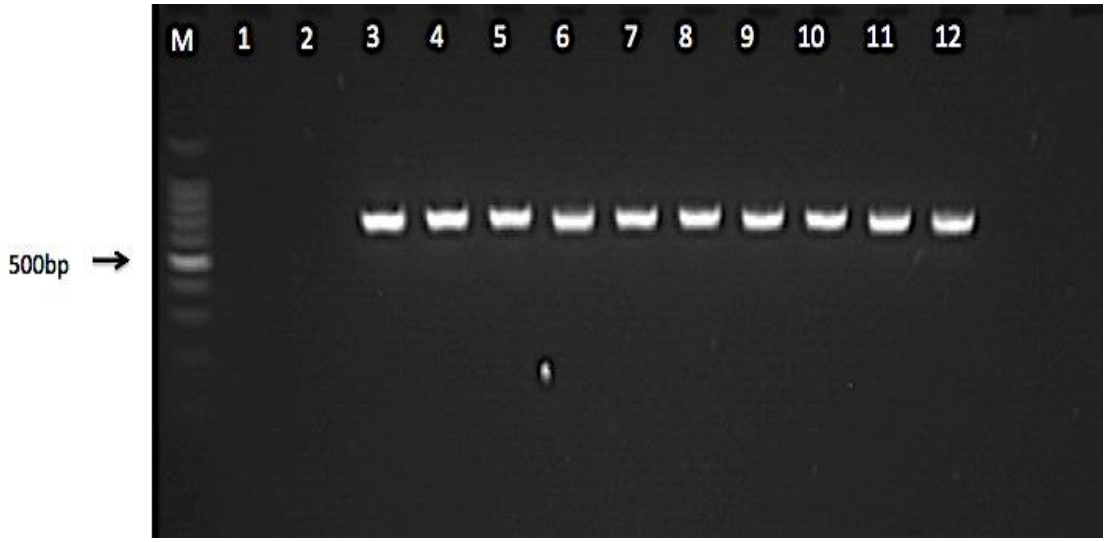


Figure 3.10: Gel electrophoresis of the *gfp* PCR products. *E. coli* strains that carried *gfp* gene had a band size of about 700bp. The strains tested in each well were as follows: M: 100bp ladder, 1: PCR negative control, 2: Aliquot of electro-competent cell with no plasmid DNA, 3: *E. coli* CFT073, 4: UTI18, 5: UTI48, 6: UTI89, 7: UTI188, 8: UTI226, 9: UTI243, 10: UTI570, 11: UTI587, 12: P5B.

3.3.5 T24 cell culture results

3.3.5.1 Association profiles of GFP-tagged *E. coli* ST131 and non-ST131 strains

Prior to the start of performing comparative cell culture assays on GFP-tagged *E. coli* strains using T24 cells, the viable cell counts of the overnight bacterial cultures were estimated in order to obtain a standard inoculum of about $2x10^6$ CFU/ml for all tested strains. This allowed us to achieve an MOI: 100, i.e. 100 bacterial cells for each T24 cell. Subsequently, the eleven GFP-tagged *E. coli* ST131 and non-ST131 strains were tested in triplicate on three independent occasions to determine and compare their ability to associate with, to invade and to form intracellular bacterial communities (IBCs) inside T24 cells. Figures 3.11 and 3.12 show representative images obtained by confocal microscope for all ST131 and non-ST131 strains after performing bacterial association assay. For ST131 strains, our data demonstrated that all these strains were able to attach to T24 cells (Figure 3.11). However there was variability in the capacity of ST131 strains to

associate with T24 cells. The levels of bacterial association in UTI18 strain were higher than that of other ST131 strains.

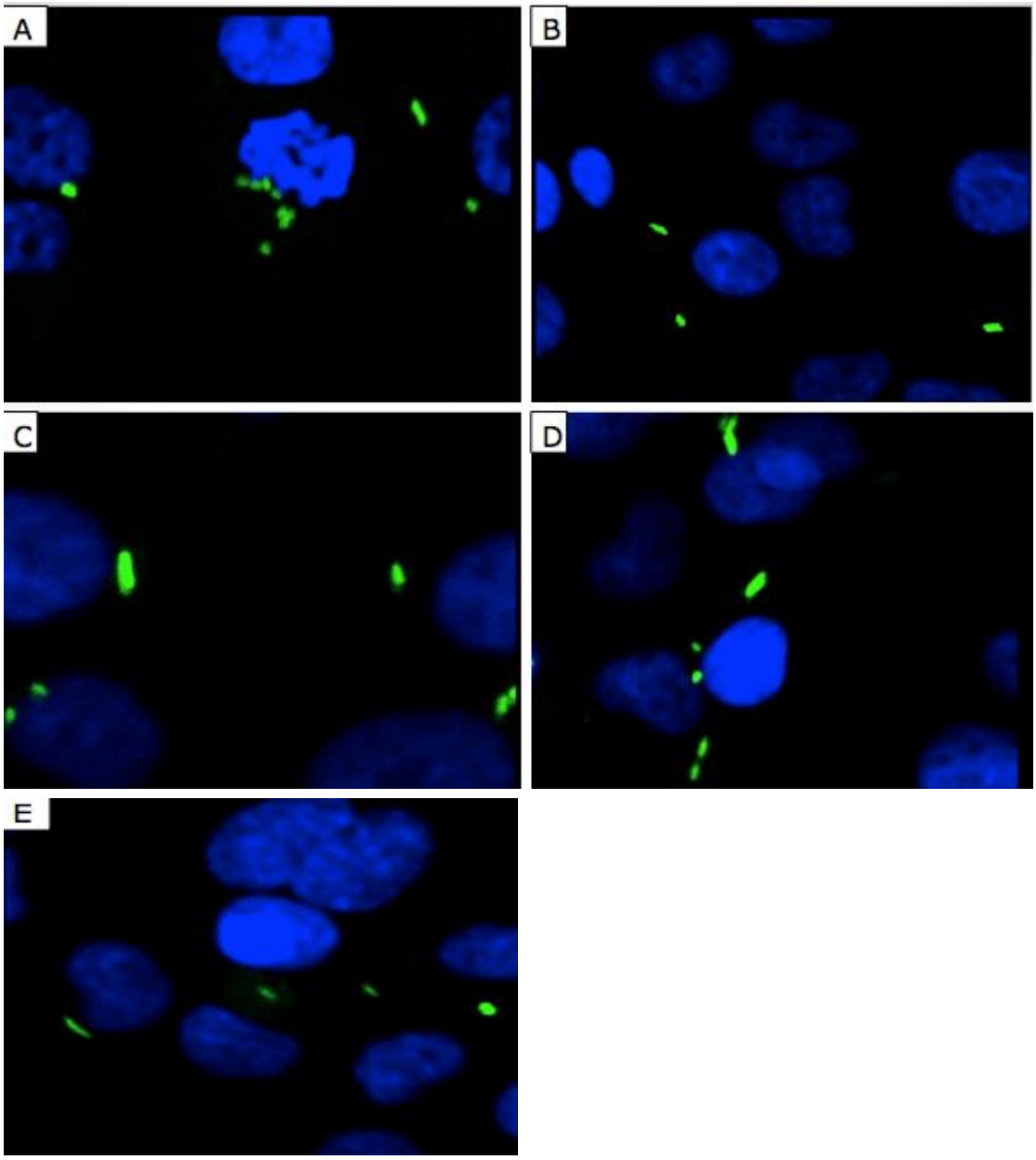


Figure 3.11: Representative florescent confocal microscope images obtained after performing association assays using T24 cells for five GFP-tagged *E. coli* ST131 strains, and show the ability of the strains to associate with the T24 cells. A: UTI18; B: UTI188; C: UTI226; D: UTI570 and E: UTI587.

With regard to the non-ST131 strains, the capacity of the positive control *E. coli* ST73 strain CFT073 and the ST69 strain UTI243 to associate with T24 cells was higher than that of other non-ST131 strains (Figure 3.12). The negative control strain MG1655 was also able to associate with T24

cells. As with ST131 strains, there was also variability in the levels of association between non-ST131 strains.

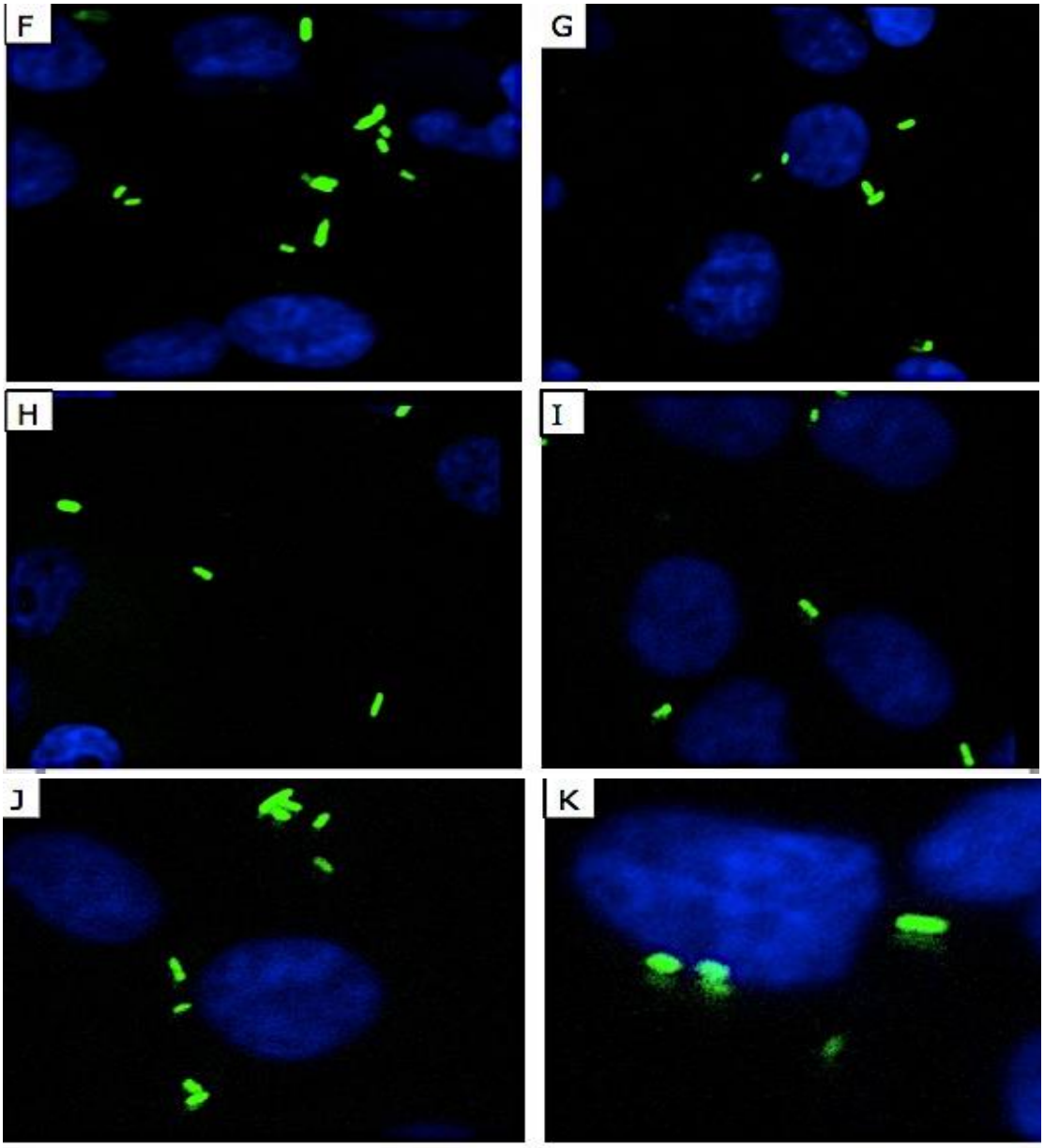


Figure 3.12: Representative florescent confocal microscope images obtained after performing association assays using T24 cells for six GFP-tagged *E. coli* non-ST131 strains, and show the ability of the strains to associate with the T24 cells. F: *E. coli* CFT073; G: UTI48; H: UTI89; I: P5B; J: UTI243 and K: MG1655.

In summary, our T24 association data for eleven *E. coli* strains belonging to ST131 and other ExPEC STs showed that all strains, including the negative control strain *E. coli* MG1655, were able to associate with T24 cells (Figures 3.11 and 3.12). It also showed that there was variability in the

association levels between the tested strains, and that the ability of three strains belonging to different STs, *E. coli* CFT073, UTI243 and UTI18, to associate with T24 cells was higher than that of other strains.

To quantitatively determine and compare the number of bacterial cells associated with T24 cells for all strains, the associated bacterial cells were counted in twenty microscopic fields. Our data showed that the CFT073, UTI18 and UTI243 strains had higher associated bacterial cell count compared to the rest of strains, and that there was a statistical significant difference in the associated bacterial cell count between these strains and the negative control strain *E. coli* MG1655 ($P<0.01$) (Figure 3.13). The average number of CFT073, UTI18 and UTI243 associated cells was 210, 203 and 195, respectively, which was almost twice the number of associated cells for the negative control strain MG1655. The number of associated cells in other ST131 and non-ST131 was also variable and it was between 93 and 150 (Figure 3.13). This confirms the variability in the association capacity of strains seen previously in the confocal microscope images and also suggests the absence of specific T24 association profile in ST131 strains.

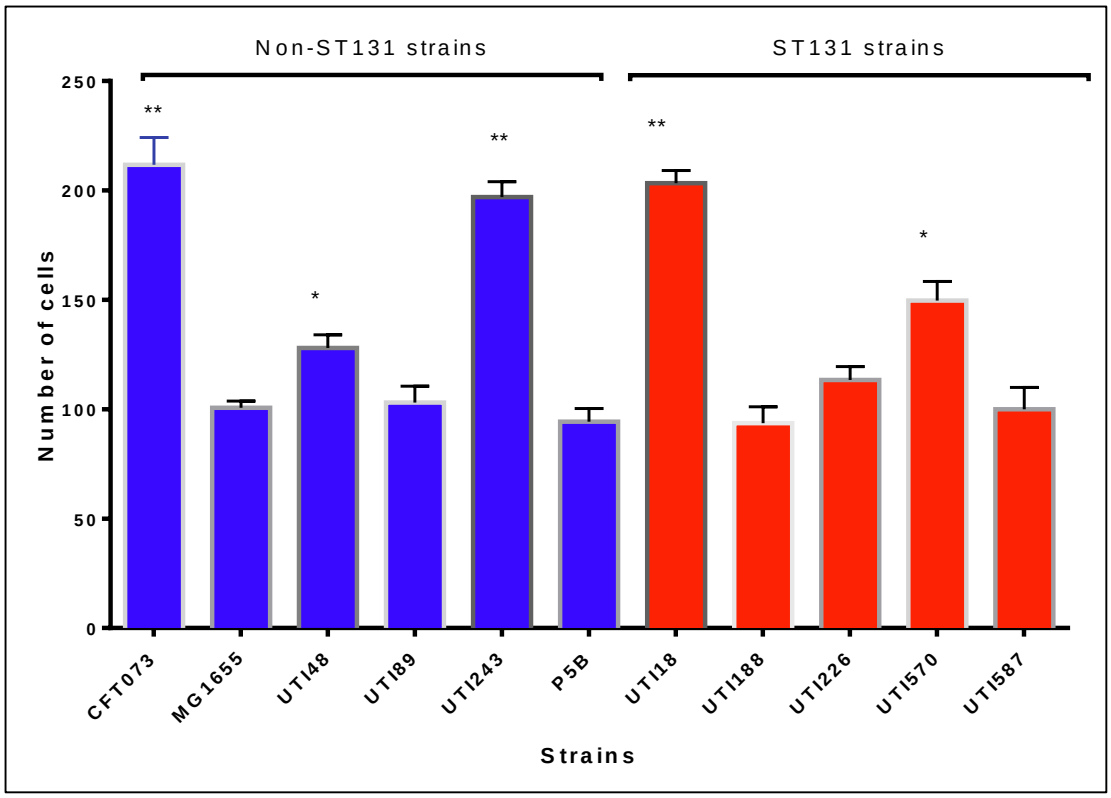


Figure 3.13: The total number of bacterial cells associated with T24 cells for each strain obtained from counting the bacterial cells in 20 microscopic fields. The associated cell count values are averaged values obtained from counting the number of associated bacterial cells in triplicate wells. The graph also shows the significance of difference between the number of associated bacterial cells for each tested strain and that of the negative control strain *E. coli* MG1655 strain as tested by the one-way ANOVA test. The presence of significant difference between is shown by the asterisks above each bar [* (P<0.05), ** (P<0.01), *** (P<0.001) and **** (P<0.0001)].

3.3.5.2 Invasion profiles of GFP-tagged *E. coli* ST131 and non-ST131 strains

To investigate the capacity of *E. coli* strains to invade T24 cells and to form intracellular bacterial communities (IBCs) inside these cells, comparative invasion assays were carried out using two different time points: 3h and 12h. The first time point allowed for assessing the ability of strains to invade T24 cells while the second one was used to determine the ability of strains to form IBCs inside T24 cells.

Figures 3.14 and 3.15 show representative images obtained for all ST131 and non-ST131 strains after performing 3h bacterial invasion assay using

T24 cells. Our data demonstrated that the all tested ST131 strains were able to invade T24 cells (Figure 3.14). However the levels of invasion were variable between these strains. UTI18 and UTI570 strains had slightly higher invasion potential than other ST131 strains (Figure 3.14).

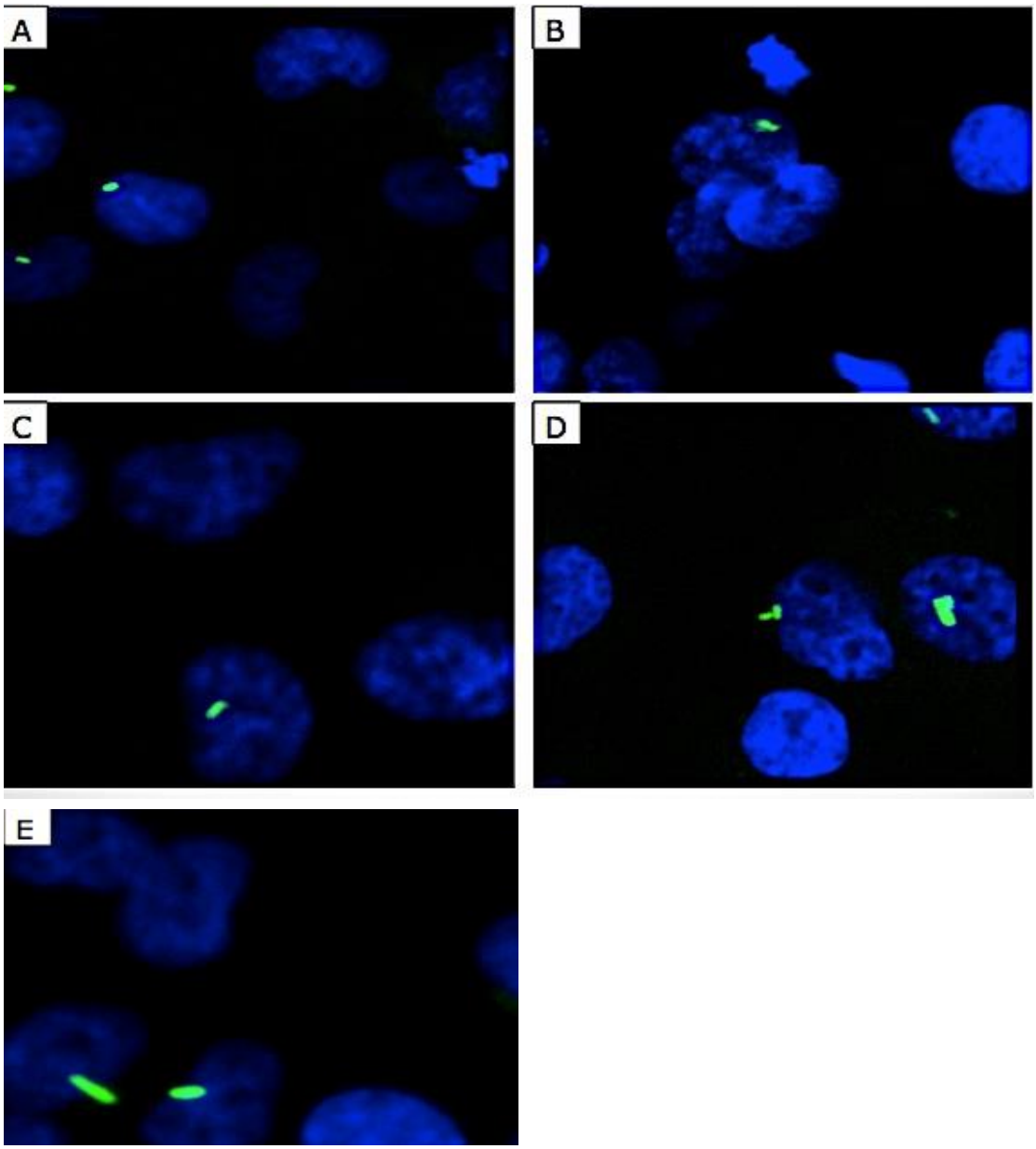


Figure 3.14: Representative florescent confocal microscope images obtained after performing 3h invasion assays using T24 cells for five GFP-tagged *E. coli* ST131 strains, and show the ability of the strains to invade the T24 cells. A: UTI18; B: UTI188; C: UTI226; D: UTI570 and E: UTI587.

With respect to the non-ST131 strains, Figure 3.15 showed that all strains, including the negative control strain MG1655, were able to invade T24 cells.

As for the ST131 strains, there was a variable invasion capacity in non-ST131 strains. The ST73 control strain *E. coli* CFT073 and the ST69 strain UTI243 had higher invasion capacity than other strains, while the negative control strain MG1655 was the lowest strain in terms of the ability to invade T24 cells (Figure 3.15).

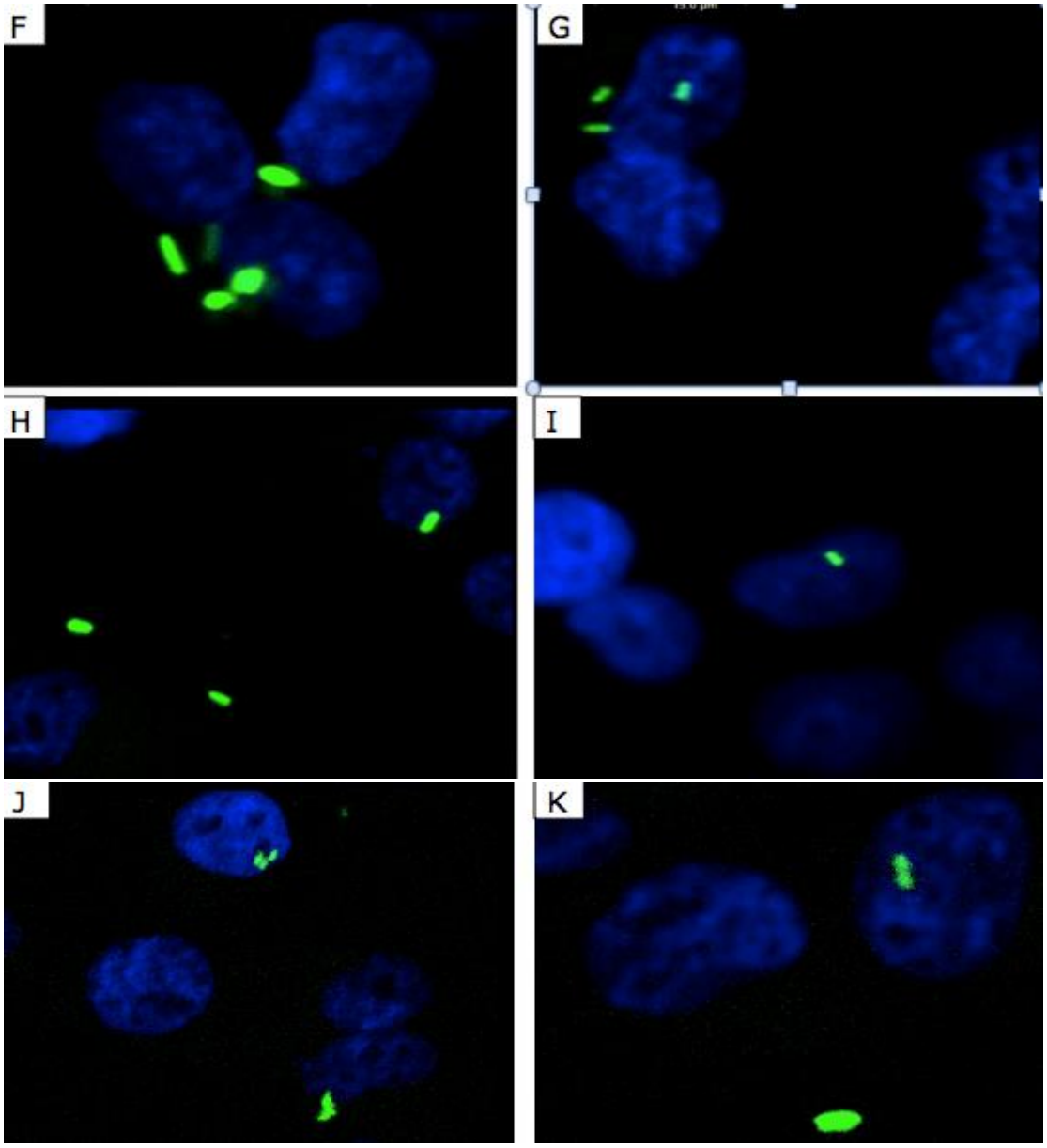


Figure 3.15: Representative florescent confocal microscope images obtained after performing 3h invasion assays using T24 cells for six GFP-tagged *E. coli* non-ST131 strains, and show the ability of the strains to invade the T24 cells. F: *E. coli* CFT073; G: UTI48; H: UTI89; I: P5B; J: UTI243 and K: MG1655.

In summary, our data obtained from testing the invasion capacity for eleven *E. coli* strains belonging to ST131 and other ExPEC STs showed that all strains, including the negative control strain *E. coli* MG1655, were able to invade T24 cells (Figures 3.14 and 3.15). It also showed variability in the invasion levels between the tested strains, and that the ability of four

strains belonging to different STs, *E. coli* CFT073, UTI243 and UTI18 to invade T24 cells was higher than that of other strains.

To quantitatively determine and compare the number of invasive bacterial cells for all strains, the invasive bacterial cells were counted in twenty microscopic fields. Our data showed that the non-ST131 strains CFT073 and UTI243, and the ST131 strain UTI18 had higher numbers of invasive bacterial cells compared to the rest of strains, and that there was a statistical significant difference in the invasive bacterial cell count between these strains and the negative control strain *E. coli* MG1655 ($P<0.01$) (Figure 3.16). The average number of invasive cells for the CFT073, UTI243 and UTI18 strains was 55, 50 and 53, respectively. However the negative control strain MG1655 was the lowest strain in terms of the number of invasive cells, with only 10 invasive bacterial cells. The number of invasive cells in other ST131 and non-ST131 was also variable and it was between 19 and 30 (Figure 3.16). This therefore confirms the variability in the invasion capacity of strains seen previously in the confocal microscope images and also suggests the absence of a specific invasion profile of ST131 strains.

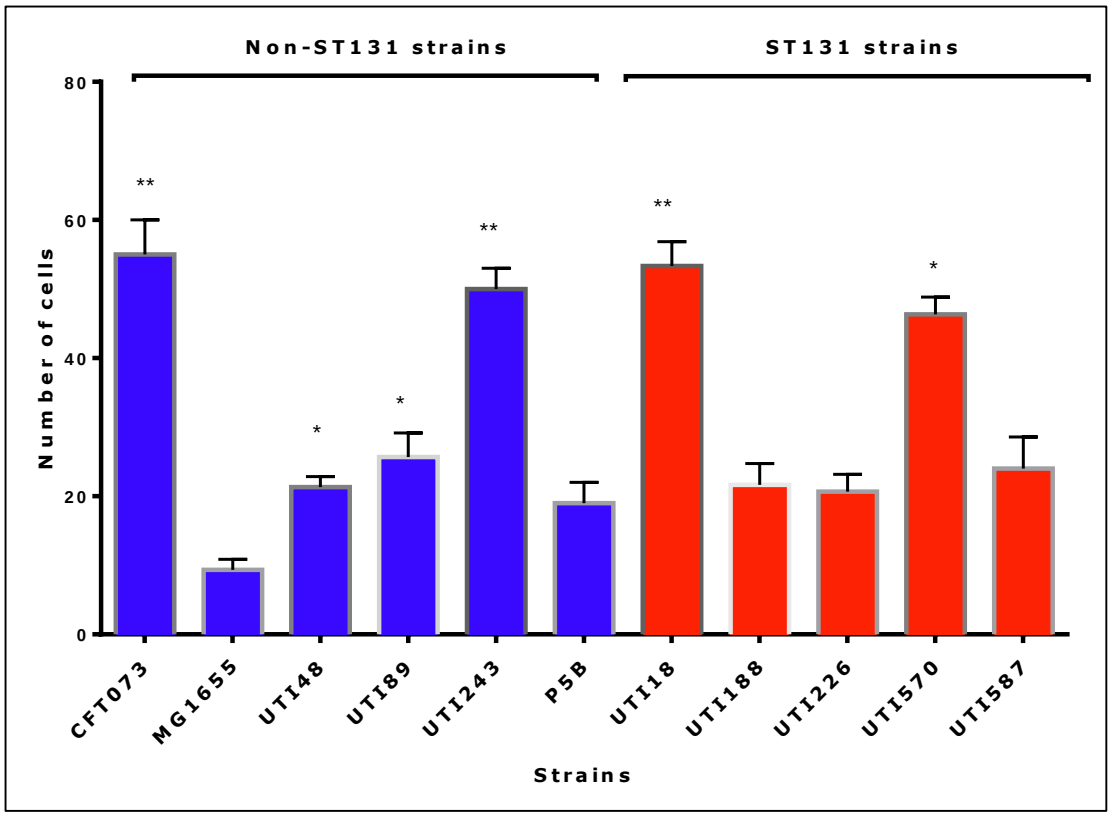


Figure 3.16: The total number of invasive bacterial cells for each strain obtained from counting the bacterial cells in 20 microscopic fields after performing 3h invasion assay. The invasive cell count values are averaged values obtained from counting the number of invasive bacterial cells in triplicate wells. The graph also shows the significance of difference between the number of invasive bacterial cells for each tested strain and that of the negative control strain *E. coli* MG1655 strain as tested by the one-way ANOVA test. The presence of significant difference between is shown by the asterisks above each bar [* (P<0.05), ** (P<0.01), *** (P<0.001) and **** (P<0.0001)].

Afterwards, to investigate the ability of the eleven *E. coli* strains to form IBCs inside T24 cells, 12h invasion assays were performed. Figures 3.17 and 3.18 show representative images obtained for all ST131 and non-ST131 strains after performing 12h bacterial invasion assays. Our data demonstrated that the all tested ST131 strains were able to form IBCs inside T24 cells (Figure 3.17). However the levels of bacterial replication were variable between these strains. UTI18 and UTI570 strains had slightly higher ability to replicate inside T24 than other ST131 strains (Figure 3.17).

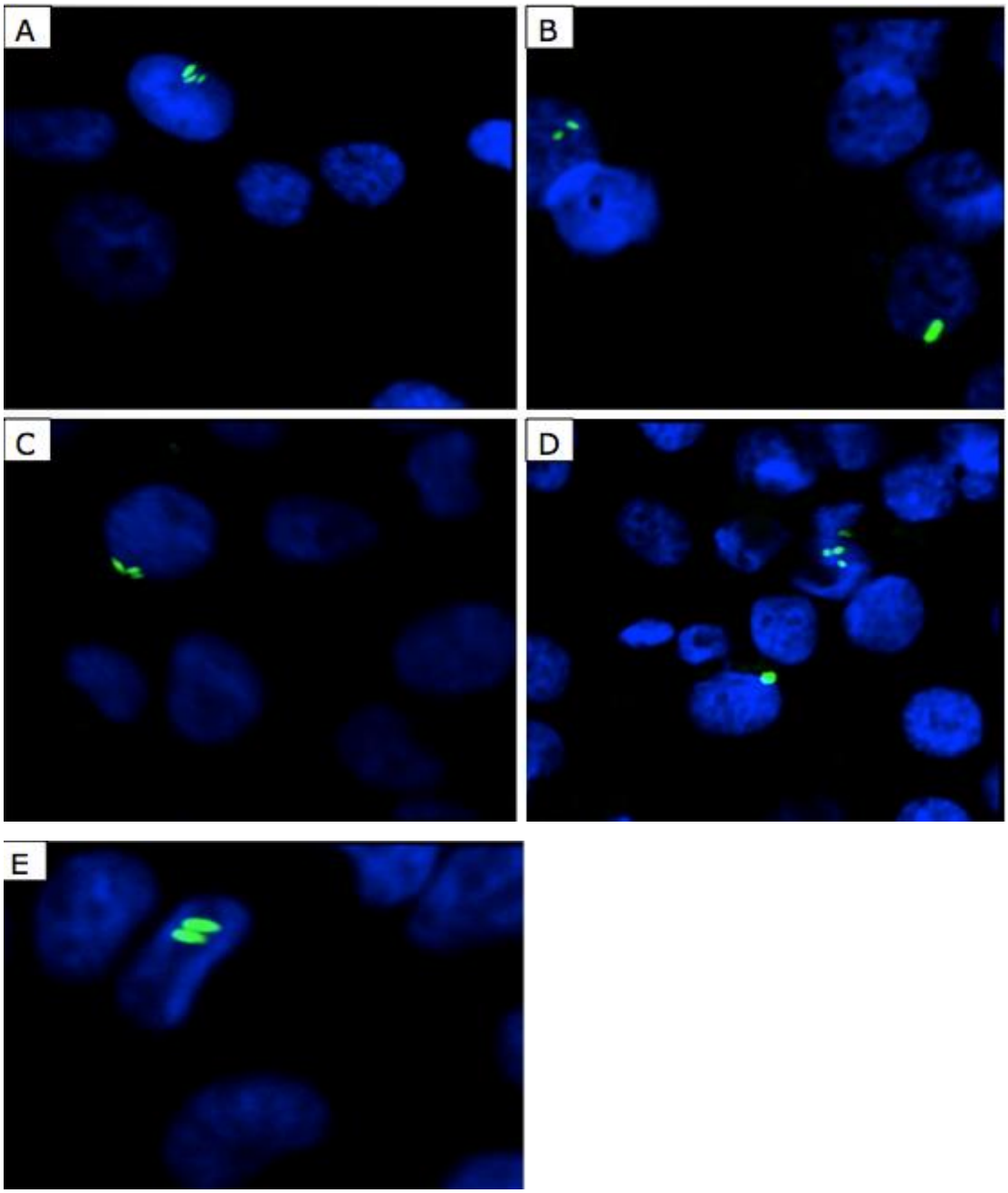


Figure 3.17: Representative florescent confocal microscope images obtained after performing 12h invasion assays using T24 cells for five GFP-tagged *E. coli* ST131 strains, and show the ability of the strains to invade the T24 cells. A: UTI18; B: UTI188; C: UTI226; D: UTI570 and E: UTI587.

With respect to the non-ST131 strains, Figure 3.18 showed that all strains, except the negative control strain MG1655, were able to replicate within T24 cells. As for the ST131 strains, there was a variable replication capacity

in non-ST131 strains. The ST73 control strain *E. coli* CFT073 had higher ability to form IBCs inside T24 cells than other strains (Figure 3.18).

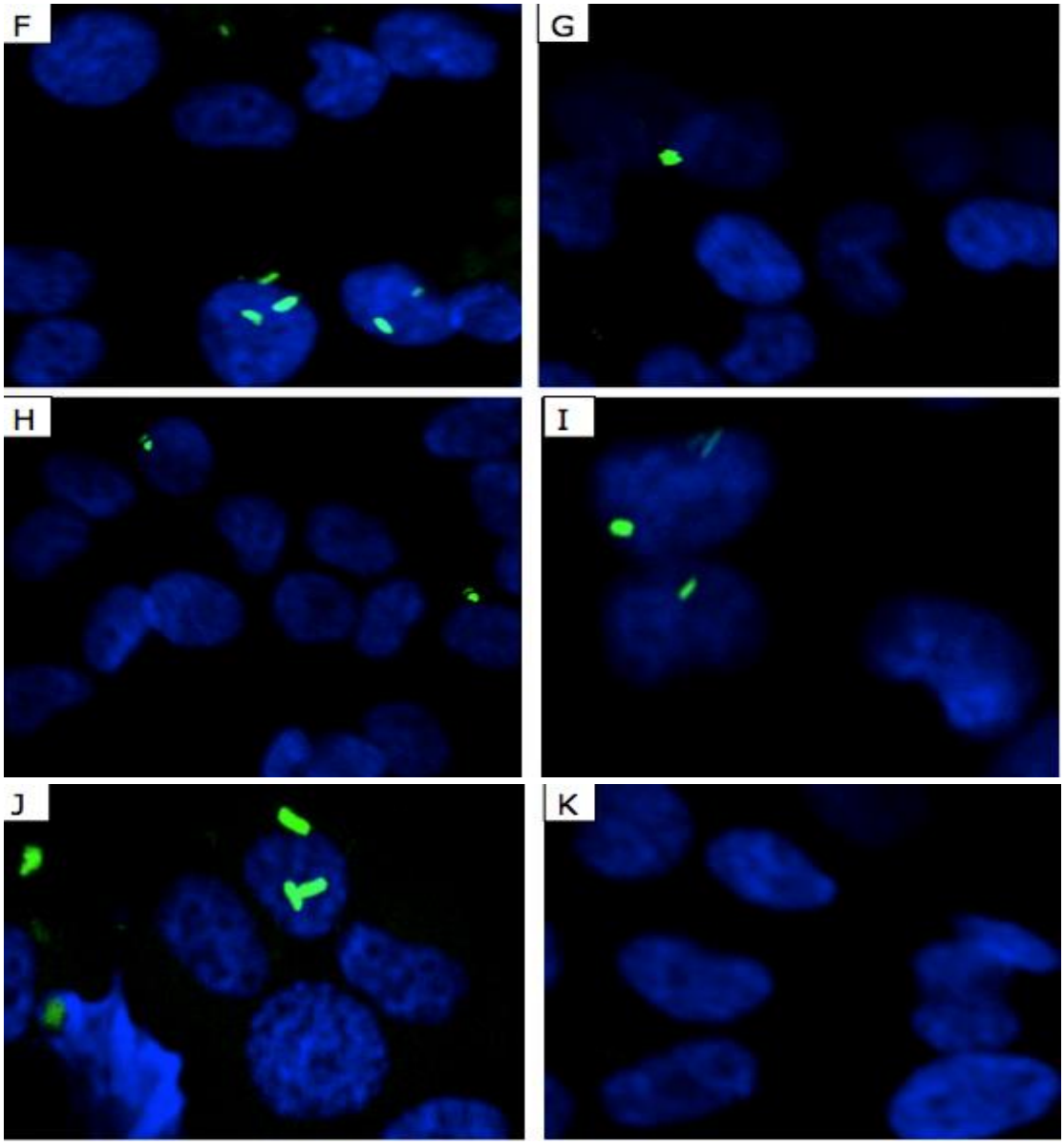


Figure 3.18: Representative florescent confocal microscope images obtained after performing 12h invasion assays using T24 cells for six GFP-tagged *E. coli* non-ST131 strains, and show the ability of the strains to invade the T24 cells. F: *E. coli* CFT073; G: UTI48; H: UTI89; I: P5B; J: UTI243 and K: MG1655.

In summary, our data obtained from testing the capacity of eleven *E. coli* strains belonging to ST131 and other ExPEC STs to form IBCs inside T24 cells showed that all strains, except the negative control strain *E. coli* MG1655, were able to replicate in T24 cells (Figures 3.17 and 3.18). It also

showed that there was variability in the replication levels between the tested strains, and that the ability of four strains belonging to different STs, *E. coli* CFT073, UTI18 and UTI570 to replicate in T24 cells was higher than that of other strains.

To quantitatively determine and compare the number of replicated bacterial cells for all strains, the replicated bacterial cells were counted in twenty microscopic fields. Our data showed that the non-ST131 positive control strain CFT073, and the ST131 strains UTI18 and UTI570 had higher numbers of replicated bacterial cells compared to the rest of strains (Figure 3.19). The average number of replicated cells for the CFT073, UTI18 and UTI570 strains was 73, 65 and 56, respectively. However the negative control strain MG1655 failed to replicate inside T24 cells. The number of replicated cells in the other ST131 and non-ST131 was also variable and it was between 33 and 46 (Figure 3.19). This is consistent with the variability in the IBCs formation capacity of strains seen previously in the confocal microscope images and also suggests the absence of specific ability of ST131 strains to form IBCs inside T24 cells.

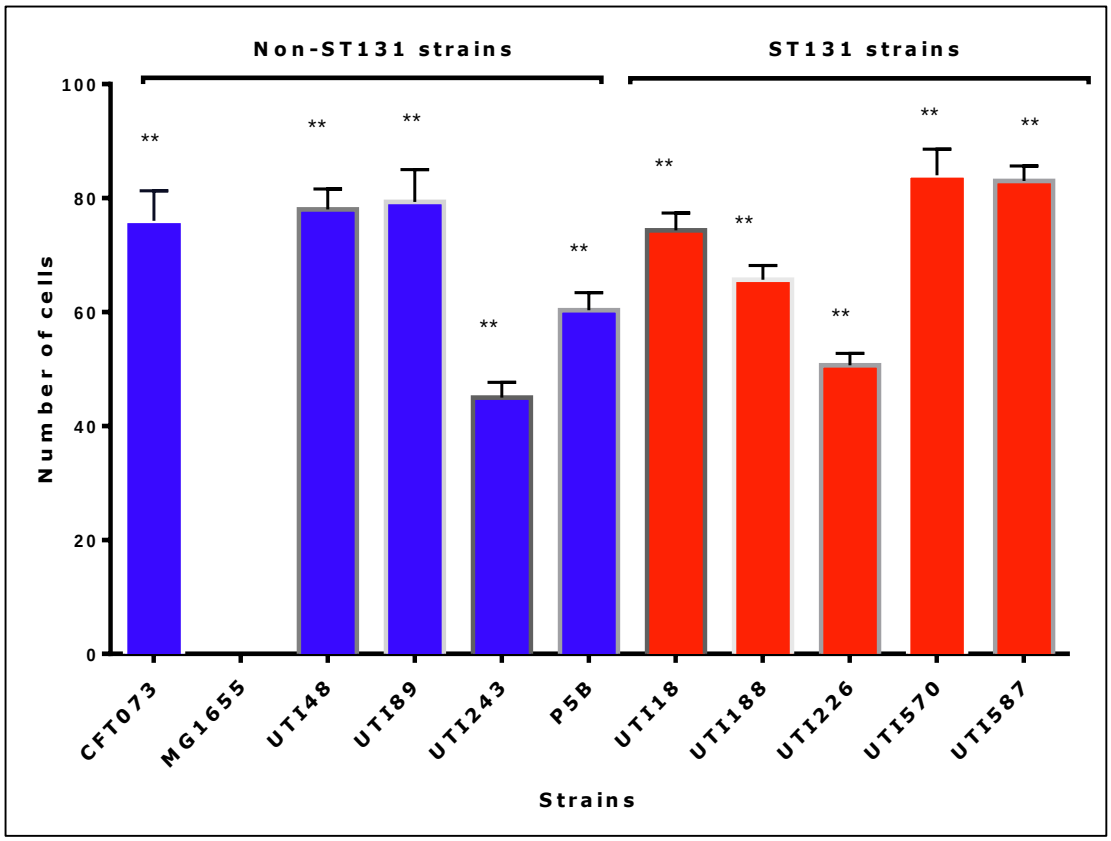


Figure 3.19: The total number of invasive bacterial cells for each strain obtained from counting the bacterial cells in 20 microscopic fields after performing 12h invasion assay. The invasive cell count values are averaged values obtained from counting the number of invasive bacterial cells in triplicate wells. The graph also shows the significance of difference between the number of invasive bacterial cells for each tested strain and that of the negative control strain *E. coli* MG1655 strain as tested by the one-way ANOVA test. The presence of significant difference between is shown by the asterisks above each bar [* (P<0.05), ** (P<0.01), *** (P<0.001) and **** (P<0.0001)].

3.3.6 Persistence of *E. coli* strains within U937 cell line results

Human differentiated U937 macrophages were used to determine and compare the ability of eleven *E. coli* strains to persist and replicate within human macrophages following phagocytosis. The *C. koseri* SMT319 strain was used as a positive control since it was previously shown to persist and replicate within U937 cells (Townsend *et al.* 2003), while the *E. coli* MG1655 strain was used as a negative control because it was found to be totally cleared by U937 cells after 24h of cell infection (Townsend *et al.* 2007). Figure 3.20 shows the U937 survival curves of all tested strains. Our data demonstrated that all tested strains, except the negative control strain *E.*

coli MG1655, were able to persist within U937 cells up to 48h after cell infection. However this ability differed between strains over the different time points.

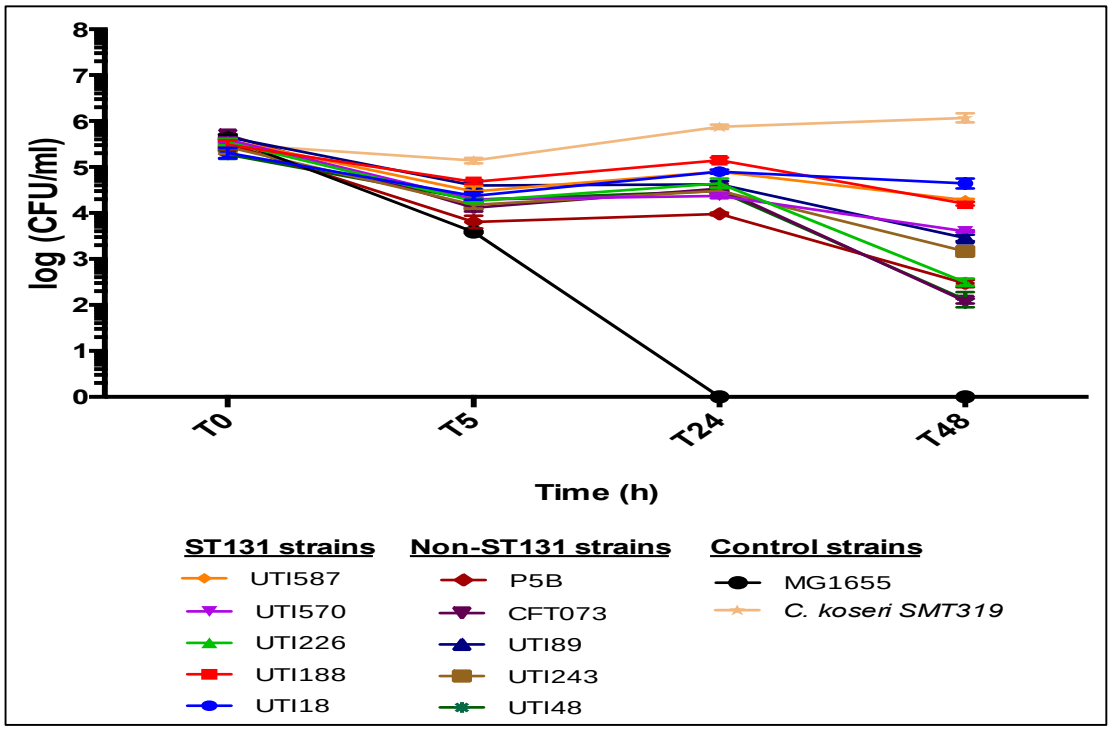


Figure 3.20: Persistence of eleven *E. coli* strains and C. koseri STT319 strain within cultured U937 human macrophages over a 48h time course. Survival curves show the number of intercellular bacterial cells at each time point, given in log (CFU/ml) values. The log (CFU/ml) values were averaged values obtained from determining the intracellular bacterial cells in triplicate wells for each strain.

At T0, the U937 uptake levels for all tested strains, including the positive and negative controls, were similar. The log CFU/ml values obtained for all strains at this time point was between 5.30 and 5.72 (Figure 3.20), and the statistical difference in the log CFU/ml values at this time point between ST131 and non-ST131 strains was insignificant (P = 0.915) (data not shown).

At T5, there was a clear reduction in the intracellular bacterial cell count of all strains compared to that at T0 (uptake), and the log CFU/ml values obtained for all strains were between 3.58 and 5.14 (Figure 3.20). Among

all strains, the positive control strain *C. koseri* SMT319 was the highest in terms of the intracellular bacterial cell count. With regard to the *E. coli* strains, although there is a little difference in the log CFU/ml values of strains at this time point, the grouped statistical analysis showed that there was no significant difference between the intracellular bacterial counts of ST131 and non-ST131 strains (P = 0.159) (data not shown).

At T24, the intracellular bacterial count for all tested strains began to increase. The negative control strain *E. coli* MG1655 was totally cleared at this time point. The number of intracellular bacterial counts of all persistent strains ranged from 4.37 to 5.87 (Figure 3.20). As with T5, the positive control strain *C. koseri* SMT319 was the highest in terms of the intracellular bacterial cell count among all tested strains. With respect to the *E. coli* ST131 and non-ST131 strains, although they were all able to persist within U937 cells, there was a slight variability in their persistence ability and this was observed even between strains belonging to the same ST. Generally, the difference in the intracellular bacterial cell count between *E. coli* ST131 strains and non-ST131 at T24 was found to be statistically insignificant (P = 0.06) (data not shown).

Finally, determining the intracellular bacterial cell count of all strains at T48 showed that there was a reduction in their log CFU/ml values. All *E. coli* strains were able to persist within U937 cells, expect the negative control strain *E. coli* MG1655. The positive control strain *C. koseri* SMT319 was the only strain to show an increased capability to persist and replicate within U937 cells, and this was shown by its higher bacterial cell count compared to other persistent *E. coli* strains (Figure 3.20). Considering the persistence

levels of *E. coli* strains, there was variability in their persistence ability and this was observed even between strains belonging to the same ST. Although the intracellular bacterial cell count of the *E. coli* ST131 strains was not significantly higher than that of non-ST131 strains (P = 0.08) (data not shown), the graph shows that there was a noticeable difference between strains in terms of their intracellular bacterial cell count at this time point (Figure 3.20). The insignificant difference obtained from performing the statistical analysis might result from the way that the statistical test used to calculate the significance of difference between strains.

In summary, determining and comparing the ability of eleven *E. coli* strains and the *C. koseri* SMT319 strain to persist and replicate within human differentiated U937 macrophages showed that all tested strains, except the negative control strain *E. coli* MG1655, were able to persist within U937 cells up to 48h after cell infection. Additionally, there was no statistical significant difference in the intracellular bacterial cell count between ST131 and non-ST131 at T0, T5, T24 and T48, suggesting that there was no specific U937 persistence profile of ST131 strains compared to non-ST131 strains.

3.4 Discussion

It has been shown that the bacterial infection process depends on the ability of the organism to grow in the host. Additionally, higher growth rates are commonly associated with the ability of pathogenic bacteria to cause recognisable infection (Barer 2001). A previous study tested a collection of urine isolates, including six *E. coli* isolates and fourteen other isolates belonging to different species such as *Staphylococcus saprophyticus (S.*

saprophyticus) biotype3 and *Proteus mirabilis* (*P. mirabilis*), to determine if differential growth rates in urine may be a discriminatory factor allowing pathogenic *E. coli* strains to cause infection, and found that the growth rates of *E. coli* isolates were significantly higher than that of isolates belonging to other species (Anderson *et al.* 1979). Firstly, this study determined and compared the growth rates of *E. coli*, including *E. coli* ST131 and non-ST131 strains, using two different growth assays, turbidity measurement assay and viable cell count assay. Our data obtained from measuring the OD value of 49 *E. coli* strains using 4 different growth conditions showed that all ST131 and non-ST131 strains had almost similar growth rates. With respect to the factors affecting the bacterial growth, testing the growth rates of the *E. coli* strains using two different types of culture media, LB medium and McCoy's 5A modified medium, showed an increased ability of the all tested strains to grow in LB medium compared to the McCoy's 5A modified medium. This was expected since the McCoy's 5A medium is a minimal-defined medium and it lacks some essential growth factors such as L-glutamine, which may affect the biosynthesis of macromolecules and energy production. This observation was supported by the previous finding showing the association between high bacterial growth rates and the availability of nutrients in culture media (Shehata and Marr 1971). In addition, the effect of using higher incubation temperature on the growth rates of strains was observed, and the growth rates of all strains incubated at 37°C were higher than those incubated at 25°C. This is consistent with many reports demonstrating the effect of temperature on the growth of bacterial strains (Madigan *et al.* 2009).

Growth curves obtained from performing viable cell counting assay on ten representative *E. coli*, five ST131 and five non-ST131 strains, showed that the difference in the growth rates of these strains remained also undetectable in all growth conditions. This confirms our observation from using turbidity measurement assay where there were almost similar growth rates of all tested strains.

With regard to the limitation of the methods used to determine the growth of *E. coli* strains, we determined the growth of a collection of ExPEC strains, including ST131 and non-ST131, using two different types of culture media in order to examine if ST131 strains are associated with differential growth rates but we did not test that in human urine. Testing the presence of differential growth rates of ST131 in human urine is important since it might be a selective factor allowing this clone to cause UTIs (Anderson *et al.* 1979). The importance of studying the bacterial growth in human urine was also highlighted by the previous finding showing that UPEC strains grown in human urine were able to express more cytoplasmic proteins than those grown in LB medium (Alteri *et al.* 2009), which suggests that these proteins can act as main carbon sources for *E. coli* during infection of the urinary tract and therefore might be of impotence in UPEC fitness *in vivo*.

Considering the expression of type 1 fimbriae in *E. coli* ST131 strains, a previous whole genome-based study by our research group identified the presence of a transposon insertion mutation in the *fimB* gene of ten clinical ST131 isolates (Clark *et al.* 2012). Our data obtained from performing PCR screening for the presence of this *fimB* gene insertion in twenty-five ST131 strains belonging to H30Rx clade identified a 1.895bp *fimB* insertion in 21

of the 25 *E. coli* ST131 strains (83.33%). In addition, one *E. coli* ST131 strain, UTI453, had an insertion in the *fimB* gene but with a lower product size obtained for the rest of strains. This *fimB* gene insertion was firstly identified in the non-pathogenic *E. coli* strain, *E. coli* Nissle 1917 (Stentebjerg-Olesen *et al.* 1999). Recently, Totiska and colleagues identified a *fimB* insertion of similar size in the *E. coli* ST131 H30Rx strain, EC958, and in 59% of the *E. coli* ST131 strains tested in that study (Totsika *et al.* 2011). Additionally, whole-genome analysis of the ST131 H30Rx strain NA114 and the ST131 strain SE15 showed the presence of insertion sequence (IS) element copy in the *fimB* gene of NA114 strain, while this insertion was undetectable in the SE15 strain (Paul *et al.* 2013). Given some previous studies showing that SE15 strain does not belong to ST131 H30Rx clade (McNally *et al.* 2013b, Petty *et al.* 2014), our data in addition to the two aforementioned studies demonstrating the presence of *fimB* insertion in ST131 strains belonging to H30Rx clade highlighted the importance of determining the effect of the *fimB* insertion on the ability of these strains to produce functional type 1 fimbriae.

The effect of the *fimB* insertion on the ability of *E. coli* ST131 strains to produce functional type 1 fimbriae was tested using yeast cell agglutination assay. All *E. coli* ST131 strains failed to express type 1 fimbriae using overnight shaking cultures. However these strains were able to produce functional type 1 fimbriae after two rounds of static incubation. It is well known that the expression of type 1 fimbriae is phase variable, and cells can switch between fimbriated and nonfimbriated states. This characteristic depends on the "on or off" orientation of an invertible 314-bp DNA switch located upstream of the *fimA* gene, encoding the major subunit protein

FimA (Abraham *et al.* 1985). The inability of all *E. coli* ST131 strains to produce type 1 fimbriae following overnight shaking incubation condition was due to the off orientation of the DNA switch. Unlike the shaking growth, static aerated growth in liquid media strongly selects for expression of type 1 fimbriae as this would turn on the orientation of the DNA switch by the action the FimB recombinase (Totsika *et al.* 2011), and this was clearly observed in the tested *E. coli* ST131 strains. Generally, the ability of all ST131 strains to produce functional type 1 fimbriae after static growth suggests that the presence of a *fimB* transposon insertion did not affect the capability of the *E. coli* ST131 strains to express functional type 1 fimbriae.

A previous study by our research group tested and compared the ability of ST131 strains to attach to and to invade T24 cells to that of the invasive *E. coli* reference strain CFT073 using gentamicin protection assay, and found the presence of a hyper-invasive phenotype in 15 out of 27 (56%) of the tested *E. coli* ST131 isolates. However when compared with the rest of the population, ST131 strains did not invade the T24 cells to significantly higher numbers (Croxall *et al.* 2011a). The gentamicin protection assay can provide only a rough estimation of the dynamics of bacterial-cell interaction processes (Saini *et al.* 1999), and so comparative association and invasion assays were performed on eleven GFP-tagged *E. coli* strains using T24 cells followed by using confocal microscopy to visualise the attached or internalised bacteria.

Our bacterial association data showed that all ST131, non-ST131 and the negative control strain *E. coli* MG1655 were able to associate with T24 cells. Confocal microscope images showed that the level of bacterial association

was variable between strains belonging to different ExPEC STs. To quantitatively assess this variability, counting the number of attached bacterial cells in 20 microscopic fields showed the presence of only strain-specific differences. Taken together, our data showed no increased ability of ST131 strains to associate with T24 cells compared to non-ST131 strains. This corroborates with our previous finding obtained from using gentamicin protection assay and showing no significant difference between ST131 and non-ST131 in terms of their T24 association profiles (Croxall *et al.* 2011a).

The capacity of eleven GFP-tagged ST131 and non-ST131 strains to invade and to form IBCs within T24 cells was also determined and compared using two different time points, 3h and 12h. Our data demonstrated that all strains were able to invade the T24 cells after 3h of incubation. The level of bacterial invasion was variable between strains belonging to different ExPEC STs. Invasive cell count showed also no difference at ST-level between strains. As with the bacterial association profiles, there was no increased ability of ST131 strains to invade T24 cells compared to non-ST131 strains which is in agreement with our previous data obtained from using gentamicin protection assay and showing no significant difference between ST131 and non-ST131 in terms of their T24 invasion profiles (Croxall *et al.* 2011a).

With respect to the 12h invasion assay, our data showed that all strains, except the negative control MG1655, were able to form IBCs within T24 cells. However the variability seen previously in the attachment and 3h invasion profiles for all strains was also detected here, and there was no difference at ST-level between strains in terms of their ability to form IBCs

within T24 cells. This finding was in agreement of a previous report showing that the pyelonephritis- and cystitis-associated *E. coli* strains CFT073 and UTI89 were more able to survive in human T24 bladder epithelial cells at 24h post-infection than the asymptomatic *E. coli* strains VR50 and 83972 (Bokil *et al.* 2011). Taken together, these observations might support the assertion that the establishment of IBCs within host epithelial cells is crucial for UPEC host evasion and reinfection, and thus UTI chronicity (Hunstad *et al.* 2005, Nielubowicz and Mobley 2010). Additionally, our data showing the ability of the negative control strain *E. coli* MG1655 to associate with T24 cells suggests the unsuitability of using this strain as a negative control for this assay. However it was shown that *E. coli* MG1655 strain was a good control for invasion assays since it was the lowest in terms of its ability to invade the T24 and it failed to form IBCs inside T24 cells.

It has been shown that macrophages are widely used by professional intracellular pathogens as an intracellular niche to evade host immune response (Bokil *et al.* 2011). Given many studies showing that macrophages might contribute to the innate immune response against UPEC (Engel *et al.* 2008), and that the intracellular survival of UPEC can contribute to pathogenesis (Blango and Mulvey 2010), the ability of *E. coli* ST131 and non-ST131 strains to persist within U937 human macrophages was determined and compared. Our data demonstrated that all *E. coli* strains, with an exception of the negative control strain MG1655, were able to persist within U937 up to 48h after cell infection. However this ability was variable between strains over the different time points. Additionally, it showed that the difference between strains was not significantly detected at ST-level, and that ST131 strains were not associated with U937

persistence profile.

Although our study was the first to investigate the UPEC persistence within U937 macrophages, there were previous studies that tested the ability of some successful UPEC strains to survive within macrophages. For example, Bokil *et al.* (2011) tested and compared the ability of the UPEC strain *E. coli* CFT073, the laboratory *E. coli* strain XL1-blue and the intracellular pathogen *Salmonella typhimurium* (*S. typhimurium*) strain SL1344, to persist and replicate within the murine bone marrow-derived macrophages (BMM). They found that intracellular CFT073, but not XL1- blue, were recovered from BMM 24h after infection. However the intracellular survival of CFT073 was lower than that of *S. typhimurium* SL1344 strain. Our data showing the ability of *E. coli* strains to persist within U937 macrophages up to 48h of infection might suggest that the long-term macrophage survival can be used by these strains to escape immune detection. However the absence of ST131 specific U937 persistence profile suggests that intramacrophage survival is not a specific trait of this clone, suggesting the presence of other factors that might contribute to its fitness.

3.5 Conclusion

On the basis of many reports showing the importance of bacterial growth in pathogenesis, this study determined and compered the growth rates of a collection of *E. coli* strains belonging to ST131 and non-ST131 using different growth measurement methods and conditions. It found that there was no difference between ST131 and non-ST131 strains in terms of their growth rates, suggesting that bacterial growth is not a factor contributing to the ST131 fitness. Screening the presence of the previously identified

fimB insertion in twenty-five *E. coli* ST131 strains showed that 21 of 25 ST131 strains (83.33%) had this insertion. Testing the effect of this insertion on the ability of *E. coli* ST131 strains to produce type 1 fimbriae showed that all strains were able to express functional type 1 fimbriae after two rounds of static growth. Testing the attachment, invasion and IBCs formation capacity of GFP-tagged ST131 and non-ST131 strains using confocal microscopy showed the presence of strain-specific differences between strains, with no difference at ST-level. Finally, determining the ability of ST131 and non-ST131 strains to persist within U937 macrophages showed also only strain-specific difference between strains, with no specific ST131 persistence profile.

Chapter four

Comparative studies on the metabolic potential and gene content of a group of *E. coli* ST131 and non-ST131 strains

4.1 Introduction

It has been suggested that the ability of ExPEC strains to colonise and cause infection in the urinary tract results from the cumulative action of specific virulence factors such as adhesins and iron acquisition determinants (Johnson 1991). Since the emergence of *E. coli* ST131 in 2008, many studies on its virulence potential have been conducted. Initial studies reported an increased virulence associated gene (VAG) carriage in *E. coli* ST131 isolates (Coelho *et al.* 2010), and also found them responsible for causing a range of severity in disease cases (Lau *et al.* 2008b, Johnson *et al.* 2010b). However many comparative whole-genome (Toh *et al.* 2010, Avasthi *et al.* 2011, Totsika *et al.* 2011, Clark *et al.* 2012) and PCR-based (Croxall *et al.* 2011a, Gibreel *et al.* 2012b, Lavigne *et al.* 2012) studies have recently demonstrated the variability in the VAGs possessed by *E. coli* ST131 strains, and that *E. coli* ST131 strains have been associated with moderate virulence capacity compared to other potential ExPEC STs such as ST73, ST69 and ST127. These findings call into question the role of VAG carriage in the current success of *E. coli* ST131, suggesting the presence of other factors that can contribute to its fitness.

4.1.1 The role of metabolism in bacterial colonisation and pathogenesis

Metabolism is an important factor in colonisation not only in the environment but also in human and animal hosts (Rohmer *et al.* 2011). Since the transition from environment to a host niche can impose challenges on bacteria in terms of pH, oxygen levels, and nutrient availability such as free iron, an extensive metabolic repertoire is required

to accommodate these changes (Rohmer *et al.* 2011). Additionally, it has recently been suggested that high metabolic capacity enhances bacterial fitness and contributes to pathogenesis (Le Bouguénec and Schouler 2011). Previous studies showed that sugar metabolism (Le Bouguénec and Schouler 2011), and the possession of specific metabolic enzymes (Pancholi and Chhatwal 2003), may increase the virulence of many bacterial species. In addition, it has been demonstrated that the utilisation of the amino acid D-serine by UPEC strains during UTI supports bacterial growth and acts as a signalling mechanism to trigger virulence gene expression (Roesch *et al.* 2003, Anfora *et al.* 2007).

4.1.2 The proposed role of *E. coli* ST131 metabolic potential in enhancing its fitness

A recent study used the Vitek2 Advanced Expert System for metabolic profiling on a collection of 300 isolates representing the major UPEC STs (Gibreel *et al.* 2012a). The authors concluded that *E. coli* ST131 isolates had higher metabolic potential and biochemical profiles compared to other ExPEC ST isolates, which may contribute to the fitness of *E. coli* ST131. They also found a correlation between metabolic activity and antibiotic susceptibility profiles, with resistance isolates showing the highest metabolic potential (Gibreel *et al.* 2012a). Another study by Vimont *et al.* tested the ability of one *E. coli* ST131 strain to colonise the intestinal tract by measuring its maximal growth rate (MGR) using three different culture conditions: a rich medium and two minimal media containing either gluconate or glucose. By comparing the metabolic activity of the ST131 strain with three other *E. coli* strains: the commensal K-12 *E. coli* strain MG1655 and the ExPEC CFT073 and HT7 strains, the study found that the

ST131 strain had a significantly higher MGR than the non-ST131 strains (Vimont *et al.* 2012).

4.1.3 Overview of the phenotypic methods used for testing the bacterial metabolic potential

4.1.3.1 Biotyping

Biotyping is a method of subtyping bacterial isolates by using their metabolic profiles (Eberle and Kiess 2012). Generally, biotyping is based on sugar utilisation and fermentation patterns and enzyme profiles. Additionally, it has traditionally been used for identification of bacteria to the species level (Katouli *et al.* 1990), and can also be used to subdivide strains below the species level (Katouli *et al.* 1990). However this greatly depends on the biochemical variability within the species and the distribution of these variations (Lockhart and Liston 1970). *E. coli* has considerable biochemical variability, thus biotyping is reliable not only in the identification of its strains but also in reflecting its potential pathogenicity in humans (Reuter 2011). A previous study demonstrated that biotyping was useful in the characterisation of *E. coli* that cause UTI in humans and identified two biotypes, biotype 5,144,552 and biotype 5,144,572, accounting for more than 40% of tested strains (Davies 1977). Another study found that using a biotyping scheme consisting of β-glucuronidase, sorbitol, raffinose and dulcitol can differentiate the enteric pathogen *E. coli* O157:H7 from other *E. coli* serotypes (Ratnam *et al.* 1988). However the use of biotyping in epidemiological investigations is considered unreliable for some bacterial pathogens due to its low reproducibility and poor discriminatory power (Eberle and Kiess 2012, Gibreel *et al.* 2012a).

4.1.3.2 Phenotypic microarray (PM) technology

Phenotypic microarray (PM) was first developed for *E. coli* by Bochner, Gadzinski and Panomitros. PM is a high-throughput technology used for testing a wide range of cellular phenotypes (Bochner *et al.* 2001). Unlike the biotyping method, PM technology allows the simultaneous testing of hundreds or even thousands of metabolites in an automated fashion (Bochner *et al.* 2001). Additionally, it has been utilised to study the metabolic flexibility of various bacterial species and provides the highest level of resolution currently available for the metabolic capacity of cells (Bochner *et al.* 2001). PM technology is based on cellular respiration that reflects cell activity (Bochner *et al.* 2001). The PM assay chemistry uses tetrazolium violet dye that is reduced in the presence of electrons leading to a colour change from colourless to purple. In the absence of nutrient transport or metabolism, there will be no electron flow and therefore no reduction of the dye (Figure 4.1).

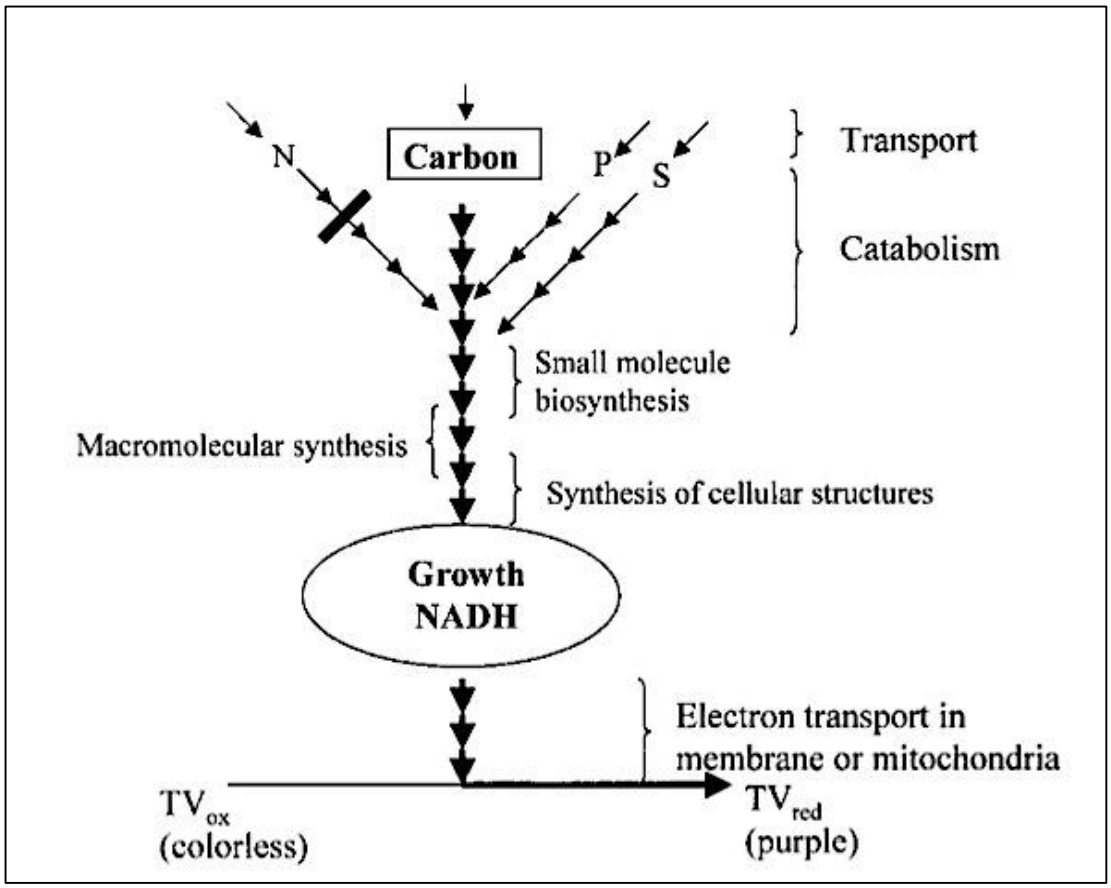


Figure 4.1: Respiration is coupled to conversion of tetrazolium violet dye (taken from (Bochner *et al.* 2001)).

4.1.4 Gene content analysis for the identification of unique ST131 loci

4.1.4.1 Bacterial "pan-genome" approach as a tool for gene content analysis

Previous analysis of sequenced bacterial genomes showed that the size and content of these genomes vary greatly, even for closely related strains (Gordienko *et al.* 2013). For example, it has been shown that only 39% of the combined, non-redundant sets of proteins were common for three sequenced *E. coli* strains (Welch *et al.* 2002). Additionally, although whole-genome sequencing can be a useful tool for studying the genetic makeup of bacterial strains (Reuter 2011), the inability to consistently assign a function to the predicted products of uncharacterised genes, such as virulence-associated genes, remains a main limitation (Rasko *et al.* 2008).

To accurately characterise and compare the gene content of large number of bacterial genomes, a pan-genome analysis approach was introduced in 2005 (Medini *et al.* 2005). The term "pan-genome" is defined as the total complement of genes from all sequenced strains of the same species (Tettelin *et al.* 2005), genus (Snipen and Ussery 2010) or a larger group (Lapierre and Gogarten 2009). Pan-genome includes a core "backbone" genome composed of genes present in all strains and a dispensable "accessory" genome composed of genes absent from one or more strains and genes that are unique to each strain (Tettelin *et al.* 2005). The accessory genome is particularly important in population studies of pathogenic bacteria, as it may contain the key virulence and antimicrobial resistance genes of the organism (Mira *et al.* 2010). The basic step in pan-genome analysis is the computation of gene families (Snipen and Ussery 2010), and it can therefore provide detailed insights into bacterial evolution, adaptation, and population structure, vaccine design and the identification of virulence genes (Mira *et al.* 2010). Nowadays, there are many pipelines available for comprehensive gene content analysis using pan-genome concept such as the pan-genome analysis pipeline (PGAP) (Zhao *et al.* 2012), Large-Scale Blast Score Ratio (LS-BSR) (Sahl *et al.* 2014) and Gegenees (Ågren *et al.* 2012).

4.1.4.2 The genetic architecture of *E. coli* ST131 H30Rx clade

Avasthi and colleagues published the first *E. coli* ST131 whole genome sequence in 2011, and it was for the Indian uropathogenic *E. coli* ST131 strain NA114 (Avasthi *et al.* 2011). The NA114 chromosome had a size of 4,935,666 bp with a G + C content of 51.16% and a coding percentage of

88.4%. It also had 4,875 protein-coding sequences, 67 tRNAs and 3 rRNAs genes (Avasthi *et al.* 2011). The NA114 chromosome was shown to carry a range of VAGs, such as *malX*, *fimH*, *kpsM* and *iutA* (Avasthi *et al.* 2011), which are commonly present in strains belonging to other major ExPEC STs such as ST73 and ST95 (Guyer *et al.* 1998, Avasthi *et al.* 2011). However it also carried other VAGs that are infrequently associated with ST131 isolates such as *cnf1*, *sfa*, *pap* and an intact polyketide synthase (*pks*) island (Johnson *et al.* 2008). Additionally, the NA114 strain harboured a single plasmid that had a size of 3.5 kb, although it has not yet been analysed with respect to its antimicrobial resistance gene profiles (Avasthi *et al.* 2011).

Subsequently, whole genome sequencing for ten ST131 strains was performed by our research group (Clark *et al.* 2012). The genome of ST131 strain UTI18 was chosen as a representative for these ten genomes since it is highly antimicrobial resistance and of average invasion capacity (Clark *et al.* 2012). The UTI18 genome is similar to NA114 in that it comprises no novel regions that may account for increased fitness or pathogenicity compared to other publically available ExPEC genomes, however UTI18 lacked the pathogenicity island comprising the *cfn1*, *sfa* and the *pap* fimbrial operon in UTI18. It was also found that UTI18 genome contained an ISL3-like transposase in the *fimB* gene of the type 1 fimbrial operon (Clark *et al.* 2012), a fully intact high pathogenicity island (HPI) encoding the yersiniabactin locus, and two flagella-encoding regions (Clark *et al.* 2012).

The genetic architecture of isolates belonging to the *E. coli* ST131 H30Rx clade has recently been examined (Petty *et al.* 2014), paying particular

attention to virulence associated genes of ExPEC and to mobile genetic elements not found in non-ST131 ExPEC. In general, the data of this study suggested no ST131 specific virulence gene repertoire, though did highlight the seemingly unique nature of the secondary flagellar cluster (Flag-2) that had been previously identified in *E. coli* ST131 genomes (Clark *et al.* 2012). Additionally, the analysis also demonstrated the role of intra-ST131 recombination in shaping the lineage with a significant overrepresentation of fimbrial adhesins and motility genes in the identified recombinant fragments. It also identified a recombinant fragment common across ST131 within the group II capsule synthesis locus (Petty *et al.* 2014).

4.1.5 Aims of the study

Given that many recent reports have proposed the role of high metabolic capacity in enhancing the fitness of *E. coli* ST131 and contributing to its pathogenesis (Gibreel *et al.* 2012a, Vimont *et al.* 2012), we aimed to test and compare the metabolic activity of a collection of ExPEC strains including ST131 and non-ST131 strains using all available API metabolic profiling substrates (Biomerieux, UK). We further sought to test the metabolic activity of ten *E. coli* strains to a more comprehensive level using the phenotypic microarray technology in order to provide a comparison between ST131 and *E. coli* of different STs in terms of their metabolic potential, and to examine the correlation between the antimicrobial resistance and the metabolic activity of ST131 strains. Additionally, given that the comparative genomics performed to date on *E. coli* ST131 have mainly focussed on virulence associated genes (Petty *et al.* 2014), and in attempt to identify the presence of ST131 unique loci that may account for increased fitness or pathogenicity compared to other publically available

ExPEC genomes, we sought to investigate the presence of other loci uniquely associated with the H30Rx clade of *E. coli* ST131 through comparing the gene content of a collection of *E. coli* strains including ST131 and non-ST131.

4.2 Material and Methods

4.2.1 Bacterial strains

A collection of fifty-two *E. coli* strains was used in this study. Fifty *E. coli* strains, twenty-five ST131 and twenty-five non-ST131, were included in the metabolic profiling assay using API test reagents. Thirty six of these strains were collected between October 2008 and June 2009 from urine samples of elderly patients from Queens Medical Centre, Nottingham (Croxall *et al.* 2011b), while the other fourteen strains were collected between March 2011 and June 2011 from urine samples of patients from Queens Medical Centre, Nottingham (Alhashash *et al.* 2013). Afterwards, a subset including ten *E. coli*, five ST131 and five non-ST131 strains, was used to carry out the Biolog phenotypic microarray assay. The five ST131 strains were previously genome sequenced (Clark *et al.* 2012), and varied in CTX-M-15 gene carriage, source of isolation, and invasion levels. The non-ST131 strains were chosen to represent the major ExPEC STs associated with human disease (Welch *et al.* 2002, Chen *et al.* 2006b, Croxall *et al.* 2011b, McNally *et al.* 2013a). Table 4.1 shows full details of the strains used in this study.

Table 4.1: The characteristics of strains used to carry out the metabolic profiling and PM assays in chapter 4.

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
UTI18	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI24	ST131	Monomicrobial UTI	Urine	-	A
UTI32	ST131	Monomicrobial UTI	Urine	CTX-M-15	A
UTI39	ST131	Polymicrobial UTI	Urine	-	A
UTI62	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI144	ST131	Polymicrobial UTI	Urine	-	A
UTI149	ST131	Polymicrobial UTI	Urine	-	A
UTI182	ST131	Polymicrobial UTI	Urine	-	A
UTI188	ST131	Monomicrobial UTI	Urine	-	A
UTI190	ST131	Polymicrobial UTI	Urine	-	A
UTI226	ST131	Polymicrobial UTI	Urine	-	A
UTI233	ST131	Polymicrobial UTI	Urine	-	A
UTI263	ST131	Polymicrobial UTI	Urine	-	A
UTI275	ST131	Polymicrobial UTI	Urine	-	A
UTI306	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI383	ST131	Polymicrobial UTI	Urine	-	A
UTI423	ST131	Polymicrobial UTI	Urine	-	A
UTI445	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI453	ST131	Polymicrobial UTI	Urine	-	A
UTI514	ST131	Polymicrobial UTI	Urine	-	A
UTI524	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI531	ST131	Polymicrobial UTI	Urine	-	A
UTI555	ST131	Polymicrobial UTI	Urine	-	A
UTI570	ST131	Polymicrobial UTI	Urine	-	A
UTI587	ST131	Polymicrobial UTI	Urine	CTX-M-15	A

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
UTI4	ST95	Monomicrobial UTI	Urine	CTX-M	A
UTI14	ST10	Polymicrobial UTI	Urine	-	A
UTI50	ST73	Polymicrobial UTI	Urine	CTX-M	A
UTI56	ST69	Polymicrobial UTI	Urine	-	A
UTI147	ST73	Polymicrobial UTI	Urine	-	A
UTI220	ST73	Polymicrobial UTI	Urine	-	A
UTI243	ST69	Polymicrobial UTI	Urine	-	A
UTI284	ST73	Monomicrobial UTI	Urine	-	A
UTI337	ST95	Monomicrobial UTI	Urine	CTX-M	A
UTI396	ST393	Polymicrobial UTI	Urine	-	A
UTI501	ST69	Polymicrobial UTI	Urine	-	A
UTI565	ST69	Polymicrobial UTI	Urine	-	A
UTI89	ST95	Uncomplicated cystitis	Urine	-	B
E. coli CFT073	ST73	Acute pyelonephritis	Urine	-	C
P5B	ST10	Bacteraemia	Blood	-	D
U22	ST95	UTI	Urine	CTX-M	E
U60	ST95	UTI	Urine	CTX-M	E
U19	ST10	UTI	Urine	-	E
U95	ST10	UTI	Urine	-	E
U9	ST69	UTI	Urine	-	E
U16	ST69	UTI	Urine	-	E
U23	ST69	UTI	Urine	CTX-M	E
U21	ST73	UTI	Urine	-	E
U24	ST73	UTI	Urine	-	E
U30	ST73	UTI	Urine	-	E

A: (Croxall *et al.* 2011b); B:(Chen *et al.* 2006b); C: (Welch *et al.* 2002); D: (McNally *et al.* 2013a); E: (Alhashash *et al.* 2013).

4.2.2 Comparative metabolic profiling assays

The metabolic potential of a collection of *E. coli* strains was determined using two assays: the metabolic profiling using API test reagents and the Biolog phenotypic microarray assays.

4.2.2.1 Metabolic profiling assay using API test reagents

The metabolic profiling of fifty *E. coli* strains, twenty-five ST131 and twenty-five non-ST131 strains, was carried out using four API kits: API 20E, ID32 E, API 50 CH and API ZYM kits (Biomerieux, UK). This resulted in a total of 120 biochemical tests, [information on these substrates are given in Appendix (chapter 7)], and allowed the measurement of carbon source utilisation, carbohydrate fermentation and enzymatic activity of *E. coli* strains. Preparation of bacterial suspensions and inoculation of test kits were performed according to the manufacturer's instructions (Biomerieux, UK). The assays were performed in duplicate on two independent occasions giving completely concordant results.

4.2.2.2 Biolog phenotypic microarray (PM) assay

The PM assay was performed, with the help of Dr. Jane Newcombe (University of Surrey), using Biolog Inc. (Hayward, CA). This assay consisted of two 96 well PM panels (PM1, PM2A), which were used to test the ability of five *E. coli* ST131 and five non-ST131 strains to utilise 190 carbon sources [information on these substrates are given in Appendix (chapter 7)]. *E. coli* strains were plated out on LB agar at 37°C prior to starting the assay. The bacterial cell suspension for each strain was prepared by transferring around 20–25 bacterial colonies into a sterile tube containing 15ml of sterile dH_2O. A uniform suspension was made until a

turbidity of 42% transmittance (T) ± 1% in the Biolog turbidimeter was obtained. 2ml of this cell suspension was then added to 10ml of inoculation fluid-0 (IF-0)+120µl dye A to yield a final cell density of 85% T. Afterwards, 100µl of the 85% T cell suspension was added to each well. The plates were then placed in the OmniLog reader (Biolog), and incubated for 48h. The OmniLog reader analyses the plates every 15 minutes, converting the pixel density in each well to a signal value reflecting cell growth and dye conversion. Figure 4.2 shows the workflow of the Biolog phenotypic microarray assay.

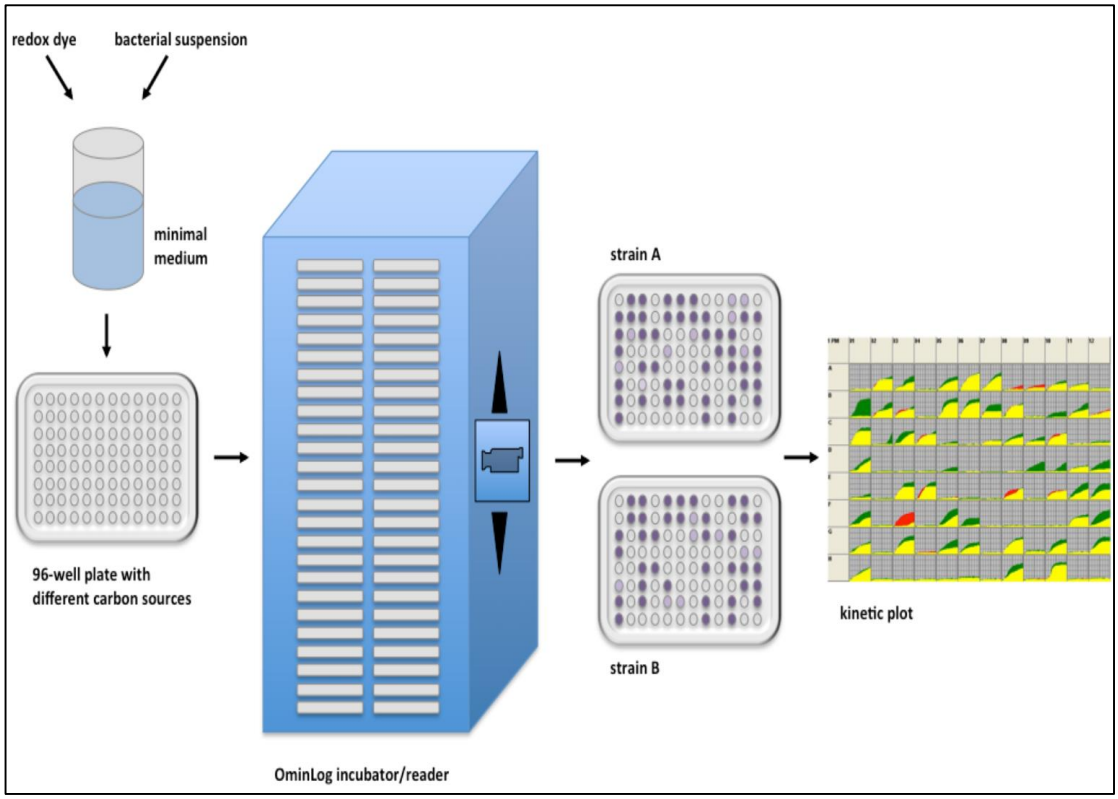


Figure 4.2: Workflow of Biolog phenotypic microarray assay [adapted from (Bochner *et al.* 2001, Bochner 2003)].

4.2.2.3 Phenotypic Microarray data analysis

Since its invention, the evaluation of phenotypic microarray data has not been standardised. Although previous studies used many approaches for phenotypic microarray data analysis (Lay Jr *et al.* 2006, Vaas *et al.* 2012),

the validity of the comparison between strains in terms of their metabolism remains one of the major challenges because substrate utilisation curves of the tested strains can show different shapes. Additionally, the possible interference of the background signal with the real utilisation signal produced by strains caused difficulties in detecting the presence of real metabolism. Therefore, this study evaluated the phenotypic microarray data set using the signal value analysis approach described previously (Homann *et al.* 2005).

4.2.2.4 Statistical analysis

The production of heat maps for API metabolic profiling results was performed using SPSS PAWS (version 20.0) statistics software. Phenotypic microarray results were analysed using R statistics package. Testing for correlation between metabolic profile and sequence type was performed by principal component analysis (PCA) in R. This part of study was performed with the help of Dr Richard Emes (University of Nottingham).

4.2.2.5 Comparative genomics for the identification of the presence or absence of metabolic-associated loci

To compare the metabolic potential between strains at a genomic level, an ST131 core genome was created to ensure any differences observed were conserved across all the ST131 strains. The ST131 genome sequences (Clark *et al.* 2012) were aligned using Mugsy (Sahl *et al.* 2011) and the core genome was extracted using Mothur (Schloss *et al.* 2009) as described previously (McNally *et al.* 2013b). The ST131 core genome was then compared against CFT073, UTI89, and P5B using ACT (Carver *et al.* 2005) in a pairwise fashion to determine metabolic loci uniquely present or uniquely absent to ST131.

4.2.3 Gene content analysis of a group of ST131 and non-ST131 strains

4.2.3.1 Bacterial genome data

A total of twenty-five publically available *E. coli* genome sequences were used in this analysis (Table 4.2). Of these genome sequences, sixteen were reference genome sequences available from National Centre of Biotechnology Information (NCBI) website, whereas nine were ST131 genomes produced during a previous study by our group (Clark *et al.* 2012). The data set of these nine ST131 genomes was used in this study.

Table 4.2: List of *E. coli* genome data used to perform gene content analysis.

Strain	ST	Pathotype	Accession number
E. coli UTI18	131[a]	ExPEC	ERP001095
E. coli EC958	131[b]	ExPEC	CAFL01000001
E. coli NA114	131[c]	ExPEC	CP002797.1
E. coli UTI24	131[a]	ExPEC	ERP001095
E. coli UTI32	131[a]	ExPEC	ERP001095
E. coli UTI62	131[a]	ExPEC	ERP001095
E. coli UTI188	131[a]	ExPEC	ERP001095
E. coli UTI226	131[a]	ExPEC	ERP001095
E. coli UTI306	131[a]	ExPEC	ERP001095
E. coli UTI423	131[a]	ExPEC	ERP001095
E. coli UTI587	131[a]	ExPEC	ERP001095
E. coli JJ1886	131	ExPEC	CP006784.1
E. coli SE15	131[a]	Human commensal	AP009378.1
E. coli LF82	135	AIEC	NC_011993.1
E. coli IHE3034	95	ST95 ExPEC	CP001969.1
E. coli UTI89	95	ST95 ExPEC	CP000243.1
E. coli S88	95	O45 ExPEC	CU928161.2
E. coli APEC01	95	APEC	CP000468.1
E. coli UM146	643	AIEC	CP002167.1
E. coli NRG857c	135	AIEC	CP001855.1
E. coli ED1a	452	O81	CU928162.2
E. coli ABU89372	73	Asymptomatic	CP001671
E. coli CFT073	73	ExPEC	AE014075.1
E. coli Di14	73	ExPEC	AE014075.1
E. coli Di12	73	ExPEC	CP002211.1

[a]2009 UK ST131 strains; [b]2004 UK ST131 strain; [c]Indian ST131 strain.

4.2.3.2 *E. coli* ST131 genome re-assembly

Due to the improvements in assembly methodology since our initial study

(Clark *et al.* 2012), the nine ST131 genomes were re-assembled using the De Bruijn graph-based de novo assembler Velvet (Zerbino and Birney 2008) and Post Assembly Genome Improvement Toolkit (PAGIT) (Swain *et al.* 2012). The De novo Velvet assembly was performed using the following parameters: a word length or k-mer length of 31, a short-paired read category, an average coverage of 75 and a coverage cut-off of 8. Afterwards, PAGIT was used to improve the quality of genome assemblies by ordering the previously assembled contigs against a reference genome (Swain 2009). In this study, the JJ1886 genome was used as a reference genome (Andersen *et al.* 2013). PAGIT was parallelised using a script written by Dr. Alan McNally including the use of four programs, ABACAS for ordering and orienting contigs previously aligned against a reference genome, IMAGE for closing the gaps within the scaffolds, ICORN for error correction at nucleotide level and RATT for annotation transfer from reference genome (Swain *et al.* 2012). All command lines used to perform this part of study are given in the appendix (Chapter 7).

4.2.3.3 Comparative gene content analysis of a group of *E. coli* ST131 and non-ST131 strains using Gegenees software

Gegenees (version 2.1.0) (Ågren *et al.* 2012) was used to compare the gene content of *E. coli* strains. Firstly, all against all fragmented alignment of all ExPEC genomes was performed and then used to look for genomic regions that are unique for the ST131 genomes. In detail, the fasta files of twenty-four ExPEC genomes (Table 4.2) were imported into Gegenees database and all against all fragmented alignment of the these genomes was performed using BLASTN method. An accurate fragmented alignment was achieved by using a fragment size of 200 and a step size of 100.

Afterwards, Gegenees alignment was used to identify genomic regions that are conserved in a specified target group (ST131 genomes) but not present in the background group (non-ST131 genomes). The ST131 JJ1886 reference genome (Andersen *et al.* 2013) was defined within the target group to provide the genome coordinates required for describing target signature and to display signatures along with genome annotation. Finally, a pan-genome was created for all ExPEC genomes (Table 4.2) representing all sequence material in these genomes. The pan-genome was exhibited by fragmenting all genomes and analysing every fragment for similarity to the pan-genome, and only unique fragment was added to the pan-genome (Ågren *et al.* 2012).

4.2.3.4 Core and pan-genome analysis of a group of *E. coli* ST131 and non-ST131 genomes using LS-BSR software

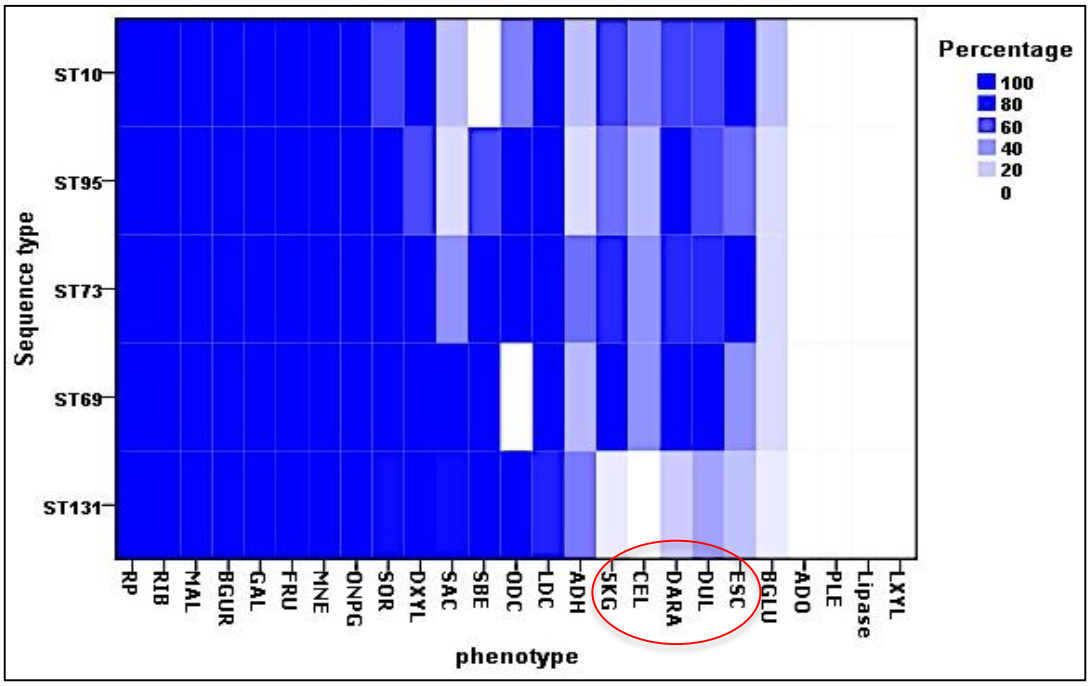
Core and pan-genome analysis of all ExPEC genomes listed in Table 4.2 was performed using LS-BSR software (Sahl *et al.* 2014). Firstly, a pan-genome for all ExPEC genomes was created by Dr. Alan McNally. Then, the "compare_BSR python" script implemented in the LS-BSR package was used to identify loci unique to genomes belonging to the ST131 H30Rx clade, with the exception of NA114 which has been shown to have known H30Rx genes missing from its assembly (Petty *et al.* 2014). Finally, the resulting loci that were only associated with ST131 H30Rx lineage were identified by performing BlastX search tool in Artemis against the genome of JJ1886 (Andersen *et al.* 2013). Information on the "compare_BSR python" script used to perform this part of study is given in the appendix (Chapter 7).

4.3 Results

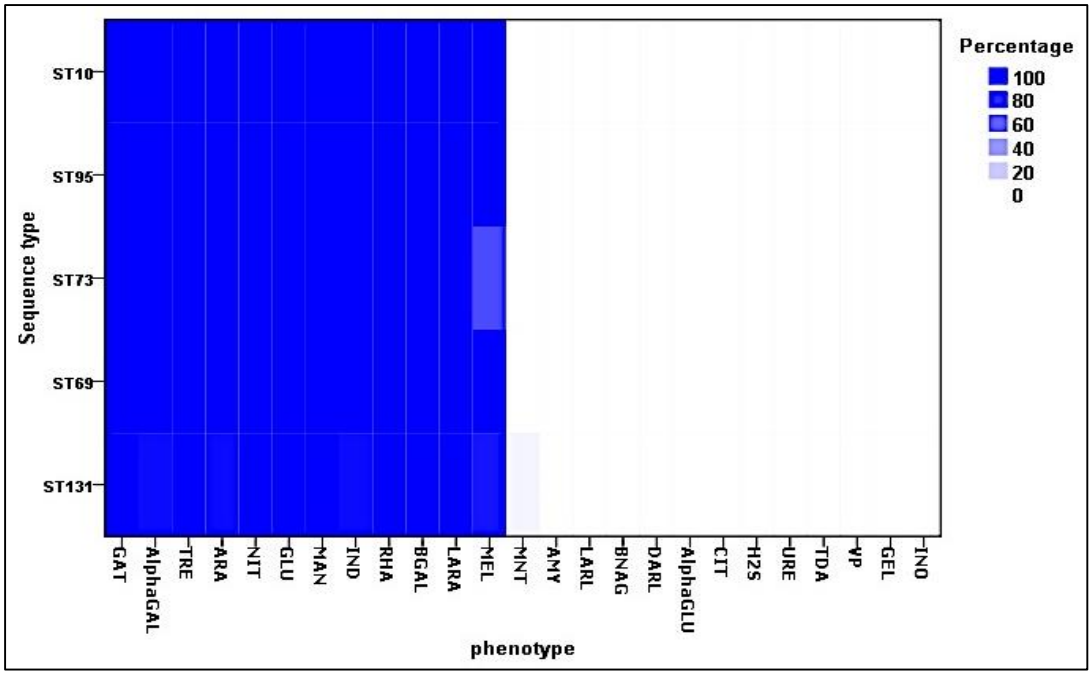
4.3.1 Metabolic activity of *E. coli* strains obtained from API test reagents

120 API test reagents were used to perform comparative metabolic profiling on twenty-five ST131 and twenty-five non-ST131 strains belonging to the four major ExPEC sequence types: ST69, ST73, ST95 and ST10. Figure 4.3 (A-D) shows a comparison of results obtained from 89 API biochemical tests for all the fifty *E. coli* ST131 and non-ST131 strains. All *E. coli* strains were positive for utilisation of thirty substrates, and negative for forty-three others, with variations between STs in terms of their capability of utilising the remaining forty-seven substrates. Examples of these differences were ST69 that was completely unable to utilise ODC, ST10 that was unable to utilise SBE and ST95 that was also unable to utilise SAC.

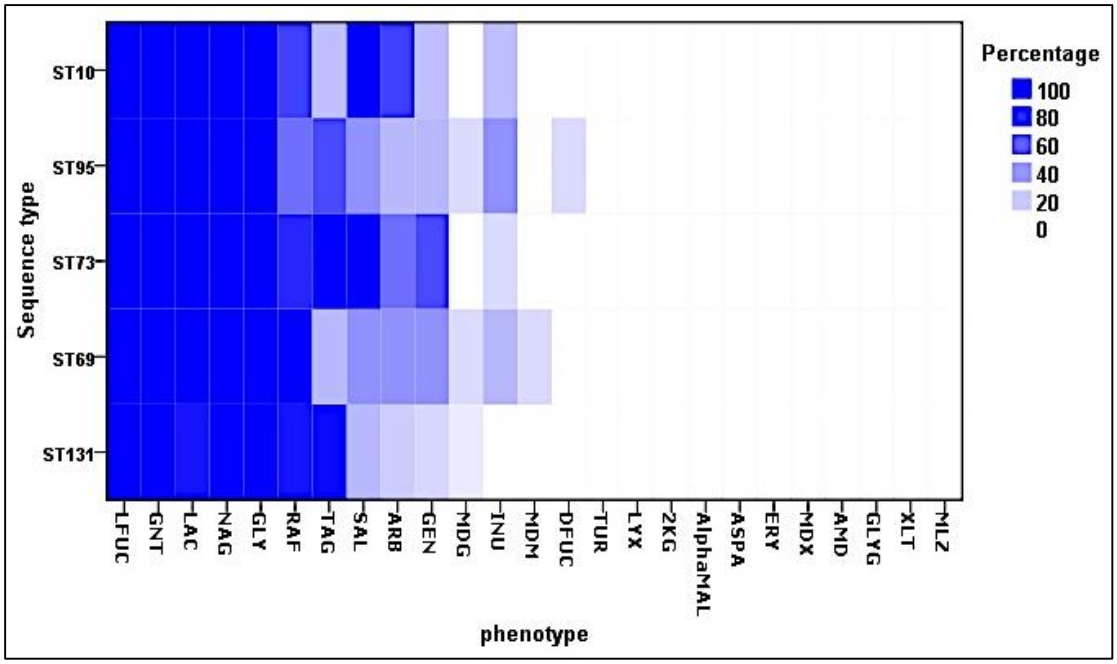
A


Sequence type
ST10
ST95
ST73
ST69
ST131
RP
RIB
MAL
BGUR
GAL
FRU
MNE
ONPG
SOR
DXYL
SAC
SBE
ODC
LDC
ADH
SKG
CEL
DARA
DUL
ESC
BGLU
ADO
PLE
Lipase
LXYL
phenotype
Percentage
100
80
60
40
20
0

B


Sequence type
ST10
ST95
ST73
ST69
ST131
GAT
AlphaGAL
TRE
ARA
NIT
GLU
MAN
IND
RHA
BGAL
LARA
MEL
MNT
AMY
LARL
BNAG
DARL
AlphaGLU
CIT
H2S
URE
TDA
VP
GEL
INO
phenotype
Percentage
100
80
60
40
20
0

C



D

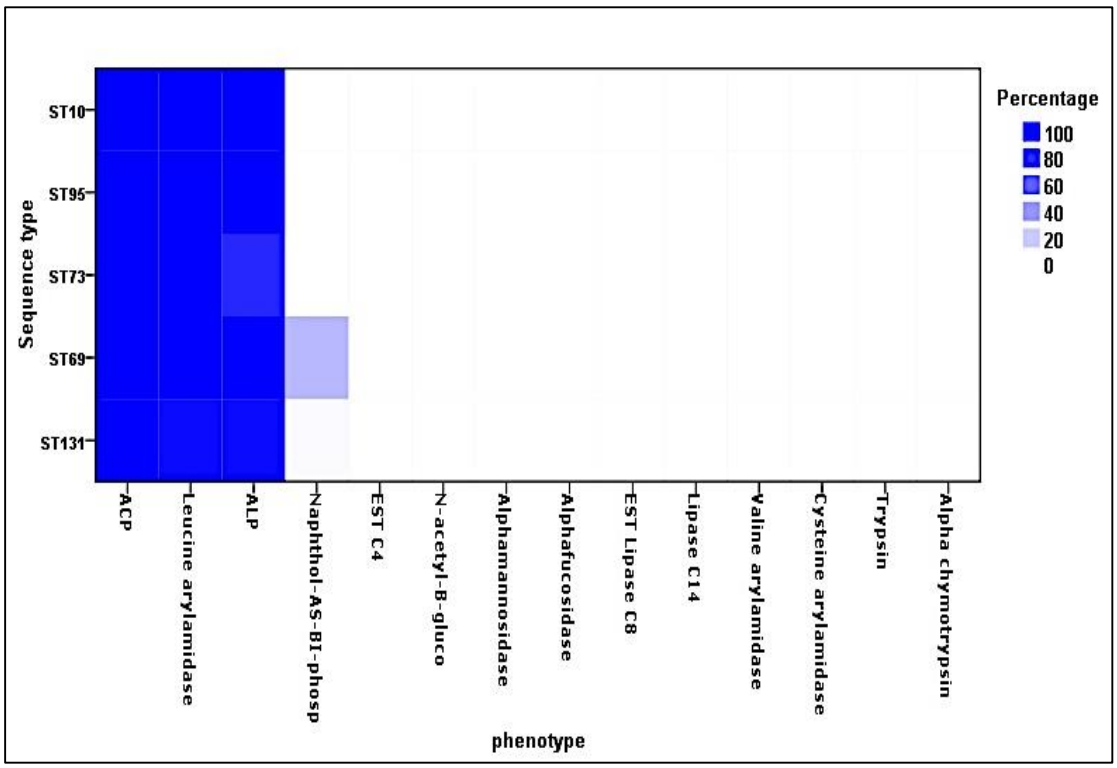


Figure 4.3 (A-D): Four heat maps show a comparison of results obtained from 89 carbon sources for *E. coli* ST131 and four major *E. coli* STs. Darker shaded areas indicate higher percentage of strains under each ST that tested positive for each phenotype. Red circle in the heat map (A) showed the substrates that were under-utilised by ST131 strains.

ST131 strains exhibited an inability to utilise five tested substrates: 5-keto-D-gluconate (5KG), D-Arabinose (DARA), esculin (ESC), cellobiose (CEL) and dulcitol (DUL) (Figure 4.3 A), in comparison to other ST strains. With regard to 5KG utilisation, only 2 of the 25 ST131 strains (8%) were able to utilise this substrate, whilst the utilisation percentages for ST69, ST73, S795 and ST10 strains were 100%, 85%, 57% and 75%, respectively. Additionally ST131 strains exhibited poor DARA utilisation compared to other ST strains. Only 20% of ST131 strains were able to utilise this substrate compared to 100% for ST69 strains, 85% for ST73 strains, 100% for ST95 strains and 75% for ST10 strains. Similarly dulcitol utilisation percentage for ST131 strains was 36% compared to 100% for ST69 strains, 85% for ST73 strains, 71% for ST95 strains and 75% for ST10 strains. The capability of ST131 strains to utilise esculin was 24% compared to 42% for ST69, 100% for ST73, 57% for ST95 and 100% for ST10 strains. The cellobiose utilisation percentage for ST131 strains was 0% compared to 42% for ST69 strains, 42% for ST73 strains, 28% for ST95 strains and 50% for ST10 strains.

ST131 strains did not show higher metabolic activity for any biochemical substrate tested when compared to other ST strains, which is in contrast to the previous study (Gibreel *et al.* 2012a). The ability of ST131 to utilise substrates for which it was more active in that study, such as ODC, βGUR and SAC, was not higher using API methodology, while some other substrates do not appear in the API substrate panel.

4.3.2 Phenotypic Microarray data

4.3.2.1 Choosing the signal value calculation approach for the PM data analysis

An example of the current phenotypic microarray data set showing a comparison between the kinetic plots of the ST131 strain UTI226, and the ST73 strain, CFT073 is given in Figure 4.4. Considering for instance the H05 well, there was a very low signal produced by the strains, which might reflect slight utilisation ability, but it was difficult to differentiate it from the background signal. This explains the reason behind the importance of subtracting the negative control signal value for each PM plate.

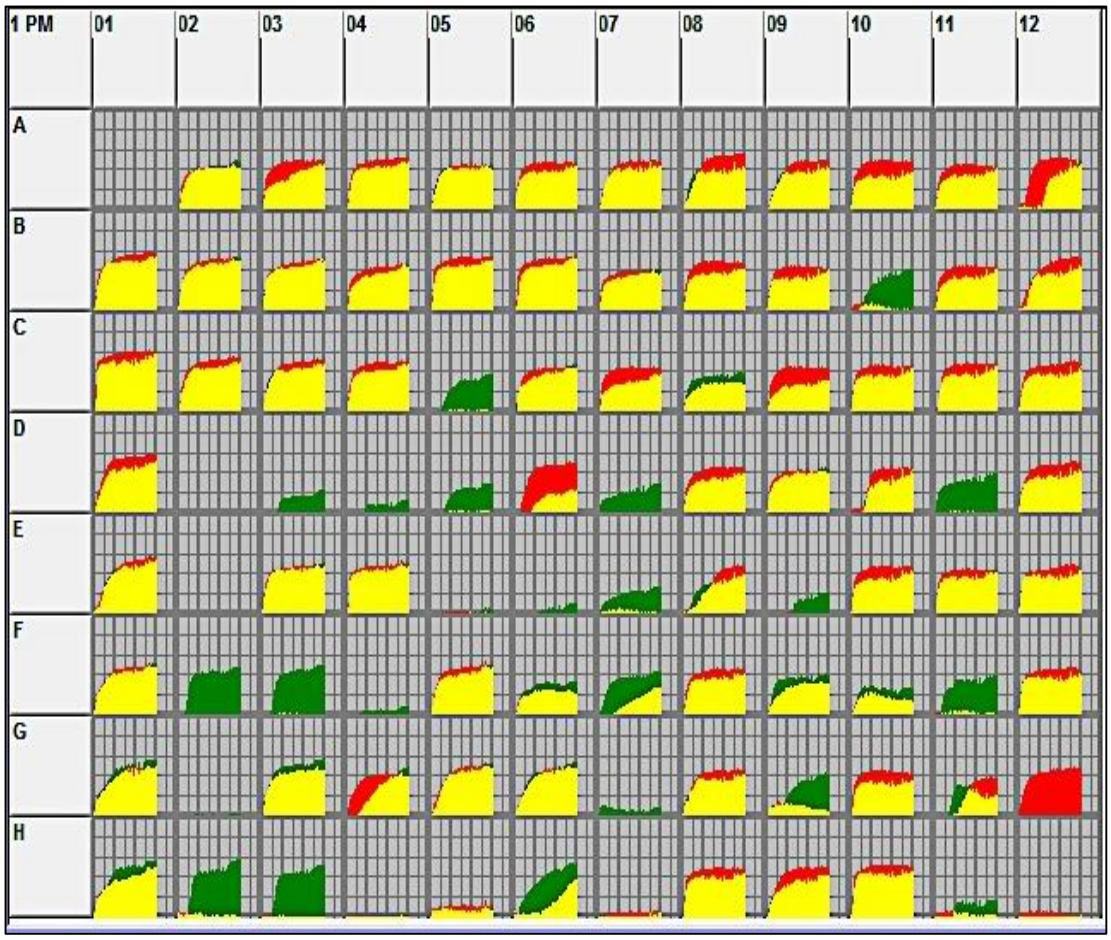


Figure 4.4: Example for utilisation of carbon sources in UTI226 and CFT073. A01 is negative control. Overlapping curves are shown in yellow. UTI226 is shown in green, CFT073 is shown in red.

In addition the curves also showed very different shapes. Some curves showed a steady rise from the start of cell respiration such as well G06, while others showed a quick rise such as well D01. Thus it was important to determine a way by which we can compare these curves and then to study the difference between them. Different approaches have been used for the analysis of PM data. For example, using the area under curve is a useful tool in comparing two utilisation curves (Vaas *et al.* 2012). However this could not be used in some cases such as in wells G04 and E08, where one strain utilised the substrate earlier while the other utilised it to a higher degree. In addition the measurement of the endpoint of the curve is also widely used for PM data analysis and it only gives a signal in the last hours

(Lay Jr *et al.* 2006), but this approach does not allow for obtaining a valid comparison between strain curves such as in well G04, where they almost have the same endpoint but their progression is completely different. For PM data analysis in our study, we chose an approach that depends on calculating the signal value to measure and compare the metabolism of strains (Homann *et al.* 2005). Firstly, the data was exported from the Biolog File Manager. Then, the average height and maximum height over 48h and the average height over 2h were extracted. The background control was then subtracted for each PM plate. The signal value was calculated using the following formula: Signal value = [(Average signal over 48h + maximum signal over 48h) /2] – average signal over 2h. This was done to firstly correct the average height for later data points, to take the progression of the curve into account and to reduce any background signals; therefore, a single number can therefore reflect the time course of metabolism.

4.3.2.2 Metabolic activity of *E. coli* strains obtained from PM

In light of the contrasting results above with a previous study showing an enhanced metabolic function for ST131 compared to non-ST131 (Gibreel *et al.* 2012a), we utilised the superior discriminatory power of the phenotypic microarray assay to fully characterise the metabolic capacity of five *E. coli* ST131 and five non-131 strains. The PM data showed that all the tested strains were positive for utilisation of seventy-nine carbon sources, and negative for thirty-six carbon sources, with variations between STs in terms of their capability of utilising the remaining seventy-three carbon sources. Among all tested strains, the ST95 strain UTI89 was the highest in terms of the metabolic activity, while the ST10 strain P5B had the lowest

metabolic capacity. Figure 4.5 shows a heat map for all metabolites tested, with values corresponding to the intensity of utilisation for the different nutrient sources. There were very few detectable differences in the metabolic capacity between ST131 and other ST strains for the majority of tested carbon sources. As with the commercial test reagents there were some strain-specific differences in metabolite utilisation but these were not consistent in either ST131 or non-ST131 groups, and ST131 strains were not associated with a specific metabolic profile (Figure 4.5).

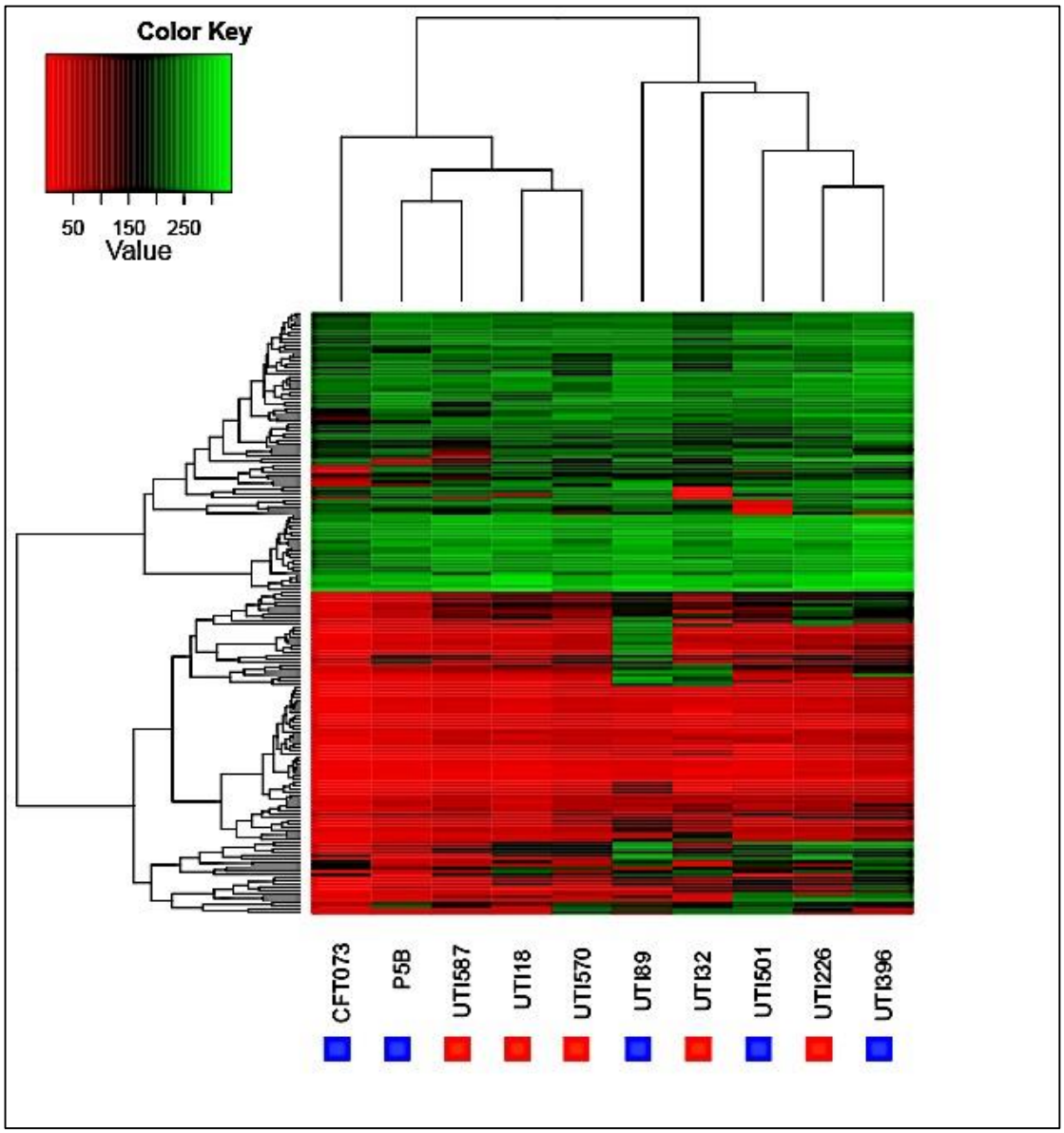


Figure 4.5: A cluster heat map showing the signal values of *E. coli* ST131 and non-ST131 strains obtained from 190 biochemical tests using PM assay, with red showing no utilisation through to green showing high levels of utilisation. A UPGMA dendrogram informed by the metabolic profile is presented above the heat map. The ST131 strains are represented by red blocks and the non-ST131 strains by blue blocks.

Considering the metabolic activity of CTX-M-15 producing ST131 strains, UTI32 showed a slightly higher metabolic potential in comparison to other ST131 strains. It was the only strain to utilise malonic acid and L-phenylalanine, and its signal values were the highest for a range of substrates such as M-inositol, palatinose, D-tagatose, N-acetyl-L-glutamic acid, L-histidine and putrescine compared to the rest of ST131 strains. In

contrast, it was the only *E. coli* strain that was unable to utilise N-acetyl-D-galactosamine, and its signal values for D-raffinose and L-malic acid were lower than that of all other tested ST131 and non-ST131 strains. UTI587 showed a lower capacity of utilising the following substrates: dulcitol, glycyl-L-glutamic acid, 3-0-b-d-galacto-pyranosyl and melibionic acid. Additionally, the UTI18 strain had the lowest signal value for α-Keto-butyric acid substrate among the tested ST131 strains.

However the ST131 CTX-M-15 negative strain UTI226 had the highest metabolic activity among the tested ST131 strains for pyroglutamic acid, chondroitin sulfate c, β-cyclodextin, amygadin and gentiobiose carbon sources. In contrast, it had the lowest ability to utilise β-hydroxy-butyric acid compared to other ST131 strains. Taken together, our data is strongly indicative that ESBL gene carriage is not associated with a specific metabolic activity profile within the ST131 strains, and that variation is on a more generic strain to strain level.

With respect to the metabolic activity of non-ST131 strains, UTI89 exhibited an increased ability to metabolise some carbon sources compared to other strains, and was the only strain capable of utilising thirteen carbon sources such as tyramine, gelatin, xylitol, G-amino-butyric acid and D, L-octopamine. Additionally, the signal values of UTI89 were higher than that of other strains for D-Arabitol, L-Arabitol and 3-methyl glucose. However the ST10 strain P5B was of low metabolic capacity compared to all other ST131 and non-ST131 strains. This was shown by its inability to utilise 102 out of the 188 carbon sources used in the PM plates. The ST73 strain

CFT073 was also unable to utilise 101 out of the 188 carbon sources, and it was also exclusively incapable of metabolising three carbon sources.

In short, although there were some differences in the metabolic potential between strains, these differences were very much strain-specific and were not detected at a sequence type level. The generation of a UPGMA dendrogram based on utilisation of metabolites confirmed this, with the ST131 and non-ST131 strains equally dispersed throughout the dendrogram (see Figure 4.5), providing strong evidence that ST131 are not a metabolically distinct group of ExPEC strains. Table 4.3 shows a summary of the strain-specific differences in the metabolic activity of the ST131 and non-ST131 strains as obtained by our PM analysis.

Table 4.3: Summary of the strain-specific differences in the metabolic activity of the ST131 and non-ST131 strains as obtained by our PM analysis.

Strain	ST	Exclusive positive tests	Exclusive negative tests
CFT073	ST73	None	Tween 40 Mono methyl succinate L-lyxose
P5B	ST10	None	α-keto-glutaric acid
UTI89	ST95	Tyramine b-Phenylethylamine Gelatine Glycogen Inulin D-fucose 3-methyl glucose Xylitol G-amino-butyric acid Quinic acid 4-Hydroxy-L-Proline L-Ornithine D, L-Octopamine	None
UTI396	ST393	None	None
UTI501	ST69	None	L-galactonic acid -g-lactone Succinamic acid
UTI18	ST131	None	None
UTI32	ST131	Malonic acid L-phenylalanine	N-acetyl-D-galactosamine
UTI226	ST131	None	None
UTI570	ST131	None	None
UTI587	ST131	None	None

4.3.2.3 Principal component analysis of PM data

To further confirm that ST131 are not a metabolically distinct group of ExPEC strains, we performed principal component analysis (PCA) on the phenotypic microarray metabolic profile data set. PCA is a technique, which uses sophisticated mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components (Richardson 2009). The PCA1/PCA2 plot (Figure 4.6)

shows clearly that ST131 strains are not grouped based on metabolic properties, but rather are dispersed throughout the PCA plot amongst the non-ST131 strains. In conjunction with the heat-map based dendrogram our data provides strong evidence that *E. coli* ST131 are not a metabolically distinct clade of ExPEC.

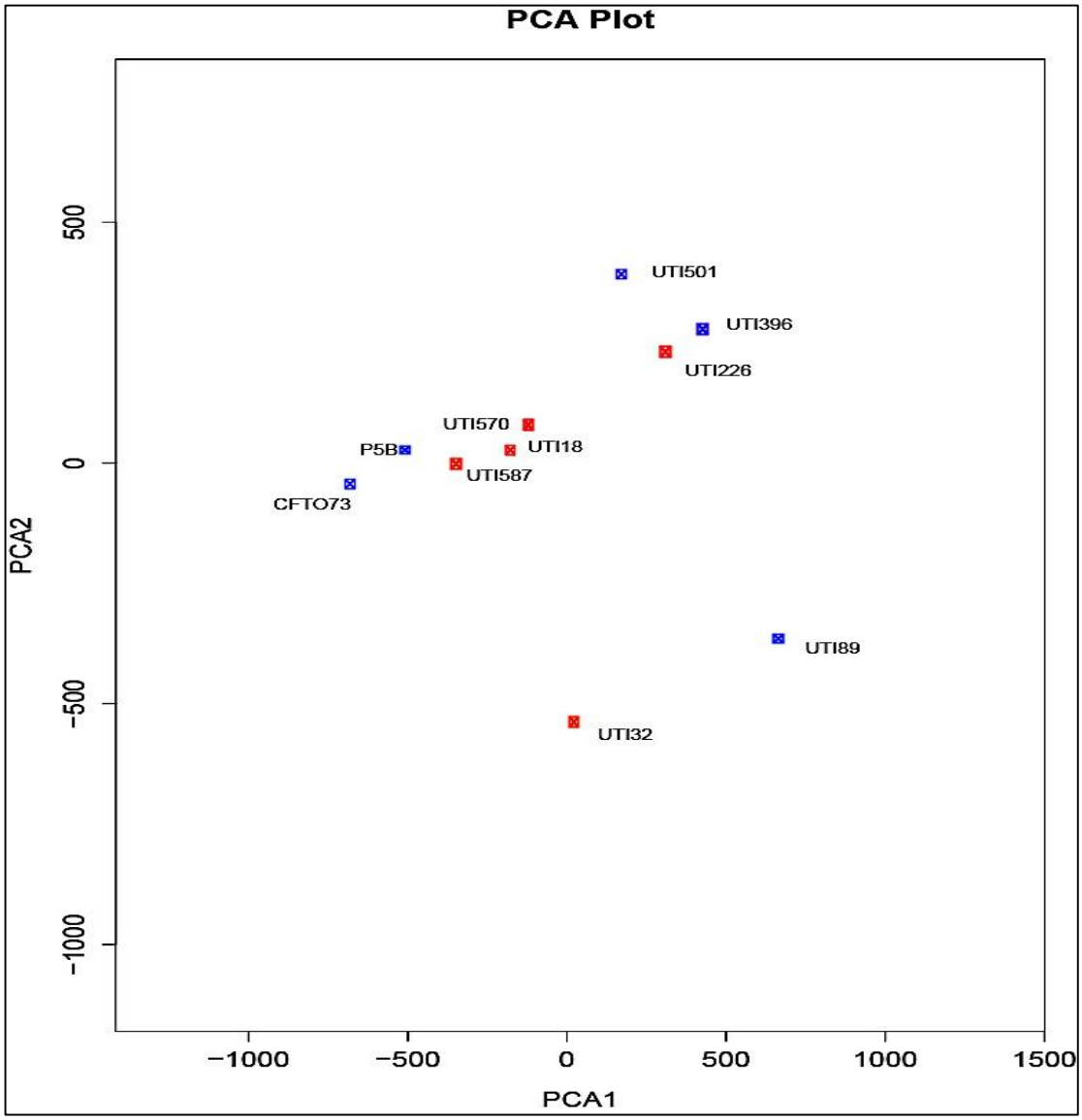


Figure 4.6: Principal component analysis for metabolic profiles obtained from an analysis of the phenotypic microarray data set. The ST131 strains are denoted by red blocks and the non-ST131 strains by blue blocks.

4.3.2.4 Differences in the carbon source utilisation by ST131 strains obtained from using two measurement methods

Since some carbon sources that are included in API test panels are also present in the PM plates, we sought to determine if there is a difference in the capacity of ST131 strains to utilise these carbon sources as a result of using two different testing methodologies. The capacity of ST131 strains to utilise twelve carbon sources that are available in API test panels and in PM plates were compared, and the data showed a similar utilisation capacity of these strains in 9 out of 11 carbon sources (Table 4.4). All the ST131 strains were able to metabolise 6 out of 11 carbon sources, while all ST131 strains were unable to metabolise the remaining three carbon sources: inulin, gelatine and D-Arabitol. However there were few detectable differences between strains in the utilisation of 2 carbon sources: cellobiose and dulcitol. As mentioned earlier, our metabolic profiling data using API method found that the ST131 strains had the lowest capacity to utilise these substrates. The PM data showed that all but one ST131 strains were able to utilise cellobiose, which is in contrast to metabolic profiling data. Additionally, PM data found that ST131 strains were capable of utilising dulcitol to a higher level than that in API method.

Table 4.4: Comparison of sugar utilisation results of ST131 strains obtained from using two testing methods.

Test	Biotyping result	PM result	Agreement
Glucose	+	+	Yes
Sucrose	+	+	Yes
L-Rhamnose	+	+	Yes
Cellobiose	-	+ (expect in one ST131 strain)	No
D-Xylose	+	+	Yes
D-Mannitol	+	+	Yes
Dulcitol	ST131 had the lowest utilisation capacity among other STs	ST131 strains had a low utilisation capacity but higher than that of biotyping	No
L-Fucose	+	+	Yes
Inulin	-	-	Yes
Gelatine	-	-	Yes
D-Arabitol	-	-	Yes

4.3.2.5 Relating the metabolism observations to the presence/absence of ST131 associated metabolic loci

Given that the vast majority of the strains analysed by phenotypic microarray in our study have been genome sequenced (Clark *et al.* 2012, McNally *et al.* 2013b), we sought to further confirm our metabolism observations by correlating them to gene presence/absence data at a whole genome level. We adapted an approach recently used to successfully identify clade-specific metabolic functions in *Campylobacter* (Sheppard *et al.* 2013) by constructing an ST131 core genome and then comparing this against the CFT073, UTI89, and P5B genomes using ACT (Carver *et al.* 2005), concentrating on genes and operons with predicted or confirmed metabolic functions.

We found only 3 differences in the *idn, ydd,* and *asc* operons, which we have previously reported (Clark *et al.* 2012). These three operons are involved in utilisation of two of the five metabolites (5KG and cellobiose) identified as being under-utilised in ST131 strains by API test panels. The

idn operon is a subsidiary pathway for the *gntII* gluconate metabolism system, and the *yddA* and *yddB* genes are transporters for the *gntI* gluconate system, whilst the *asc* operon encodes a combined arbutin/salicin/cellobiose uptake and metabolism pathway. Figure 4.7 shows ACT comparisons of the three loci from CFT073 and the core ST131 genome. Gene content differences relating to the other three differential metabolites could not be found. Therefore, our data suggests that presence or absence of loci does not correlate well with metabolic profile. This indicates that the subtle differences that have been observed in metabolism of substrates in ST131 compared to other ExPEC STs may not be clearly distinguishable at a genome level and may be down to discreet mutations in other upstream or downstream metabolic pathways which then impinge on metabolism of the identified differential substrate.

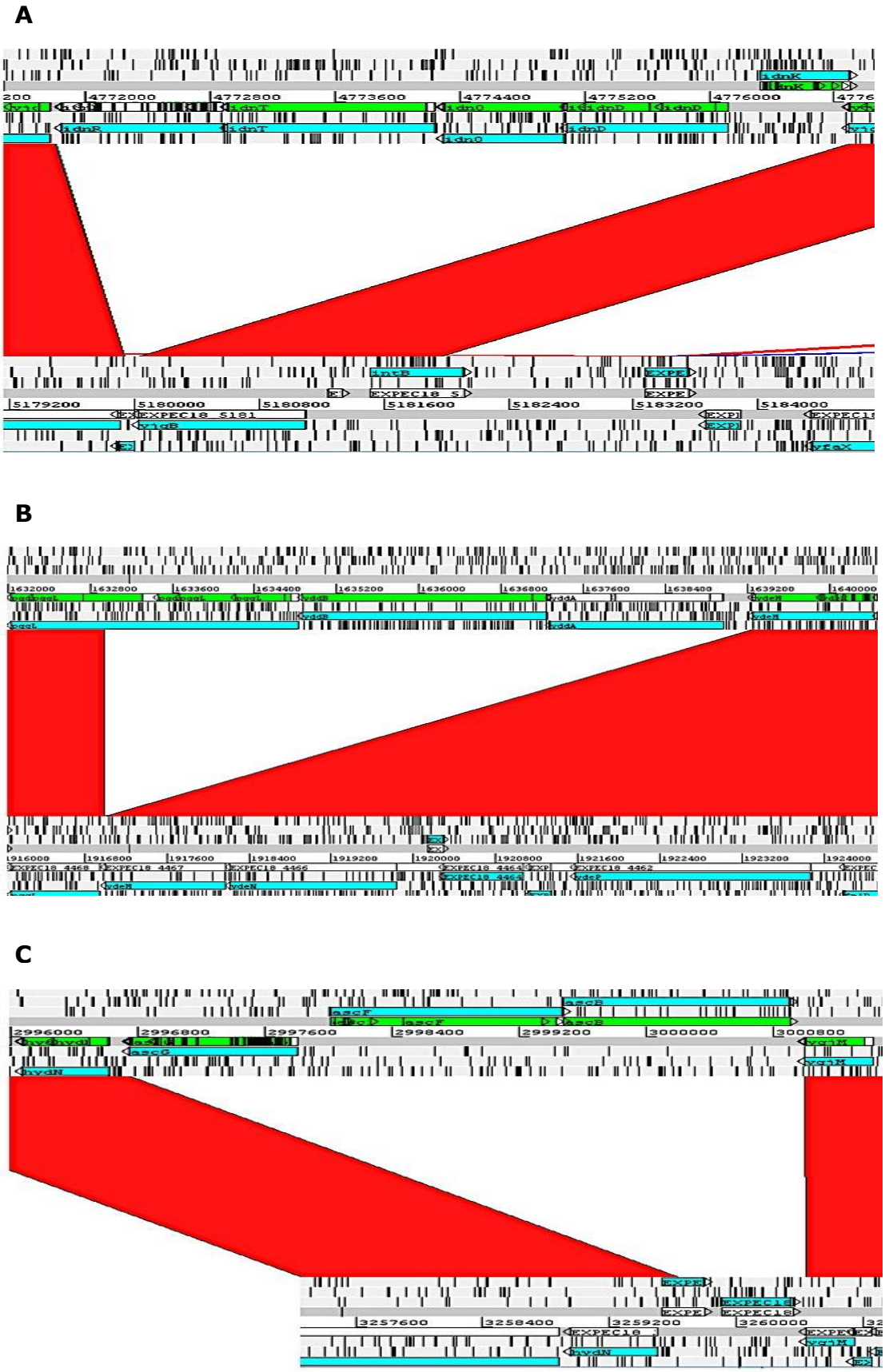


Figure 4.7 (A-C): ACT comparisons of the (A) *idn* (B) *ydd* and (C) *asc* loci from CFT073 and the core ST131 genome.

4.3.3 Gene content analysis of *E. coli* strains

4.3.3.1 *E. coli* ST131 genome re-assembly data

The whole genome sequences for nine *E. coli* ST131 strains (Clark *et al.* 2012) were re-assembled using Velvet (Zerbino and Birney 2008) and PAGIT (Swain *et al.* 2012) in order to improve the quality of ST131 genome assemblies prior to using them in the gene content analysis. Our final genome assembly data showed that the size of all ST131 genomes ranged from 5Mbp to 5.5Mbp (Table 4.5). Additionally, the number of contigs of ST131 genomes ranged from 57 to 787 contigs, while the N50 values of these contigs ranged from 10987 to 191894bp (Table 4.5).

Table 4.5: Final genome assembly statistics for the nine *E. coli* ST131 strains genomes used in this study.

Strain ID	Sequencing platform	Total contig length (bp)	N50 of contigs (bp)	Number of contigs
UTI18	lllumina 454	5150118	14394	592
UTI24	lllumina	5272012	95669	144
UTI32	lllumina	5203330	119581	110
UTI62	lllumina	5307017	50730	229
UTI188	lllumina	5182230	12544	657
UTI266	lllumina	4958223	191894	57
UTI306	lllumina	5292101	122653	112
UTI423	lllumina	5262577	10987	787
UTI587	lllumina	5454260	13172	691

4.3.3.2 Comparative genome analysis of *E. coli* strains using Gegenees software

In order to examine and compare a collection of ExPEC genomes including ST131 and non-ST131 (Table 4.2), Gegenees software (Ågren *et al.* 2012) was used. The overall relationship between the ExPEC genomes based on performing all against all fragmented alignment was determined and shown in Figure 4.8. For the twelve ST131 genomes, the data demonstrated a high average similarity between them ranging from 85% to 99%. Similarly, the average similarity between the twelve non-ST131 genomes was high and it ranged from 80% to 99%. However our data shows the presence of some differences in the average similarity between genomes when compared with each others (Figure 4.8). For example when the UTI18 genome was compared with the UTI32 genome, the heat plot shows different average similarity between these strains. The average similarity between UTI18 and UTI32 was 93% while it was 94% when the UTI32 genome was compared with the UTI18 genome. This might indicate the presence of limitations in Gegenees software and call into question its reliability as a tool for comparative genome analysis.

A phylogenomic overview of the average similarity between genomes in the alignment was also visualised in a heat-plot (Figure 4.8). All ST131 genomes were grouped together at the top of the heat-plot. The non-ST131 genomes were phylogenomically grouped together at the bottom of the heat-plot. Additionally, the heat-plot showed a further subgrouping for non-ST131 genomes belonging to particular STs and this was observed for the four genomes belonging to ST73 (*E. coli* CFT073, *E. coli* ABU89372, *E. coli* Di12 and *E. coli* Di14), the four genomes belonging to ST95 (*E. coli* S88, *E.*

coli APEC01, *E. coli* UTI89 and *E. coli* IHE3034) and the two genomes belonging to ST135 (*E. coli* LF82 and *E. coli* NRG857c).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1: 18	100	96	96	95	93	94	90	92	87	93	94	95	76	79	80	79	79	79	78	79	78	77	76	76
2: 62	95	100	97	94	93	93	90	92	87	92	94	96	76	79	80	79	79	79	78	79	78	77	76	76
3: 306	96	98	100	96	94	94	90	93	87	93	95	97	76	79	80	79	79	79	78	79	78	78	77	77
4: 24	95	96	96	100	94	94	91	93	88	94	94	95	77	79	81	80	80	79	79	79	79	78	77	77
5: 32	94	95	95	95	100	96	92	94	88	95	95	95	77	81	82	81	81	80	79	79	80	79	78	78
6: EC958	96	97	97	97	98	100	93	96	90	96	96	97	79	82	83	82	82	81	80	81	81	80	79	79
7: NA114	97	97	97	97	97	97	100	97	91	96	97	97	81	84	85	84	84	84	84	84	83	83	81	81
8: JJ1886	98	99	98	99	98	99	96	100	91	98	98	98	81	84	85	84	84	84	83	83	83	82	81	81
9: 226	93	93	93	93	93	93	90	93	100	92	93	93	80	84	84	84	84	83	85	85	84	84	83	83
10: 188	92	93	93	93	93	93	89	92	86	100	94	93	75	79	80	78	78	78	77	78	78	78	76	76
11: 423	91	93	93	91	91	91	88	90	84	92	100	93	74	76	78	77	77	76	75	76	76	75	74	74
12: 587	90	92	92	90	88	89	85	88	82	88	90	100	72	75	76	75	75	75	74	75	74	74	72	72
13: E.coliED1a	79	80	80	80	80	80	79	80	77	80	80	80	100	83	83	82	82	83	83	82	82	82	80	80
14: CFT073	82	82	82	82	83	83	82	83	81	83	82	82	83	100	95	95	95	86	86	87	88	88	83	83
15: E.coliABU83972	84	84	84	85	84	85	83	85	82	84	84	84	83	95	100	96	96	85	85	87	89	89	84	84
16: E.coliDi14	84	85	85	85	85	86	84	85	84	85	85	85	84	98	98	100	100	87	87	90	90	90	86	86
17: E. coliDi2	84	85	85	85	85	86	84	85	84	85	85	85	84	98	98	100	100	87	87	90	90	90	86	86
18: E.coliS88	84	85	85	85	84	84	84	85	83	84	84	85	86	89	87	87	87	100	96	94	94	93	86	86
19: E.coliAPECO1	83	83	83	83	82	83	83	83	84	82	83	83	84	88	87	86	86	95	100	95	94	93	87	86
20: E.coliIHE3034	83	84	83	84	82	83	83	84	84	82	83	84	83	89	88	88	88	92	94	100	97	96	86	86
21: UTI89	82	83	83	83	83	83	82	84	83	84	82	83	84	90	90	89	89	92	94	97	100	98	86	87
22: E.coliUM146	83	84	83	84	84	84	83	84	84	84	83	83	85	91	91	90	90	93	94	97	99	100	87	87
23: E.coliLF82	85	86	86	86	86	86	85	86	88	85	85	86	88	91	90	90	90	91	92	92	92	91	100	99
24: E.coliNRG857C	86	86	86	87	86	87	85	87	88	86	86	86	88	91	91	91	91	91	92	92	92	92	99	100

Figure 4.8: A heat-plot showing the average similarity between ST131 and non-ST131 genomes, based on a fragmented alignment using BLASTN made with settings 200/100 and using the default cut-off threshold for non-conserved material. Phylogenomic grouping for the ExPEC genomes is represented on the left of the heat-plot and the ST131 genome ID was shown in green while the non-ST131 genome ID was shown in red.

A Gegenees alignment for the ExPEC genomes was then mined to determine the unique regions associated only with ST131 genomes. The basis for the signature analysis was to compare the conservation pattern between the target (ST131) and background (non-ST131) groups for every fragment from the selected reference genome. All fragments were then given a biomarker score that can be plotted along the reference genome (Ågren *et al.* 2012). In order to obtain the stringent form of the biomarker score, a fragment's maximum score against the background group (worst false positive) was divided by the minimum BLAST score against the target group genomes (worst false negative) and finally subtracted from 1 to yield a

value ranging from 1 (perfect conservation and no cross-reaction) to -∞ (Ågren *et al.* 2012). Prior to performing signature analysis in Gegenees, ST131 genomes were selected as a target group, non-ST131 genomes were selected as a background group, and the ST131 JJ1886 genome was used as a reference genome. Figure 4.9 shows a signature analysis of the target group (ST131 genomes) using two stringent forms of biomarker score: (maximum/minimum) that showed only the fully unique regions for the target group and (maximum/average) that showed regions in the signature which might be absent or less conserved in a portion of the target group. Interestingly, our data demonstrated the presence of few regions unique to ST131, and the green lines in the tubular view were indicative of this observation (Figure 4.9).

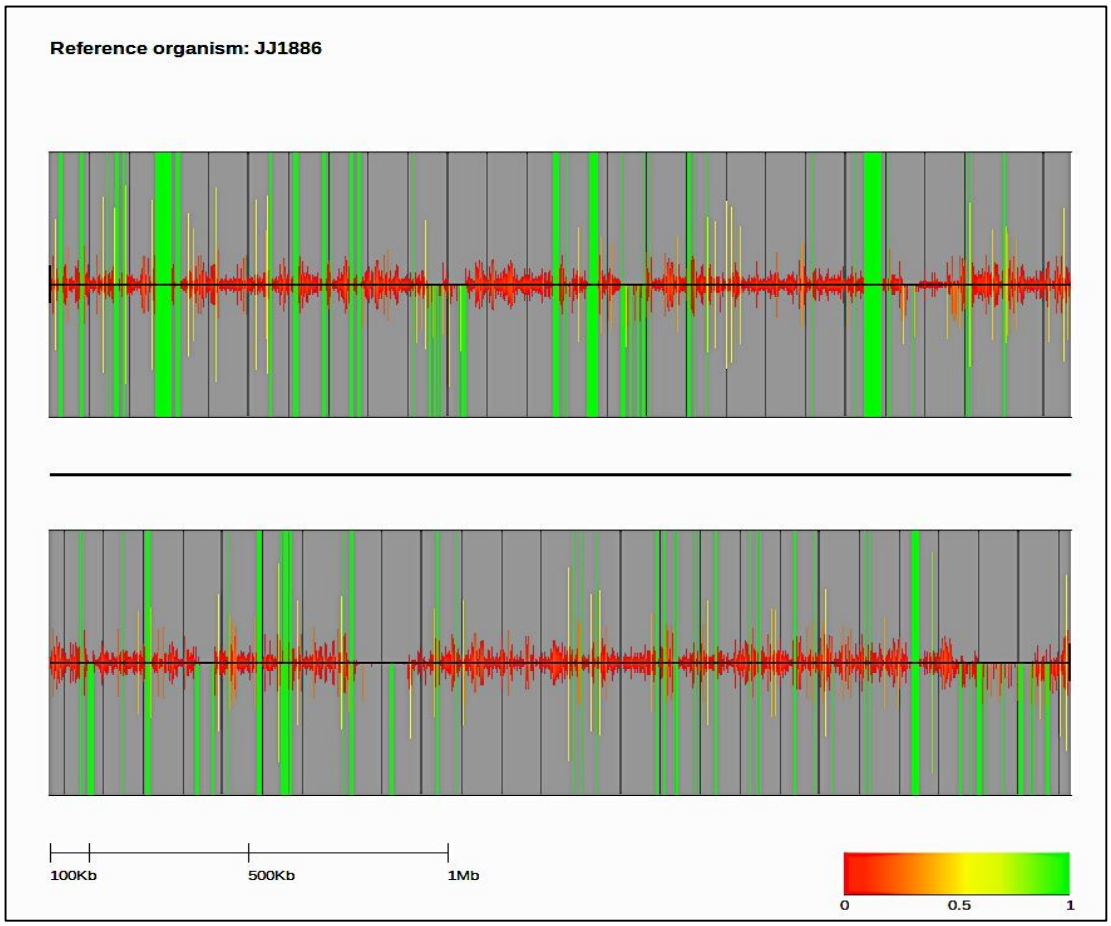


Figure 4.9: A tubular view showing the signature analysis of ST131 genomes. A fragmented alignment was firstly performed with 200/100 settings using BLASTN (BLAST+). Target group was defined to include the ST131 genomes and the background group was defined to include the non-ST131 genomes. The "maximum background/minimum target" and "maximum background/average target" settings were used for biomarker score calculation. The ST131 genome JJ1886 was used as a reference genome. Scale bar represents the size of the reference genome. The colour bar represents the biomarker score values where 1 is the highest biomarker score value and shown by green colour and 0 is the lowest biomarker score value and shown by red colour.

Afterwards, a pan-genome for all the twenty-four ExPEC strains was created representing all sequence material of the ExPEC strains. The pan-genome statistics showed that it contained 43871 fragments and 8774200 nucleotides. Figure 4.10 shows the pan-genome created for the ExPEC strains. The horizontal black line represents the pan-genome, while the vertical lines crossing it separate the sub-sequences of the pan-genome. The horizontal lines below the pan-genome line represent the genomes used in the analysis. If a fragment of the pan-genome is similar enough

(i.e. a BLAST score over the draw parameter, which was adjusted to 0.99) to a genome, Gegenees draws a dot over the genome. Several dots in a row become a peak. The colour of the peak indicates the mean distance over the draw parameter for all fragments represented by that peak. Thus, green peaks represent the similarity of the fragments between the strains and the pan-genome. The pan-genome viewer (Figure 4.10) showed that the core genome of ST131 strains formed more than 50% of overall size of the pan-genome. Additionally, comparing the similarity peaks of ExPEC genomes showed the presence of different gene content of ST131 compared to that of non-ST131.

Figure 4.10: A pan-genome viewer showing the pan-genome created for the 24 ExPEC strains. The horizontal black line at the top represents the pan-genome while the vertical lines crossing it separate the sub-sequences of the pan-genome. The horizontal lines below the pan-genome line represent the genomes used in the analysis. The draw parameter used in this run was 0.99.

4.3.3.3 Identification of genetic loci unique to the *E. coli* ST131 H30Rx clade as shown by pan-genome analysis

Given the focus on virulence-associated genes in previous gene content studies (Petty *et al.* 2014), and on the basis of our aforementioned findings obtained from using Gegenees showing the presence of some regions unique to ST131 H30Rx strains as well as a different gene content of these strains, this study aimed to determine genetic loci unique to ST131 H30Rx strains without bias for functionality of the encoding loci. An ExPEC pan-

genome was constructed using the blast-score ratio method implemented in LS-BSR (Sahl *et al.* 2014) containing twelve ST131 genomes and all available non-ST131 reference genome sequences (Table 4.2). Using the resulting pan-genome matrix, the genetic loci uniquely associated with the ST131 group were determined versus the non-ST131 group using the compare_BSR python script implemented in the LS-BSR package. To define the ST131 group, the genome of NA114 was excluded on the basis that previous work has suggested the methodology used to assemble the genome has resulted in regions missing from that genome which are present in all other H30Rx strains (Petty *et al.* 2014). The analysis was performed with SE15 as a non-ST131 H30Rx strain to determine loci unique to the H30Rx clade to which SE15 does not belong (McNally *et al.* 2013b, Petty *et al.* 2014). The resulting data set identified a total of 150 loci unique to ST131 H30Rx strains in comparison to other ExPEC (Table 4.6), dominated by three phages common across the lineage which have most probably been acquired by the common ancestral H30Rx progenitor and then maintained in the lineage.

Table 4.6: List of loci identified as unique to *E. coli* ST131 H30Rx strains by our pan-genome analysis.

Locus label	Gene product	Accession number
centroid_24082	Flagellar export protein (FliJ)	YP_ 006137319.1
centroid_24111	Flagellar basal body-associated protein (FliL)	YP_006137345.1
centroid_15348	Hypothetical protein	YP_006141838.1
centroid_31178	Hypothetical protein	YP_006140981.1
centroid_15320	Hypothetical protein	YP_006139096.1
centroid_16303	Putative TRAP-type transport system	YP_006140118.1
centroid_20468	Putative phage tail protein PTP	YP_006139077.1
centroid_37714	Hypothetical protein	YP_003348574.1
centroid_21100	Hypothetical protein	YP_006139841.1
centroid_92180	Hypothetical protein	YP_006138463.1
centroid_32019	Hypothetical protein	YP_006138454.1
centroid_34817	Hypothetical protein	YP_002293511.1
centroid_92780	Hypothetical protein	YP_006139086.1
centroid_29292	Hypothetical protein	YP_006138557.1
centroid_32007	Hypothetical protein	YP_006138465.1
centroid_82289	Putative altronate hydrolyse (PAH)	YP_006137747.1
centroid_24504	Putative altronate hydrolyse (PAH)	YP_006137748.1
centroid_24000	Putative fimbrial usher protein (PFUP)	YP_006137236.1
centroid_35546	Sucrose operon repressor (SOR)	YP_670644.1
centroid_69872	Glycosyl hydrolase, Sucrose-6-phosphate hydrolase (S-6-PH)	YP_006139921.1
centroid_71285	Putative manganese transporter protein (mntH)	YP_006141326.1
centroid_18755	Flagellar hook protein (FliD)	YP_006137341.1
centroid_29649	Hypothetical protein	YP_006138914.1
centroid_82164	Transketolase C-terminal subunit	YP_006137619.1
centroid_23066	Ribose ABC transport ATP-binding component	YP_006137615.1
centroid_39946	Putative baseplate phage assembly protein (PBAP)	YP_006139083.1
centroid_33272	Hypothetical protein	YP_006137366.1
centroid_39971	Hypothetical protein	YP_008722088.1
centroid_67103	Hypothetical protein	YP_006137152.1
centroid_23723	Hypothetical protein	YP_006137366.1
centroid_19102	Flagellar motor protein (motA)	YP_001742390.1
centroid_92163	Hypothetical protein	YP_008721382.1
centroid_30405	Hypothetical protein	YP_008723083.1
centroid_239	Flagellar assembly protein (FliH)	CP006784.1
centroid_82288	3-oxoacyl-ACP reductase (OAR)	YP_006137746.1
centroid_67710	Putative demethylmenaquinone methyltransferase (PDM)	YP_006137744.1

Locus label	Gene product	Accession number
centroid_25685	Hypothetical protein	YP_006139098.1
centroid_83611	Hypothetical protein	YP_006141849.1
centroid_25661	Hypothetical protein	YP_006139076.1
centroid_29383	Hypothetical protein	YP_002292777.1
centroid_68972	Hypothetical protein	WP_001554155.1
centroid_34308	Hypothetical protein	WP_001531630.1
centroid_39984	Hypothetical protein	YP_008722100.1
centroid_3465	Putative phage DNA-packaging protein (Phage DNA-PP)	YP_006139087.1
centroid_91089	Flagellar protein (FliS)	YP_006137342.1
centroid_42743	Hypothetical protein	YP_006138449.1
centroid_4705	Hypothetical protein	YP_006137153.1
centroid_19083	Flagellar basal body rod protein (FlgC)	YP_001726316.1
centroid_29752	Putative bacteriophage lambda head decoration protein (BLHDP)	YP_006139089.1
centroid_11234	Putative DNA-binding protein (DNA-BP)	YP_006140930.1
centroid_24109	Hypothetical protein	YP_006137343.1
centroid_14081	Flagellar motor switch protein (FliN)	YP_006167278.1
centroid_34813	Hypothetical protein	YP_006139097.1
centroid_33222	Flagellar-hook basal body complex protein (FliE)	YP_006137314.1
centroid_34291	Hypothetical protein	YP_006138445.1
centroid_22108	Putative permease	YP_006140852.1
centroid_10576	Putative replication protein P (PRPP)	YP_006141845.1
centroid_18855	Putative transport protein (PTP)	YP_006137103.1
centroid_34792	Putative phage tail sheath protein (PPTSP)	YP_006139078.1
centroid_8361	Hypothetical protein	YP_003348575.1
centroid_33515	Lateral flagellar chaperon protein (FlgN)	YP_008720325.1
centroid_3436	HP	YP_006141842.1
centroid_91474	Putative transcriptional regulator (PTR)	YP_006137743.1
centroid_32613	Hypothetical protein	YP_001742765.1
centroid_67290	Putative lateral flagellar hook assembly protein (FlgD)	YP_006137328.1
centroid_18750	Putative RNA polymerase sigma factor for flagellar operon (FliA)	YP_006137346.1
centroid_6096	Hypothetical protein	YP_006139085.1
centroid_39980	Putative replication protein (PRP)	YP_006141844.1
centroid_16833	Hypothetical protein	YP_006140650.1
centroid_37776	Ribose ABC transporter ATP-binding component (Ribose ATP-bind co)	YP_006137613.1
centroid_29736	Hypothetical protein	YP_006139074.1

Locus label	Gene product	Accession number
centroid_24127	Hypothetical protein	YP_003348238.1
centroid_81857	Putative glycosyltransferase (PGT)	YP_006137321.1
centroid_90982	Putative outer membrane protein (omp)	YP_006137235.1
centroid_86055	Putative type 1 fimbriae anchoring protein (fimD)	YP_006141432.1
centroid_92998	Putative molybdate metabolism regulator (PMMR)	YP_006139310.1
centroid_29737	Hypothetical protein	YP_008722053.1
centroid_33503	Flagellar M-ring protein (FliF)	YP_006137315.1
centroid_3844	Putative oxidoreductase (ORD)	YP_006137610.1
centroid_24075	Putative Flagellar motor switch protein (FliM)	YP_006137311.1
centroid_9932	Hypothetical protein	YP_006138455.1
centroid_27155	Putative transcriptional regulator (PTR)	YP_006140851.1
centroid_67711	Alcohol dehydrogenase (ADH)	YP_006137745.1
centroid_9922	Hypothetical protein	YP_006138446.1
centroid_17334	Hypothetical protein	YP_006141106.1
centroid_18765	Flagellar basal-body rod protein (FlgF)	YP_006137331.1
centroid_82304	Transcriptional regulator of succinyl CoA synthetase operon (TRSCoA)	YP_006137761.1
centroid_27365	Hypothetical protein	YP_006141053.1
centroid_69020	Adenylosuccinate synthase (ADSS)	YP_006141843.1
centroid_14918	AraC family transcriptional regulator (AraC)	WP_001296050.1
centroid_19090	Hypothetical protein	YP_006137335.1
centroid_20505	Hypothetical protein	YP_006141848.1
centroid_24102	Flagellar hook-associated protein (FlgK)	YP_006137336.1
centroid_16115	PTS system, sucrose-specific IIBC component	YP_670642.1
centroid_26480	Maltoporin protein (MPP)	YP_006139919.1
centroid_33398	Putative glycosyl hydrolase (PPH)	YP_006137102.1
centroid_67730	Putative glycosyl hydrolase (PPH)	YP_006137763.1
centroid_2535	Hypothetical protein	YP_006140853.1
centroid_8303	Putative transcriptional regulator (PTR)	YP_006137617.1
centroid_3845	Putative oxidoreductase (ORD)	YP_006137611.1
centroid_82163	Transketolase N-terminal subunit (TK)	YP_002383623.1
centroid_68323	Phage integrase	YP_006138452.1
centroid_67582	Ribose ABC transporter permease component (RiboseABCTP)	YP_006137612.1
centroid_9391	Dihydrodipicolinate synthase (DHDPS)	YP_006137151.1

Locus label	Gene product	Accession number
centroid_27018	Hypothetical protein	YP_006140681.1
centroid_34807	Hypothetical protein	YP_006139092.1
centroid_29757	Hypothetical protein	YP_008722074.1
centroid_68331	Hypothetical protein	YP_006138458.1
centroid_38627	Hypothetical protein	YP_006137364.1
centroid_15349	Hypothetical protein	YP_006141837.1
centroid_81902	Putative phage integrase (PPI)	YP_006137368.1
centroid_82896	Hypothetical protein	YP_003349147.1
centroid_20507	Hypothetical protein	YP_006141846.1
centroid_9573	Anti-sigma-28 factor (FlgM)	YP_006137324.1
centroid_29781	Hypothetical protein	YP_006141840.1
centroid_19324	Hypothetical protein	YP_006137575.1
centroid_68999	Hypothetical protein	YP_008722077.1
centroid_84549	Fructokinase (FK)	YP_003350555.1
centroid_9611	Hypothetical protein	YP_006137365.1
centroid_34823	Hypothetical protein	YP_008722089.1
centroid_38568	Flagellar biosynthesis protein (FlhB)	YP_006137306.1
centroid_38574	Hypothetical protein	YP_006137313.1
centroid_82907	Hypothetical protein	YP_008721389.1
centroid_12469	Hypothetical protein	YP_006137698.1
centroid_27243	Hypothetical protein	YP_006140929.1
centroid_245	Hypothetical protein	WP_001531418.1
centroid_10531	Hypothetical protein	YP_006139073.1
centroid_20473	Putative phage baseplate assembly protein (PBAP)	YP_006139081.1
centroid_23065	Hypothetical protein	YP_006137614.1
centroid_15304	Putative phage tail protein (PTP)	YP_006139080.1
centroid_9128	Transposase IS30	YP_008721392.1
centroid_68324	Hypothetical protein	YP_006138453.1
centroid_9926	Hypothetical protein	YP_003349153.1
centroid_3468	Hypothetical protein	YP_006139084.1
centroid_36304	Hypothetical protein	YP_006140680.1
centroid_25680	Hypothetical protein	YP_008722073.1
centroid_29782	Hypothetical protein	YP_006141839.1
centroid_81841	Lateral flagellar export/assembly protein (FliQ)	YP_001726334.1
centroid_42732	Hypothetical protein	YP_006138459.1
centroid_39935	Hypothetical protein	YP_008722051.1
centroid_69012	Hypothetical protein	YP_006141850.1
centroid_29742	Hypothetical protein	YP_006139079.1
centroid_19913	Hypothetical protein	YP_006138460.1
centroid_12407	PTS system, fructose-specific IIABC component	YP_006137762.1
centroid_6770	Hypothetical protein	YP_006139974.1
centroid_13920140	Hypothetical protein	YP_006137154.1
centroid_23067	Glycerol kinase (GK)	YP_006137616.1
centroid_23136	Chaperone protein (hscC)	YP_006137685.1
centroid_67323	Hypothetical protein	YP_006137363.1
centroid_4877	Flagella basal body P-ring formation protein (FlgA)	YP_008720327.1

4.3.3.4 Distribution of functional categories of genes unique to *E. coli* ST131 H30Rx strains

With respect to the distribution of functional categories of genes unique to the ST131 H30Rx strains, our analysis showed that hypothetical proteins were the dominant functional category and they constituted 52.6% of the 150 unique genes. Interestingly, the Flag-2 accessory flagella locus genes were the second most common functional category and they constituted 13.3% of the total genes, followed by a small number of metabolic genes and they constituted 11.3% of the total 150 genes (Figure 4.11). Additionally, we identified a very small set of genes encoding for fimbrial proteins, transcriptional regulators, phage proteins and other proteins (Figure 4.11).

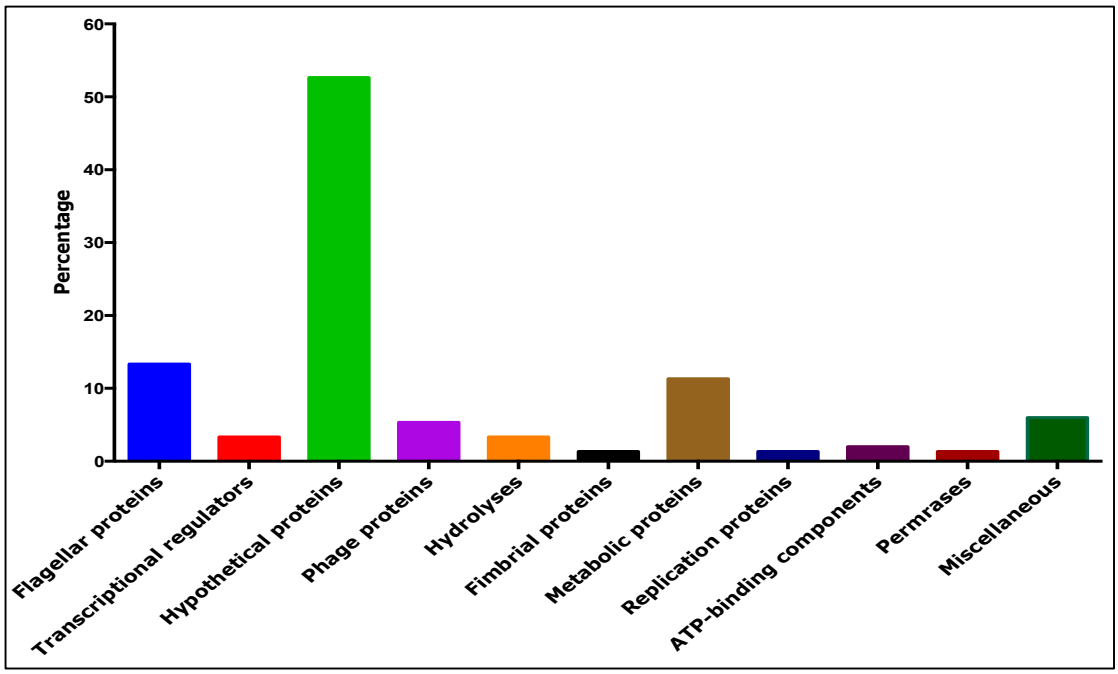


Figure 4.11: Distribution of functional categories of genes unique to *E. coli* ST131 H30Rx compared to a collection of ExPEC reference genomes.

With respect to the flagellar genes, our pan-genome analysis identified the presence of 20 flagellar genes (13.3%) unique to the ST131 strains belonging to H30Rx (Table 4.7). These genes encode for a second flagellar locus (Flag-2 locus) that has also been identified in several bacterial species including *Vibrio parahaemolyticus* (Stewart and McCarter 2003), *E. coli* (Ren *et al.* 2005), *Yersinia enterolitica* (Bresolin *et al.* 2008), *Citrobacter rodentium* (Petty *et al*. 2010) and *Aeromonas hydrophila* (Canals *et al.* 2006). The vast majority of the Flag-2 genes, such as *flhB, fliAJLHSNEFQM* and *flgCNFADKM*, are responsible for the synthesis of the flagellar basal body and hook. While one gene, the *motA* gene encoding the H^+ motor protein A, is involved in motility (Parthasarathy *et al.* 2007).

Table 4.7: The flagellar genes identified as unique to the ST131 H30Rx strains and their products.

Flagellar gene name	Gene product
fliJ	Flagellar export protein (FliJ)
fliL	Flagellar basal body-associated protein (FliL)
fliD	Flagellar hook protein (FliD)
motA	Flagellar motor protein (motA)
fliH	Flagellar assembly protein (FliH)
fliS	Flagellar protein (FliS)
flgC	Flagellar basal body rod protein (FlgC)
fliN	Flagellar motor switch protein (FliN)
fliE	Flagellar-hook basal body complex protein (FliE)
flgN	Lateral flagellar chaperon protein (FlgN)
flgD	Putative lateral flagellar hook assembly protein (FlgD)
fliA	Putative RNA polymerase sigma factor for flagellar operon (FliA)
fliF	Flagellar M-ring protein (FliF)
fliM	Putative Flagellar motor switch protein (FliM)
flgF	Flagellar basal body rod protein (FlgF)
flgK	Flagellar hook-associated protein (FlgK)
flhB	Flagellar biosynthesis protein (FlhB)
fliQ	Lateral flagellar export/assembly protein (FliQ)
flgA	Flagella basal body P-ring formation protein (FlgA)
flgM	Anti-sigma-28 factor (FlgM)

With regard to the presence of Flag-2 locus in *E. coli* strains, a previous whole-genome based study on the enteroaggregative *E. coli* 042 strain demonstrated the presence of Flag-2 locus in this strain, although this locus was found to be inactive due to a frameshift mutation in the *lfgC* gene (*flgC* in other genera and in our data set). This gene encodes for a basal body rod protein that is required for protein export through the outer membrane (Ren *et al.* 2005). However some studies have shown the absence of the Flag-2 locus in many other *E. coli* strains. For example, the genome analysis of the *E. coli* K-12 MG1655 strain has found a deletion of a large region containing 44 genes encoding for the entire Flag-2 locus (Ren *et al.* 2005). This deletion was also identified in some other commensal, intestinal pathogenic and extraintestinal pathogenic *E. coli* strains (Ren *et al.* 2005). Taken together, these findings suggest that the entire Flag-2 locus was originally present in the last common ancestor of the species, which is thought to be the enteroaggregative *E. coli* 042 strain (Ren *et al.* 2005), although the reason of the Flag-2 locus deletion found in many *E. coli* strains remains unknown. The presence or absence of Flag-2 locus in different *E. coli* strains is shown in Table 4.8.

Table 4.8: Summary showing the presence/absence of Flag-2 locus in *E. coli* strains as shown by previous whole-genome sequencing studies (Ren *et al.* 2005).

***E. coli* strain**	**Pathotype**	**Flag-2 locus**
MG1655	Commensal	Absent
ATCC 8739	Commensal	Absent
W3110	Commensal	Absent
DH10B	Commensal	Absent
O157:H7 EDL933	EHEC	Absent
O157:H7 Sakai	EHEC	Absent
042	EAEC	Present, but inactive
E2348/69	EPEC	Absent
CFT073	ExPEC	Absent

4.4 Discussion

It has been known that metabolism is an important factor in triggering bacterial colonisation in any given environment particularly in animal and human hosts (Rohmer *et al.* 2011). Several studies have demonstrated the role of metabolism in enhancing the colonisation and virulence of *E. coli*. For example, a recent study has shown that enterohaemorrhagic *E. coli* (EHEC) uses fucose to modulate its virulence and metabolic genes (Pacheco *et al.* 2012).

A recent study focussed on the metabolic potential of *E. coli* ST131, the most commonly isolated cause of ExPEC infections worldwide and a recently emerged clone of ExPEC associated with multi-drug resistance (Gibreel *et al.* 2012a). The study examined the ability of ST131 drug resistant and sensitive isolates to utilise 47 biochemical substrates in comparison to non-ST131 ExPEC strains, using the Vitek2 Advanced Expert System for metabolic profiling. In total 300 UPEC isolates were compared, with the study concluding that ST131 isolates have higher metabolic potential in comparison to other ST isolates on the basis of their ability to utilise more of the substrates than non-ST131 isolates. The study showed a significant association between ST131 isolates and utilisation of eight biochemical tests including those for peptidase, decarboxylase, and alkalinisation activity. Additionally, the study also described a correlation between metabolic activity and antibiotic susceptibility profiles, with ESBL carrying, multi-drug resistant isolates showing the highest metabolic potential (Gibreel *et al.* 2012a).

In the study we report here, we tested and compared the metabolic activity of a collection of ExPEC strains including ST131 and non-ST131 strains, employing another commonly used commercial metabolic profiling system (Biomerieux, UK). Our data agrees with that of Gibreel *et al.* (2012a) in that analysis of a limited number of metabolites can produce differential profiles for different sequence types. Our study showed reduced metabolic capacity of ST69 to utilise ODC, of ST95 to utilise SAC and of ST10 to utilise SBE.

With respect to ST131, our data showed lower metabolic activity for five substrates, namely 5-keto-D-gluconate (5KG), D-Arabinose (DARA), esculin (ESC), cellobiose (CEL) and dulcitol (DUL) in comparison to other major ExPEC STs. Some of these findings are in direct agreement with the Gibreel study that also demonstrated a significant negative association between ST69 and ODC, between ST95 and SAC, and between ST131 and 5KG (Gibreel *et al.* 2012a). However the major discussion point of the Gibreel study, that ST131 are more metabolically flexible and ‘‘fit’’ is in disagreement with the data produced using API biochemical tests. Of the Vitek positive association tests which also feature in the API test panel (ODC, βGUR and SAC), none showed higher levels of metabolism in ST131 strains compared to non-ST131 strains. The most likely explanation for this is due to differences in the way substrate utilisation is measured between the fully automated Vitek system and the API test strip method. The validity of this as a potential explanation for the discrepancy is highlighted in this study by comparison of our results for some carbon source utilisation in ST131 when measured by API and by phenotypic microarray. For example, our PM data showed very little difference in dulcitol utilisation between

ST131 and other ST strains, in complete contrast to the API data. Additionally, our PM data demonstrated that all but one ST131 strains were able to utilise cellobiose, which is in opposition to the API data. This is almost certainly due to the fact that the PM plates are used to measure carbon source oxidation and not sugar fermentation as in the API test (Durso *et al.* 2004). Therefore, we suggest that different utilisation capabilities between strains could be observed depending on the principle of assay used to determine the metabolic activity.

Using a total of 47 biochemical tests on 300 UPEC strains, the Gibreel study concluded ST131 strains are metabolically more flexible (Gibreel *et al.* 2012a). However our data of 120 biochemical tests on 50 strains is suggestive of ST131 strains having slightly reduced metabolic potential. In an attempt to further determine the metabolic profile of ST131 *E. coli* compared to other ExPEC STs, we compared ten strains, five ST131 and five non-ST131 strains, using Biolog phenotypic microarray technology. Phenotypic microarrays have been utilised to study the metabolic flexibility of various bacterial species (Bochner 2003) and provides the highest level of resolution currently available for the metabolic capacity of cells. Heat map visualisation of our data indicates there is no distinct metabolic signature for *E. coli* ST131 and that any differences in metabolic repertoire are at an individual strain level rather than by sequence type grouping. This is further supported by a principal component analysis of the microarray data, showing there is no clustering of ST131 based on metabolic repertoire. An obvious caveat to our findings is the limited number of strains examined. However our comparison is of five ST131 strains, all of which have previously been shown to be genetically homogeneous and

phylogenetically clustered with the NA114 and EC958 genome sequenced strains (Avasthi *et al.* 2011, Totsika *et al.* 2011). Therefore, if ST131 were a metabolically distinct clade of ExPEC with enhanced metabolic potential we would expect to see some form of clustering of our 5 strains by principal component analysis of the phenotypic microarray data.

Given that the genome sequences of almost all of the strains subjected to phenotypic microarray are available, we sought to contextualise our metabolism findings at a whole genome level. We created an ST131 core genome as previously described by our group (McNally *et al.* 2013b) the rationale being that any ST131-discriminating metabolic repertoire would have an accompanying genetic signature associated with the ST131 clade. This was then compared to the genomes of non-ST131 strains CFT073, UTI89, and P5B. Our data failed to uncover any substantial differences in the metabolic gene repertoire of ST131 compared to the three other ExPEC genomes. However when we focussed on operons responsible for the 5 discriminatory tests identified by API (5KG, DARA, esculin, cellobiose and dulcitol) we could find genetic deletions that would account for only two of those phenotypes, in the *idn* and *ydd* operons involved in 5KG metabolism and the *asc* operon involved in cellobiose uptake. Our data appears to suggest that differences in metabolic repertoire between bacteria cannot simply be mapped to genome data and gene presence or absence. It is likely that small mutations in pathways complexly linked to the metabolic function observed could result in knock on effects which would be extremely difficult to pinpoint and associate with the phenotype under investigation.

It is our opinion that the previous study of *E. coli* ST131 metabolism

performed by Gibreel *et al.* (2012a) excellently identifies metabolic markers that could have enormous importance in rapid identification or selection of ST131 isolates in human samples. In our study, we concurred that the use of a limited number of biochemical tests can produce differential profiles for different sequence types. Our metabolic profiling data using API tests supports this as we found clear differences in the metabolic activity between ST131 and non-ST131 strains for 5 of 120 biochemical substrates. However when we comprehensively tested the global metabolic repertoire of a limited number of *E. coli* strains using the phenotypic microarray system, we found no difference in the overall metabolic fitness between ST131 and non-ST131 strains. In addition, comparison of genomic data also suggested very little difference in the repertoire of metabolic gene loci between ExPEC sequence types.

Our previous work describing the genetic homogeneity of ST131 (Clark *et al.* 2012) combined with the data presented here provide yet another confounding hypothesis that *E. coli* ST131 do not display altered metabolic fitness to other closely related ExPEC from a global metabolic and genomic perspective. An obvious caveat to both our findings and those of Gibreel *et al.* (2012a) is that different strain sets have been compared using different methodologies, and indeed this may be applicable to many of the dichotomous results in the literature regarding ST131. We propose that a coordinated international effort to fully understand the evolutionary mechanisms behind the emergence of *E. coli* ST131 is now imperative in order to combat this most serious of bacterial pandemics, and of future episodes of novel *E. coli* lineage emergence.

With regard to the gene content analysis of ST131, recent genomic studies have led to a deep understanding of the phylogeography of this lineage of ExPEC (Price *et al.* 2013, Petty *et al.* 2014) and the discovery of a sub-clade of ST131 which is globally dominant and associated with the CTX-M-15 genotype which has been termed the H30Rx clade (Price *et al.* 2013). Despite these extensive studies, the only efforts at comparative genomics of the ST131 lineage have focussed solely on virulence associated genes and large mobile genetic elements unique to this lineage, and this was achieved by creating a pan-genome for 99 ST131 strains which was then compared with other non-ST131 genomes (Petty *et al.* 2014). Here we present an approach where we created an ExPEC pan-genome for 24 ST131 and non-ST131 strains and then identified loci uniquely associated with the ST131 H30Rx clade.

Firstly, all the *E. coli* ST131 genomes produced by our group (Clark *et al.* 2012) were re-assembled using Velvet and PAGIT assemblers in order to improve the quality of ST131 genome assemblies prior to using them in the gene content analysis. The genome assembly data identified no difference between the size of the ST131 genomes and that of some previously published genomes of ST131 strains EC958 and JJ1886 (Totsika *et al.* 2011, Andersen *et al.* 2013). With respect to comparative genome analysis of a group of ST131 and non-ST131 strains, Gegenees software was used to perform all against all alignment for these strains in order to determine the phylogenomic relationship between these genomes, and our data showed that strains belonging to each ExPEC ST are phylogenomically grouped together. Mining the Gegenees alignment for *E. coli* strains identified the presence of few unique regions associated only with ST131 genomes. The

pan-genome created for all twenty-four ExPEC strains showed the presence of different gene content in ST131 compared to non-ST131 genomes.

In order to identify the unique loci associated with ST131 and the functional categories they belong to, we present an approach where we created an ExPEC pan-genome using LS-BSR and then identified loci uniquely associated with the ST131 H30Rx clade. Although the pan-genome analysis identified the presence of 150 loci uniquely associated with ST131 H30Rx clade, our data is further suggestive that at a gene content level this clade is rather unremarkable in comparison to other ExPEC, as recently suggested for the clade at a metabolic level (Alqasim *et al.* 2014a), with the secondary flagellar locus Flag-2 the standout region unique to ST131 within ExPEC.

Many previous whole-genome based studies examined the presence of Flag-2 locus in *E. coli* strains, and found that it was present in the enteroaggregative *E. coli* 042 strain although it was inactive due to a frameshift mutation in the *lfgC* gene encoding for a basal body rod protein (Ren *et al.* 2005). However others demonstrated the absence of the Flag-2 locus in many other *E. coli* strains such as *E. coli* K-12 MG1655 as well as in some other commensal, intestinal pathogenic and extraintestinal pathogenic *E. coli* strains (Ren *et al.* 2005). Taken together, these findings suggest that the entire Flag-2 locus was originally present in the last common ancestor of the species, which is thought to be the enteroaggregative *E. coli* 042 strain (Ren *et al.* 2005), although the reason of the Flag-2 locus deletion found in many *E. coli* strains remains unknown.

Our analysis showing the unique presence of secondary flagellar locus Flag-2 loci in the ST131 H30Rx strains corroborates with some previous reports

that identified these loci in the ST131 H30Rx clade amongst ExPEC strains (Totsika *et al.* 2011, Clark *et al.* 2012, Petty *et al.* 2014). Additionally, it is likely that this is ancestral to the ST131 H30Rx clade but its acquisition within the larger ST131 lineage is suggestive of a possible role in the formation and dissemination of the H30Rx clade and merits a fuller bacterial genetics investigation of its importance and role in the H30Rx clade. Given the major role of flagella in the pathogenesis of *E. coli* strains (Haiko and Westerlund-Wikström 2013), the possession of functional Flag-2 locus may be a key factor for the survival of these strains in specific niches and/or hosts (Kirov 2003).

Additionally, our data identified the presence of a small set of metabolic loci that confirms our previous finding demonstrating that ST131 H30Rx is not a metabolically distinct lineage of ExPEC (Alqasim *et al.* 2014a). Finally, we suggest that a fuller genetic investigation of all of the H30Rx loci identified as clade associated may be of merit. A saturated transposon mutant library has been constructed in an H30Rx strain and was utilised to determine the essential gene set for serum resistance (Phan *et al.* 2013). Using such a library to test a wider set of environmental and infection conditions would undoubtedly elucidate if the H30Rx unique loci do indeed play a formative role in the success of the lineage.

4.5 Conclusion

Given many recent reports showing the variable virulence potential of *E. coli* ST131 strains (Toh *et al.* 2010, Avasthi *et al.* 2011, Croxall *et al.* 2011a, Totsika *et al.* 2011, Gibreel *et al.* 2012b, Lavigne *et al.* 2012), and on the basis of many previous studies that proposed the association between high

metabolic capacity and enhanced bacterial fitness (Le Bouguénec and Schouler 2011), this study tested and compared the metabolic activity of a collection of ST131 and non-ST131 strains using two different testing methodologies: API strips and PM technology. Our API data showed that ST131 strains had a lower metabolic activity for 5 substrates, which is in contrary to a recent finding by Gibreel *et al.* (2012a) that showed ST131 strains with an increased metabolic potential compared to other ExPEC STs. Further testing of the metabolic activity of *E. coli* using phenotypic microarray system demonstrated the absence of specific metabolic profile for ST131 strains suggesting that ST131 is not a metabolically distinct lineage of ExPEC and thus metabolism might not contribute to the fitness of this clone.

We investigated the gene content of a group of *E. coli* including ST131 and non-ST131 strains in order to identify the presence of other loci that are uniquely associated with ST131 H30Rx clade. Our gene content analysis of *E. coli* strains performed using Gegenees (Ågren *et al.* 2012) showed the presence of few genetic loci unique to ST131 H30Rx clade. Further investigation of these loci using LS-BSR (Sahl *et al.* 2014) identified the presence of 150 loci, the majority of which were hypothetical proteins, uniquely associated with this clade. Generally, our data demonstrated that this clade is rather unremarkable in terms of its gene content compared to other ExPEC. It also identified the presence of a small set of metabolic loci that confirms our previous finding demonstrating that ST131 H30Rx is not a metabolically distinct lineage of ExPEC (Alqasim *et al.* 2014a). However the most striking finding at a genomic-level was identification of the secondary flagellar locus Flag as a region uniquely associated with ST131

H30Rx strains (Totsika *et al.* 2011, Clark *et al.* 2012, Petty *et al.* 2014). This region merits further detailed bacterial genetics analysis to uncover its true importance to the emergence and success of the H30Rx clade.

Chapter five

Comparative studies on the serum resistance and capsule genetics of a group of *E. coli* ST131 and non-ST131 strains

5.1 Introduction

ExPEC is a major etiological agent of bloodstream infections (BSIs) (Rogers *et al.* 2011), and it represents 17-37% of total bacteria isolated from patients with BSIs (Russo and Johnson 2003). The capability of ExPEC to cause these infections in humans can be attributed to the presence of many virulence determinants that allow ExPEC to survive in the blood and evade the host innate defences (Miajlovic and Smith 2014).

5.1.1 *E. coli* resistance to human serum

The ability to resist the bactericidal activity of serum, and therefore survive in the bloodstream, represents a crucial virulence determinant for ExPEC strains that cause bacteraemia (McCabe *et al.* 1978, Martinez *et al.* 1986, Jacobson *et al.* 1992). It has been suggested that *E. coli* strains isolated from blood are more serum-resistant than those associated with causing urinary tract infections or isolated from faecal samples (Johnson 1991, Jacobson *et al.* 1992). *E. coli* strains use many mechanisms to evade the innate defences present in serum including complement and antimicrobial peptides (Miajlovic and Smith 2014). For example, the role of cell surface polysaccharides, O antigens and K capsules in conferring resistance to human serum has been identified (Feingold 1969, Gemski *et al.* 1980, Opal *et al.* 1982). Other factors such as the major outer membrane protein OmpA (Weiser and Gotschlich 1991), plasmid-encoded proteins TraT (Moll *et al.* 1980, Montenegro *et al.* 1985) and Iss (Smith 1974) and the phage membrane protein Bor (Barondess and Beckwith 1995), have also been reported to contribute to serum resistance in *E. coli*. Additionally, a recent study on the serum resistance of the ExPEC strain *E. coli* CFT073 has identified the role of three extracytoplasmic stress response pathways Rcs,

Cpx two-component systems and the alternate sigma factor σE in providing protection against the bactericidal effect of serum (Miajlovic *et al.* 2014). It has been reported that the presence of all of these mechanisms in a single bacterial strain is not necessarily required to make it serum-resistant since the presence of one or more factors can lead to serum resistance in *E. coli* (Cross *et al.* 1986, Stawski *et al.* 1990).

5.1.2 The proposed role of *E. coli* cell surface polysaccharides in serum resistance

Bacterial lipopolysaccharide (LPS) comprises the highly conserved lipid A-core and repeating O antigen subunits (Stenutz *et al.* 2006). More than 180 distinct O antigen serotypes have been identified for *E. coli* (Orskov *et al.* 1977, Stenutz *et al.* 2006). ExPEC strains are commonly associated with particular O antigen types such as O1, O2, O4, O6, O7, O8, O16, O18, O25, and O75 (Totsika *et al.* 2012). Previous reports proposed the role of O antigen in providing protection against serum killing either by activating complement away from its target sites on the bacterial outer membrane or by blocking antibody-binding sites (Sansano *et al.* 1985, Johnson 1991). A recent study has shown that the surface expression of some O antigen types, such as O1, O6 and O18, is important for bacterial survival in human serum (Sarkar *et al.* 2014).

Exopolysaccharides (EPSs) have a distinct sugar composition and are loosely associated with the cell surface (Whitfield 2006). The exopolysaccharide colanic acid is regulated by the two-component Rcs phosphorelay, and has been associated with biofilm formation and protection from desiccation (Clarke 2010). A recent study has shown that colanic acid was protective against the bactericidal effect of human serum

in the ExPEC strains CFT073 and RS218 (Miajlovic *et al.* 2014).

E. coli strains that cause bacteraemia are commonly associated with the expression of surface-associated capsule polysaccharide (Miajlovic and Smith 2014). Bacterial capsule acts as a barrier protecting the bacterial outer membrane from various host defences involving deposition of complement factors (Buckles *et al.* 2009). Capsules frequently contain capsular (K) antigen, and like the O antigens, K antigens have a remarkable diversity with more than 80 serologically different K antigens identified in *E. coli* (Whitfield 2006). K antigens belong to four distinct groups that are defined on the basis of genetic assembly, biochemical and physical attributes (Whitfield and Roberts 1999). Although the group I capsules are associated with commensal *E. coli* strains, group II and III capsules are associated with ExPEC strains (Johnson and O'Bryan 2004). The group 4 capsules are frequently identical to that of the cognate lipopolysaccharide O side chain and have, therefore, also been called the O-antigen capsule. These capsules are associated with strains belonging to some IPEC pathotypes such as EPEC (Peleg *et al.* 2005). Comparison of the capsule locus of the ExPEC strain CFT073 with that of other *E. coli* strains is shown in Figure 5.1.

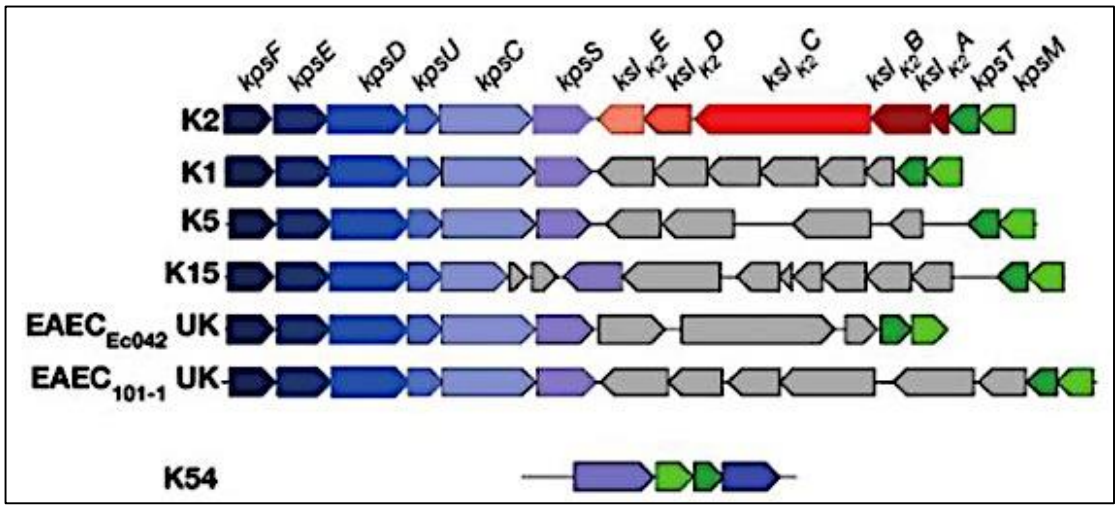


Figure 5.1: Schematic representation of the capsule gene cluster of *E. coli* CFT073 compared with that of other *E. coli* strains. The K1 capsule is represented by the UTI89 isolate; K5 by the Nissle 1917 isolate and K15 by extraintestinal pathogenic *E. coli* strain 536. Enteroaggregative *E. coli* (EAEC) isolates Ec042 and 101–1 K serotype are also included. The ExPEC strain CP9 K54 group 3 capsule cluster is shown for comparison. Blue genes denote region 1; red genes denote region 2 and green genes denote region 3 of the *kps* locus of CFT073 strain [Taken from (Buckles *et al.* 2009)].

Previous studies demonstrated the role of capsule in protecting ExPEC strains from phagocytosis (Howard and Glynn 1971, Weiss *et al.* 1982, Sarkar *et al.* 2014). However its role in complement-mediated killing is unclear. Previous reports investigated the role of many capsule types in protecting ExPEC strains from the bactericidal effect of human serum, and identified the role of some capsule types such as the K2 capsule of the ExPEC strain CFT073 and the K1 capsule associated with other strains causing neonatal meningitis in human serum survival (Leying *et al.* 1990, Buckles *et al.* 2009). However other capsule types such as K27 and K15 have been found to confer no protective role against serum-mediated killing (Opal *et al.* 1982, Schneider *et al.* 2004).

5.1.3 *E. coli* ST131 resistance to human serum

The recent emergence of *E. coli* ST131 as a multi-drug-resistant pathogen with an increased capability of causing bloodstream infections (Phan *et al.*

2013) highlights the importance of studying the mechanisms that underpin the fitness of this clone at a molecular-level. The combination of next generation sequencing and transposon mutagenesis approaches has recently been used to study the essential genes required for many bacterial species to survive in particular niches (Langridge *et al.* 2009, Christen *et al.* 2011). For example, Langridge *et al.* (2009) used the transposon directed insertion-site sequencing (TraDIS) method to identify the essential genes required for the survival of *S. typhi* in the presence of bile salts.

A recent study has used TraDIS to define the essential gene set required for growth and survival of the *E. coli* ST131 strain EC958 in human serum (Phan *et al.* 2013), and identified a set of fifty-six genes required for the survival of this strain in human serum. Of these fifty-six genes, twenty-two belong to three operons responsible for LPS production (including O25-antigen biosynthesis and lipid A-core biosynthesis) and enterobacterial common antigen (ECA) biosynthesis. The TraDIS screen also identified three genes encoding for colanic acid biosynthesis, however K capsular biosynthesis genes were not present in the essential gene set (Phan *et al.* 2013).

5.1.4 Genetic and biochemical diversity of *E. coli* ST131 capsule locus

A recent study has examined the genetic architecture of *E. coli* ST131 H30Rx clade strains and has identified a recombinant fragment, encoding the group II capsule synthesis locus, common across *E. coli* ST131, with several variant region 2 gene clusters were observed between region 1 (*kpsFEDUCS*) and region 3 (*kpsTM*) of ST131 H30Rx genomes (Petty *et al.* 2014). Additionally, although PCR-based methods have identified K2

antigen type in *E. coli* ST131 via using *kpsMII* primers (Johnson and O'Bryan 2004, Croxall *et al.* 2011a), classical capsule typing of a collection of *E. coli* ST131 strains, many of which belong to the H30Rx clade, has shown a high degree of diversity in the biochemical profile of K antigens (Olesen *et al.* 2013). There were a total of 7 different K antigen types identified within the 44 *E. coli* ST131 isolates tested. Indeed, none of the strains biochemically tested were identified as K2 but rather as K100 despite testing K2 positive by PCR (Olesen *et al.* 2013).

5.1.5 Aims of the study

Since this PhD project mainly aimed to determine the factors that give *E. coli* ST131 clone a competitive fitness advantage over other potential ExPEC clones, and given the importance of serum resistance in contributing to the virulence of ExPEC strains that cause bacteraemia, we aimed to test and compare the ability of a collection of ExPEC strains, including *E. coli* ST131 and non-ST131 strains, to resist the bactericidal effect of human serum.

Additionally, given that some previous studies proposed the important role of ExPEC capsular polysaccharides in providing protection against the serum-mediated killing, we aimed to examine the capsule region of a collection of ExPEC strains, including ST131 and non-ST131 strains, at a genomic level in order to identify and compare loci present in the capsule region of these strains. Moreover, on the basis of the confusing data available to date on the diversity of the capsule locus of *E. coli* ST131 H30Rx clade, we sought to investigate the genetic architecture for the diversity observed in the capsule locus of *E. coli* ST131 strains belonging to the H30Rx clade using previously published genomes (Avasthi *et al.* 2011,

Totsika *et al.* 2011, Clark *et al.* 2012, Andersen *et al.* 2013). We also sought to investigate the diversity in the biochemical profile of K antigens of these strains using the classical capsule typing method (Olesen *et al.* 2013).

5.2 Material and Methods

5.2.1 Bacterial strains

Forty-two *E. coli*, twenty *E. coli* ST131 and twenty *E. coli* non-ST131 strains, were used to carry out the serum resistance assay. The highly invasive ST73 strain CFT073 was used as a positive control, while the *E. coli* K12 strain MG1655 was used a negative control. Afterwards, nine whole-genome sequenced *E. coli* ST131 strains were used to perform the capsule production and biofilm formation assays. Of these nine ST131 strains, five strains were selected to perform the classical capsule-typing assay representing each of the genetic capsule type. Information on *E. coli* strains used in this study is given in Table 5.1.

Table 5.1: The characteristics of strains used to perform the experimental work in chapter 5.

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
E. coli CFT073	ST73	UTI	Urine	-	A
UTI18	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI32	ST131	Monomicrobial UTI	Urine	CTX-M-15	A
UTI62	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI188	ST131	Monomicrobial UTI	Urine	-	A
UTI226	ST131	Polymicrobial UTI	Urine	-	A
UTI587	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
B3	ST131	Bacteraemia	Blood	-	B
B5	ST131	Bacteraemia	Blood	CTX-M-15	B
B16	ST131	Bacteraemia	Blood	CTX-M-15	B
B22	ST131	Bacteraemia	Blood	-	B
B37	ST131	Bacteraemia	Blood	-	B
B44	ST131	Bacteraemia	Blood	-	B
B64	ST131	Bacteraemia	Blood	CTX-M-15	B
B65	ST131	Bacteraemia	Blood	CTX-M-15	B
B75	ST131	Bacteraemia	Blood	CTX-M-15	B
B89	ST131	Bacteraemia	Blood	CTX-M-15	B
U2	ST131	UTI	Urine	CTX-M-15	B
U5	ST131	UTI	Urine	-	B
U12	ST131	UTI	Urine	CTX-M-15	B
U79	ST131	UTI	Urine	CTX-M-15	B
B12	ST73	Bacteraemia	Blood	-	B
B14	ST73	Bacteraemia	Blood	-	B
U1	ST73	UTI	Urine	-	B
U7	ST73	UTI	Urine	-	B

Strain ID	ST	Strain history	Sample source	CTX-M carriage	Reference
B31	ST69	Bacteraemia	Blood	-	B
B33	ST69	Bacteraemia	Blood	-	B
U64	ST69	UTI	Urine	-	B
U67	ST69	UTI	Urine	-	B
UTI89	ST95	Uncomplicated cystitis	Urine	-	C
B34	ST95	Bacteraemia	Blood	-	B
B38	ST95	Bacteraemia	Blood	-	B
U22	ST95	UTI	Urine	-	B
U60	ST95	UTI	Urine	-	B
B9	ST10	Bacteraemia	Blood	-	B
B20	ST10	Bacteraemia	Blood	-	B
U19	ST10	UTI	Urine	-	B
U95	ST10	UTI	Urine	-	B
U18	ST3451	UTI	Urine	-	B
U58	ST91	UTI	Urine	-	B
U104	ST3452	UTI	Urine	-	B
E. coli MG1655	ST10	na[1]	na	-	A
UTI24	ST131	Monomicrobial UTI	Urine	-	A
UTI306	ST131	Polymicrobial UTI	Urine	CTX-M-15	A
UTI423	ST131	Polymicrobial UTI	Urine	-	A
B31	ST69	Bacteraemia	Blood	-	B
B33	ST69	Bacteraemia	Blood	-	B
U64	ST69	UTI	Urine	-	B
U67	ST69	UTI	Urine	-	B

UTI89	ST95	Uncomplicated cystitis	Urine	-	C

[1]na: not available; A: (Croxall *et al.* 2011b); B: (Alhashash *et al.* 2013); C: (Chen *et al.* 2006b).

5.2.2 Bacterial genome data

A total of twenty-one publically available ExPEC genome sequences were used in this analysis. Of these genome sequences, nine were reference non-ST131 genome sequences available from National Centre of Biotechnology Information (NCBI) website, while twelve were ST131 genomes produced during previous studies in our group (Clark *et al.* 2012, Alqasim *et al.* 2014b) and from other sources (Totsika *et al.* 2011, Andersen *et al.* 2013). Information on genomes used in this study is shown in Table 5.2.

Table 5.2: List of *E. coli* genomes involved in capsule genetics work.

Strain	ST	Pathotype	Accession number
E. coli UTI18	131[a]	ExPEC	ERP001095
E. coli EC958	131[b]	ExPEC	CAFL01000001
E. coli UTI24	131[a]	ExPEC	ERP001095
E. coli UTI32	131[a]	ExPEC	ERP001095
E. coli UTI62	131[a]	ExPEC	ERP001095
E. coli UTI188	131[a]	ExPEC	ERP001095
E. coli UTI226	131[a]	ExPEC	ERP001095
E. coli UTI306	131[a]	ExPEC	ERP001095
E. coli UTI423	131[a]	ExPEC	ERP001095
E. coli UTI587	131[a]	ExPEC	ERP001095
E. coli JJ1886	131	ExPEC	CP006784.1
E. coli JIE168	131	ExPEC	ERP001095
E. coli IHE3034	95	ExPEC	CP001969.1
E. coli UTI89	95	ExPEC	CP000243.1
E. coli S88	95	O45 ExPEC	CU928161.2
E. coli SMS35	8	Multidrug-resistant ExPEC	CP000970.1
E. coli ED1a	452	O81	CU928162.2
E. coli CFT073	73	ExPEC	AE014075.1
E. coli Di14	73	ExPEC	AE014075.1
E. coli Di12	73	ExPEC	CP002211.1
E. coli CE10	62	K1 ExPEC	NC_017646

[a]2009 UK ST131 strains; [b]2004 UK ST131 strain.

5.2.3 Identification of capsule loci in *E. coli* ST131 genomes

The nine ST131 genomes produced in our group (Clark *et al.* 2012) were re-assembled using the De Bruijn graph-based de novo assembler Velvet (Zerbino and Birney 2008) and Post Assembly Genome Improvement Toolkit (PAGIT) (Swain *et al.* 2012). The genome re-assembly method was described previously in this thesis (Materials and methods section, chapter 4). Afterwards, the assembled genomes were then annotated using Prokka (Seemann 2014) and the capsule regions written to new embl files using Artemis. The capsule encoding regions were visually compared by Dr. Alan McNally using Easyfig (Sullivan *et al.* 2011) and variable genes were searched against the non-redundant database by BlastX search. All software scripts and command lines used to perform this part of study are given in the appendix (Chapter 7).

5.2.4 Whole genome phylogeny

This part of study was performed by Dr. Alan McNally. All ST131 genome sequences were aligned using Mugsy (Sahl *et al.* 2011) and the core genome extracted using Mothur (Schloss *et al.* 2009) with the default parameters of the methods. Maximum likelihood phylogeny was determined using RaxML (Stamatakis *et al.* 2005) using the rapid bootstrap function and the general time reversible (GTR) model with Gamma correction, with 100 bootstraps performed. The resulting phylogeny was visualised using Figtree. All command lines used to perform this part of study are given in the appendix (Chapter 7).

5.2.5 Phenotypic assays

5.2.5.1 Classical capsule typing

Serotyping was done according to the method of Ørskov and Ørskov (Orskov and Orskov 1984) at Statens Serum Institute, Denmark using five *E. coli* ST131 strains. The K antigen was determined by countercurrent immunoelectrophoresis involving K-specific antisera, except for the K1 and K5 antigens, which were detected using K1- and K5-specific bacteriophages (Olesen *et al.* 2013).

5.2.5.2 Comparative serum resistance assay

Bacterial strains were plated out on LB agar plates and incubated overnight at 37°C. A single colony of each strain was inoculated into 5ml of LB broth followed by overnight incubation for 18h at 37°C with shaking. Cells were harvested by centrifugation at 16000 rcf and resuspended in 1ml of PBS to ~ 1 x 10^6 CFU/ml. 0.5ml of diluted bacterial cells was added to 1.5ml of undiluted human serum (Sigma Aldrich, UK). Viable cell count (CFU/ml) values were obtained immediately after adding the serum (T0) and after 3 hours of incubation (T3). To assess the levels of serum resistance, the log CFU/ml value obtained after T3 was compared with that obtained at T0. The serum resistance values were representative values of log CFU/ml at T0 and T3 obtained from performing this assay in duplicate on three independent occasions.

5.2.5.3 Capsule production assay

Bacterial strains were grown on LB and CLED media and incubated overnight at 37°C and 25°C. Capsule production was determined using a scoring system through testing the ability of each strain to form mucoid colonies in LB and in CLED agar plates using two incubation temperatures, 37°C and 25°C. Each strain was tested for its ability to produce capsule in

triplicate on three independent occasions.

5.2.5.4 Biofilm formation assay

Bacterial strains were grown on LB medium and incubated overnight at 37°C. Then, a single colony of each strain was inoculated into 2ml of LB and BHI broth and incubated statically at 37°C for 24h. Bacterial cultures were then diluted 1/100 and aliquots of 130µl were placed in a 96-well plate, with 5 wells per strain followed by overnight static incubation at 37°C. The bacterial growth was estimated by obtaining the OD values at 600nm. Wells were washed once using 200µl sterile PBS and stained with 130µl of 1% crystal violet solubilised in ethanol for 5 minutes at room temperature without shaking. Wells were washed five times using 200µl sterile PBS and 130µl of absolute ethanol was added to each well followed by incubation for 5 minutes at room temperature without shaking. Biofilm formation values were determined using scoring system and they were representative values of measured levels of crystal violet retention as measured at A_{600} (Chassaing and Darfeuille-Michaud 2013). Assays were performed in triplicate on three independent occasions.

5.2.6 Statistical analysis for serum resistance data

The significance of difference in the levels of serum resistance between *E. coli* ST131 and non-ST131 strains was determined using the two-way analysis of variance (ANOVA) test available in Prism software (version 6.0e), and the threshold for statistical significance was a *P* value of ≤0.05. Firstly, the significance of difference in the levels of serum resistance values at different time points was determined for each tested strain. Then, group analysis was carried out using two-way ANOVA test in order to determine

the significance of difference in the levels of serum resistance at different time points between ST131 and non-ST131 strains, and between ST131 UTI and bacteraemia strains.

5.3 Results

5.3.1 Serum resistance profiles of *E. coli* strains

Forty-two *E. coli* strains, twenty *E. coli* ST131 and twenty *E. coli* of different ST type isolated from urine and blood samples, were tested for their ability to resist the human serum. The highly invasive ST73 strain CFT073 was used as a positive control, while the *E. coli* K12 strain MG1655 was used a negative control. The data showed that the CFT073 strain was serum-resistant and this was exhibited by the increase in the log CFU/ml value after 3h of incubating the bacterial suspension with human serum compared to that at T0, while the MG1655 strain failed to survive in human serum, as it did not show any sign of growth on agar plates at T3 (Figure 5.2). Additionally, all the twenty *E. coli* ST131 strains (100%) were associated with high serum resistance phenotype, shown by the almost 2-log increase in the log/CFU values at T3 compared to that at T0 (Figure 5.1). The data also showed that the levels of serum resistance in the *E. coli* ST131 strains isolated from blood were almost similar to those isolated from urine. With respect to the *E. coli* non-ST131 strains, all four *E. coli* ST73 strains (100%) were serum-resistant, while there was variation in the levels of serum resistance in the ExPEC strains belonging to other STs such as ST69, ST95 and ST10. Additionally, three non-ST131 strains U18, U58 and U104 that belong to ST3451, ST91, and ST3452 respectively were serum-sensitive (Figure 5.2).

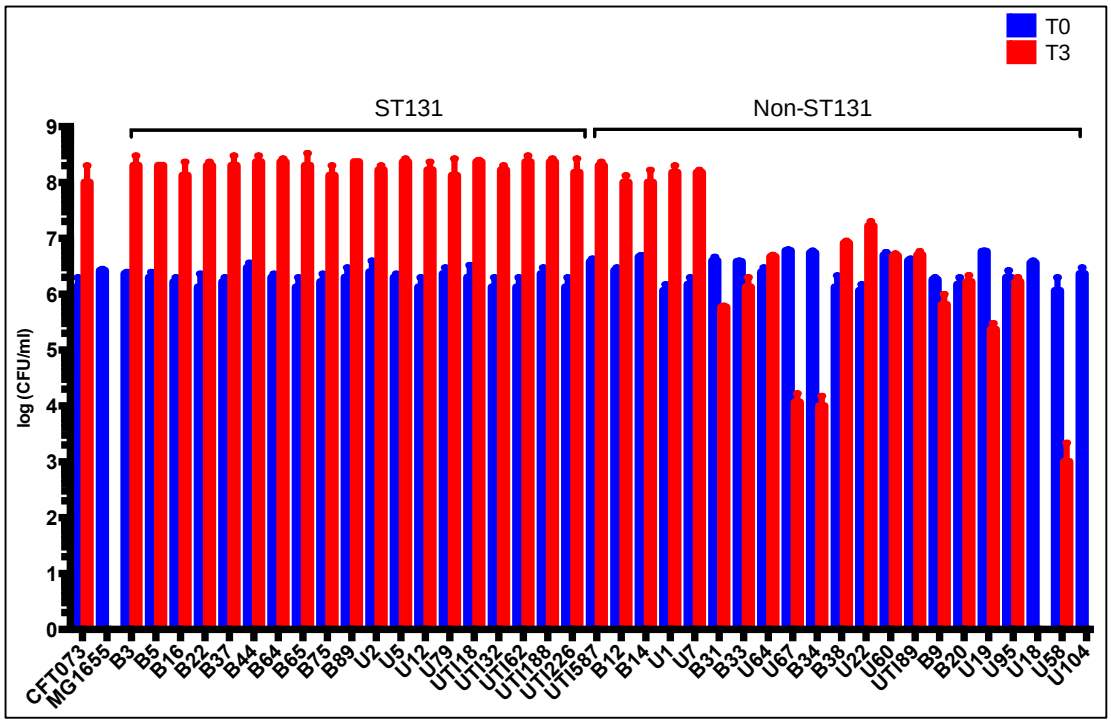


Figure 5.2: Serum resistance levels of *E. coli* strains given in log (CFU/ml). These values obtained from calculating the average of 3 log (CFU/ml) values for each *E. coli* strain. Blue bars indicate the value of serum resistance at the beginning of the assay (T0) and red bars indicate the value of serum resistance after 3h of incubation (T3).

In summary, the data showed that all tested *E. coli* ST131 and ST73 strains were highly serum-resistant, and the levels of serum resistance in strains belonging to these two STs were higher than that of other *E. coli* strains of different STs (Table 5.3). It also showed that there was variability in the levels of serum resistance of strains belonging to other major ExPEC STs such as ST69, ST95 and ST10.

Table 5.3: A summary of serum resistance profiles of *E. coli* strains belonging to different STs.

***E. coli* ST**	**Number of tested strains**	**Number of serum-resistant strains**	**Percentage**
E. coli ST131	20	20	100%
E. coli ST73	4	4	100%
E. coli ST69	4	1	25%
E. coli ST95	5	3	60%
E. coli ST10	4	1	25%
E. coli ST3451	1	0	0%
E. coli ST91	1	0	0%
E. coli ST3452	1	0	0%

5.3.2 Comparative statistical analysis of the serum resistance profiles of *E. coli* strains

The significance of difference between the levels of serum resistance at T0 and T3 of each *E. coli* strain tested in this study was determined using the two-way ANOVA test. Firstly, the levels of serum resistance at T0 and T3 of each *E. coli* strain were compared to that of the negative control *E. coli* MG1655 strain. The data showed that there was a statistically significant difference between the levels of serum resistance at T0 and T3 for the positive control CFT073 strain and for all the twenty ST131 strains (Figure 5.3A). With respect to the non-ST131 strains, there was also a significant difference between the levels of serum resistance at T0 and T3 for the four *E. coli* strains belonging to ST73 (Figure 5.3B). However the difference between the levels of serum resistance at different time points for the strains belonging to other ExPEC STs was insignificant (Figure 5.3B). The graph shows that this insignificant difference is surprising given the detectable difference in the levels of serum resistance at different time points for some non-ST131 serum-sensitive strains such as U67 and U58. This might be explained by the way used by the statistics software to calculate the significance of difference in the levels of serum resistance for strains at different time points.

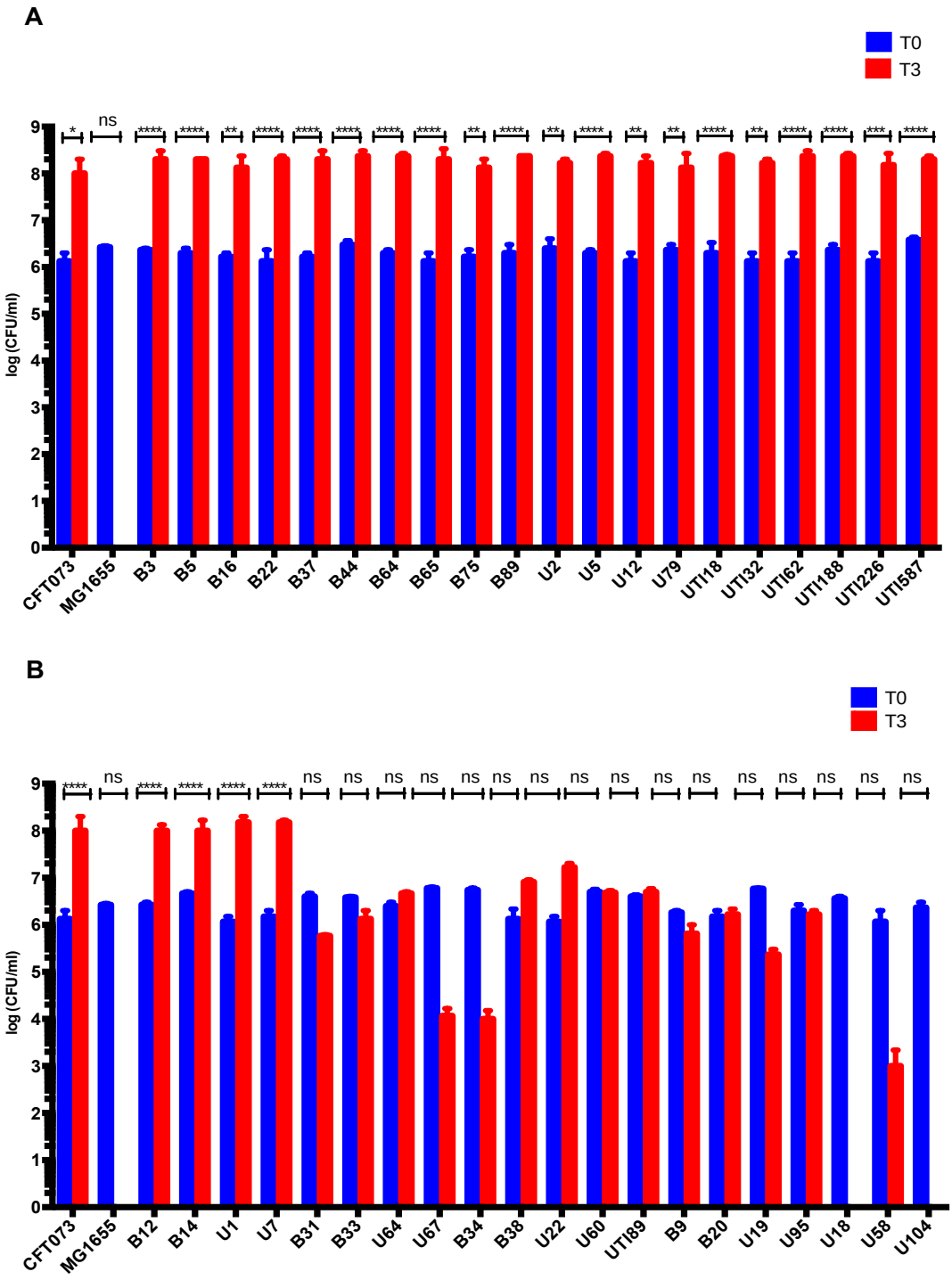


Figure 5.3 (A-B): Statistical analysis of the levels of serum resistance at T0 and T3 for each *E. coli* strain compared to that of the negative control strain *E. coli* MG1655. (A) The graph shows the significance of difference between the levels of serum resistance at T0 and T3 for the positive control strain *E. coli* CFT073 and the 20 *E. coli* ST131 strains as tested by the two-way ANOVA test. (B) The graph shows the significance of difference between the levels of serum resistance at T0 and T3 for the positive control strain *E. coli* CFT073 and the 20 non-ST131 strains as tested by the two-way ANOVA test. The presence of significant difference between the levels of serum resistance at T0 and T3 is shown by the asterisks [* (P<0.05), ** (P<0.01), *** (P<0.001) and **** (P<0.0001)]. ns means not significant.

Afterwards, the significance of difference between the levels of serum resistance of *E. coli* ST131 and non-ST131 strains was determined by a group analysis using two-way ANOVA test. Our data showed that the difference in the levels of serum resistance at T3 between ST131 and non-ST131 strains was statically significant (*P* value < 0.0001), whereas there was no significant difference in these levels between ST131 and non-ST131 strains at T0 (Figure 5.4A). Further group analysis showed that there was no statistical difference between ST131 strains isolated from blood and ST131 strains isolated from urine in terms of their serum resistance levels at T0 and T3 (Figure 5.4B).

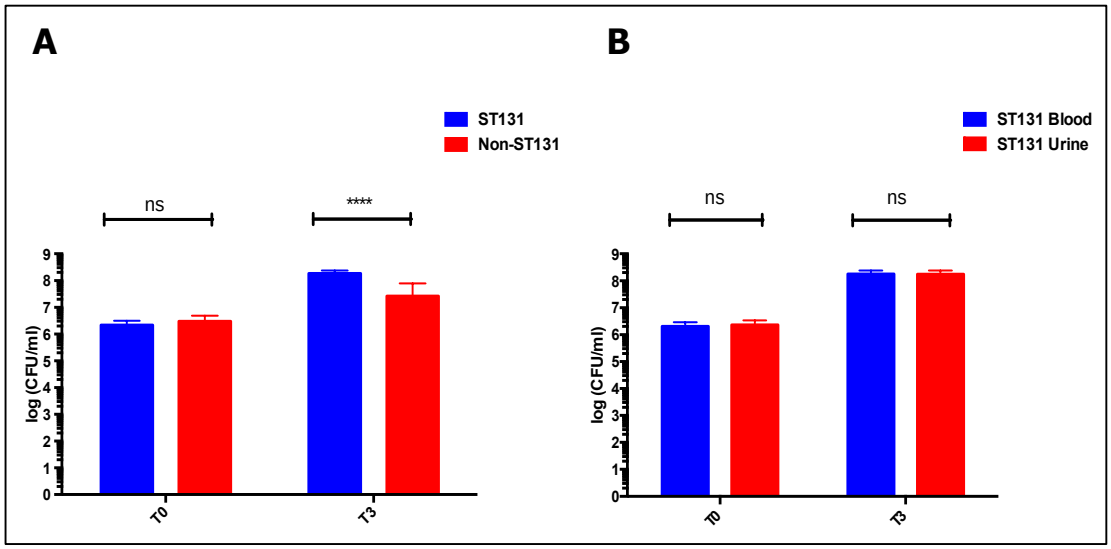


Figure 5.4 (A-B): Serum resistance group analysis performed using the two-way ANOVA test. (A) Group analysis showing the significance of difference between the levels of serum resistance at T0 and T3 for ST131 and non-ST131 strains as tested by the two-way ANOVA test. (B) Group analysis showing the significance of difference between the levels of serum resistance at T0 and T3 for ST131 blood isolates and ST131 urine isolates as tested by the two-way ANOVA test. The presence of significant difference between the levels of serum resistance at T0 and T3 is shown by the asterisks [* (*P*<0.05), ** (*P*<0.01), *** (*P*<0.001) and **** (*P*<0.0001)]. ns means not significant.

In short, our data demonstrated the presence of a statistical significant difference in the levels of serum resistance between ST131 and non-ST131 strains at T3 but not in T0, while there was no significant difference between

ST131 strains isolated from blood and ST131 strains isolated from urine in terms of the serum resistance levels at different time points.

5.3.3 Identification and comparison of genetic loci present in the capsule region of a collection of ExPEC strains

Given the proposed role of group II capsule polysaccharides of some ExPEC strains in conferring resistance the human serum, we examined the capsule region of twenty ExPEC strains, including ST131 and non-ST131 (Table 5.4), at a genomic level to identify and compare genetic loci present in the capsule region of these strains. Using Artemis, the capsule region of each strain was extracted from each genome and a separate embl file for the capsule locus of each strain was created. For all ExPEC strains, twenty-seven capsule loci, either with known or putative functions, were identified in the capsule region. Additionally, there were a number of other loci encoding for hypothetical proteins of unknown functions. Table 5.4 shows the twenty-seven capsule loci identified for ExPEC strains with their known or putative functions.

Table 5.4: The capsule loci identified for the twenty ExPEC strains and their known or putative functions.

Locus tag	Function
KpsF	Arabinose 5-phosphate isomerase
KpsE	Vi polysaccharide export inner membrane protein
KpsD	Polysialic acid transport protein precursor
KpsU	3-deoxy-manno-octulosonate cytidylyltransferase
KpsC	Capsule polysaccharide biosynthesis protein
KpsS	Capsule polysaccharide biosynthesis protein
KpsT	Polysialic acid transport ATP-binding protein
KpsM	Polysialic acid transport protein
HAD superfamily	Putative hydrolase
LPSBP	LPS biosynthesis protein
TagD_2	Glycerol-3-phosphate cytidylyltransferase
VatD	Streptogramin A acetyltransferase
CBP	Capsule biosynthesis phosphatase
DG1P	D-glucose-1-phosphatase
POLNAT	Putative O-linked N-acetylglucosamine Transferase
Ugd_1	UDP-glucose 6-dehydrogenase
SpsA	Pore coat polysaccharide biosynthesis protein
IspD_2	Putative 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase 2
tagB	Putative CDP-glycerol:glycerophosphate glycerophosphotransferase
galE_2	UDP-glucose 4-epimerase
galE_3	UDP-glucose 4-epimerase
neu family	-
neuS	Poly-alpha-2, 8 sialosyl sialyltransferase
neuE	Polysialic acid biosynthesis protein
neuC	Polysialic acid biosynthesis protein P7
neuA	Acylneuraminate cytidylyltransferase
nueB	Sialic acid synthase
neuD	Sialic acid synthase

Subsequently, the distribution of these twenty-seven capsule loci was compared between all strains and our data showed that eight loci were found to be present in all ExPEC strains (Figure 5.5). These were the highly conserved *kps* loci, including the region 1 (*kpsFEDUCS*) and the region 3 (*kpsTM*) loci, which encode for the capsule polysaccharides biosynthesis and export proteins. However the other nineteen region 2 loci were variably distributed among ExPEC strains (Figure 5.5). Although some of these loci, such as the *neu* gene family, were only found to be present in the non-

ST131 strains, the vast majority of them were found to be only present in the ST131 strains belonging to H30Rx clade.

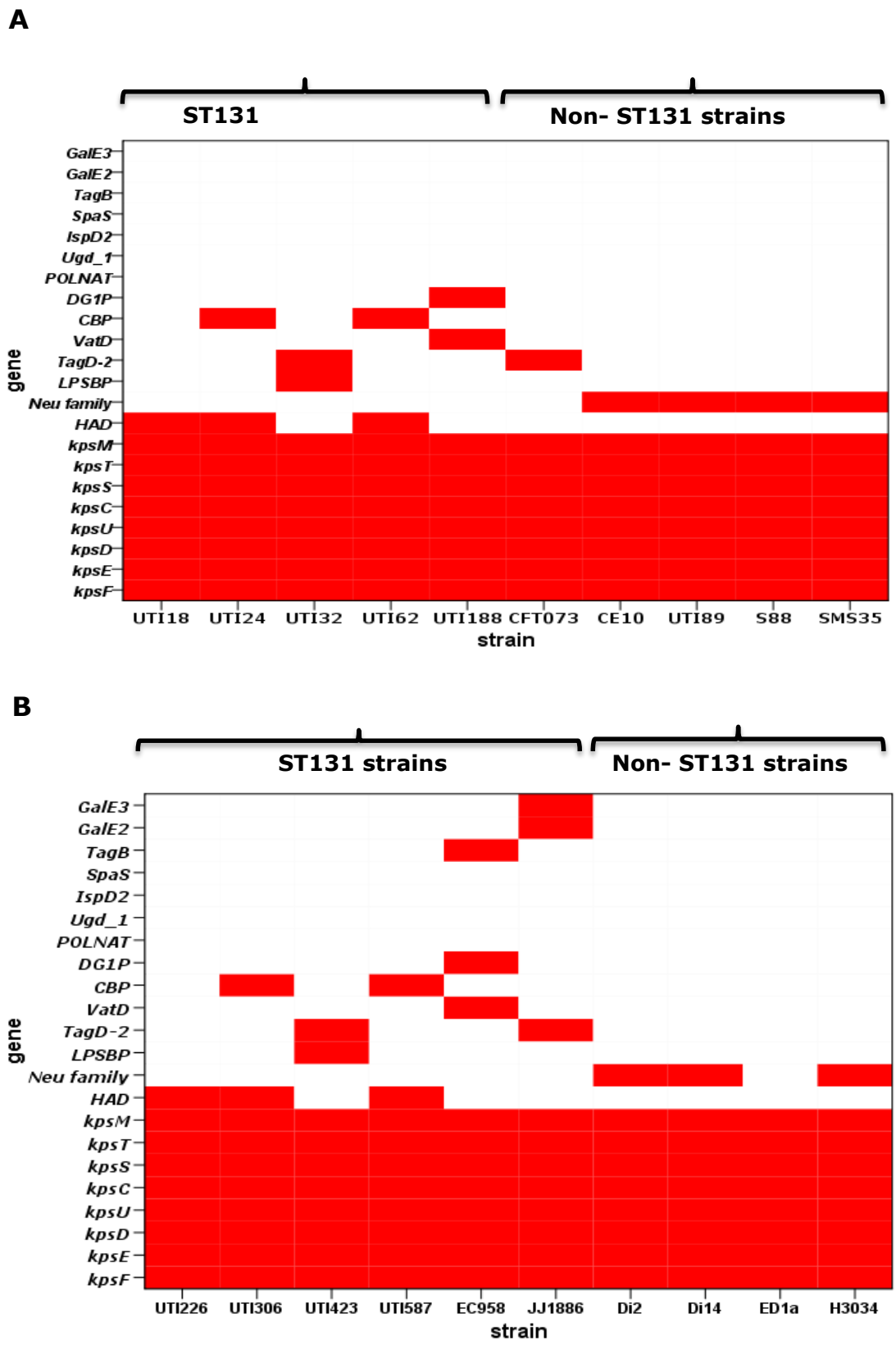


Figure 5.5 (A-B): Heat maps showing the distribution of loci identified in the capsule region of twenty ExPEC strains including ST131 and non-ST131 strains. Red blocks indicate the presence of the loci in the capsule region of each strain, while the white blocks indicate the absence of loci in the capsule region of each strain.

5.3.4 Genetic and biochemical diversity of capsule locus in the *E. coli* ST131 H30Rx clade

On the basis of our result above showing the presence of several region 2 loci that were only associated with ST131 H30Rx strains, and given the reported variability of capsule loci (Petty *et al.* 2014) and capsular antigen type (Olesen *et al.* 2013) in *E. coli* ST131 H30Rx clade strains, these loci were investigated in more detail. The recently released JJ1886 genome (Andersen *et al.* 2013) was selected as a reference genome given it is the only *E. coli* ST131 genome sequenced and assembled to a standard of quality commensurate with being a high quality genome (Chain *et al.* 2009). Using this reference, the contigs of the ST131 genomes previously reported by our group (Clark *et al.* 2012) were re-ordered to ensure the genome architecture was as accurate as possible. The capsule loci of all of the ST131 genomes were identified and separate embl files for each capsule locus of each strain were created which were then compared using Easyfig (Sullivan *et al.* 2011). The comparison of the capsule loci (Figure 5.6) showed a high degree of diversity between the conserved *kpsFEDUCS* and *kpsTM* regions, with no observable similarity between strains in the variable central genes. To ascertain how this genetic architecture reflected upon biochemical typing, the K antigen type of each of the genetic capsule types for strains was determined by classical capsule typing (Figure 5.6), and also overlaid any available capsule type information on the other sequenced strains. The data revealed the presence a high degree of diversity in the biochemical profiles of K antigens. This diversity was reflected by the presence of several K antigen variants in the tested ST131 strains belonging to H30Rx clade, despite the previous use of PCR-based methods has identified K2 antigen type in *E. coli* ST131 via using *kpsMII* primers (Johnson and O'Bryan 2004,

Croxall *et al.* 2011a).

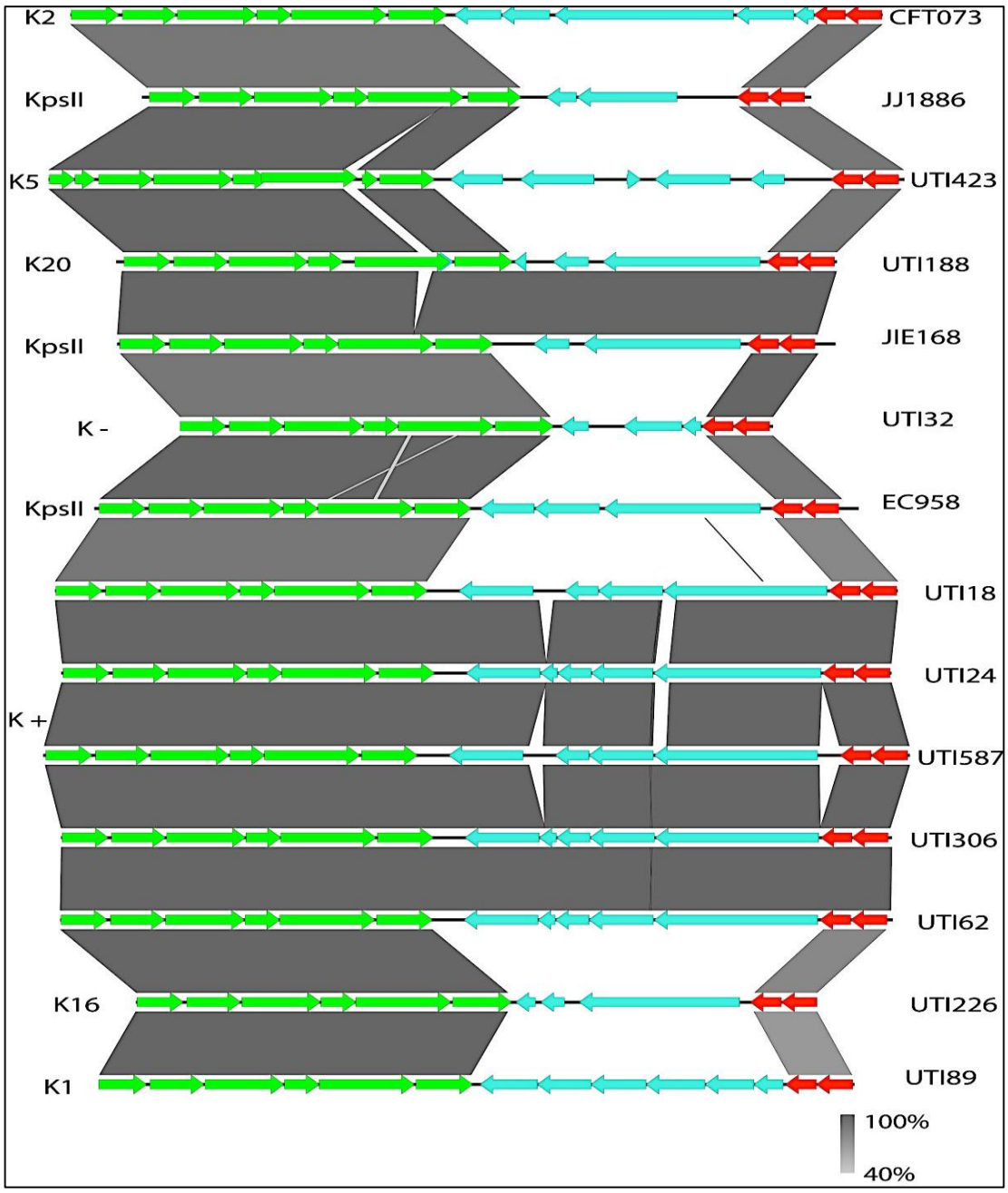


Figure 5.6: Capsule locus genetics of a collection of *E. coli* ST131 H30Rx genomes. The figure shows all of the capsule loci identified in available *E. coli* ST131 genomes. The CDS marked in green are the *kpsFEDUCS* cluster, and those in red the *kpsTM* cluster, both of which are highly conserved across all *E. coli* capsules. The CDS marked in cyan are unique CDS to that capsule type. The grey shading indicates the level of identity between any given pair of CDS. The K antigen identified for each genetic capsule type is indicated in letters to the left, whilst the strain that the capsule locus belongs to is indicated in letters to the right.

Generally, blast analysis of each of the variable central genes in each genetic capsule type returned no significant hits with any reference *E. coli* sequences (Table 5.5). The only exception here was the presence of the *TagD-2* gene that was found in the ST131 UTI32 strain and also in the non-ST131 reference strain CFT073. More importantly for this study, our data showing the presence of significant genetic diversity within the capsule locus in *E. coli* ST131 H30Rx strains suggests frequent and targeted recombination in this region that was previously suggested (Petty *et al.* 2014).

Table 5.5: Blast search results for the variable central genes identified in the capsule locus of *E. coli* ST131 strains.

Gene	ST131 strain	Present in	ID %
TagD-2	UTI32	*E. coli* CFT073	100%
		E. coli KTE220	100%
LPSBP	UTI32	*E. coli* KTE87	99%
HAD	UTI18	*E. coli* KTE76	100%
superfamily	UTI62	*E. coli* PCN033	99%
	UTI226	*E. coli* IS29	99%
	UTI306	*E. coli* HVH190	99%
	UTI587		
CBP	UTI24	*E. coli* KTE76	100%
	UTI62		
	UTI306		
DG1P	UTI188	*E. coli* LAU-EC8	100%
VatD	UTI188	*E. coli* LAU-EC8	100%
	JJ1886	*E. coli* HVH 186	100%
POLNAT	UTI226	*E. coli* HVH177	100%
		E. coli UMEA3656-1	100%
		E. coli UMEA3200-1	100%
Ugd_1	UTI423	*E. coli* KTE21	100%
		E. coli sp_1_1_43	99%
SpsA	UTI423	KFiC *E. coli* str. clone Di14	100%
		E. coli HVH7	
			99%
IspD2	EC958	*E. coli* KTE173	100%
		E. coli KTE175	100%
TagB	EC958	*E. coli* BIDMC 38	100%
		E. coli 907701	100%
galE_2	JJ1886	*E. coli* HVH 186	100%
		E. coli 908632	100%
galE_3	JJ1886	*E. coli* HVH 186	100%
		E. coli HVH 152	100%

5.3.5 Whole genome phylogeny of *E. coli* ST131 H30Rx strains

To examine the diversity of capsule locus in more detail, a core genome phylogeny for *E. coli* ST131 was created using all the ST131 genome sequences in Table 5.2 using SE15 as the root of the phylogeny given its phylogenetic position relative to ST131 H30Rx clade (Petty *et al.* 2014). The resulting phylogenetic tree (Figure 5.7) confirms that all of the strains in the analysis, including those previously sequenced by our group prior to the discovery of H30Rx (Clark *et al.* 2012), did indeed belong the H30Rx clade. More importantly when the capsule loci genetics were superimposed on the phylogenetic tree it clearly demonstrated that the capsule loci were randomly distributed across the phylogeny. The only exceptions to this were the small cluster of strains containing UTI18 that have been previously shown to be essentially a single clone (Clark *et al.* 2012).

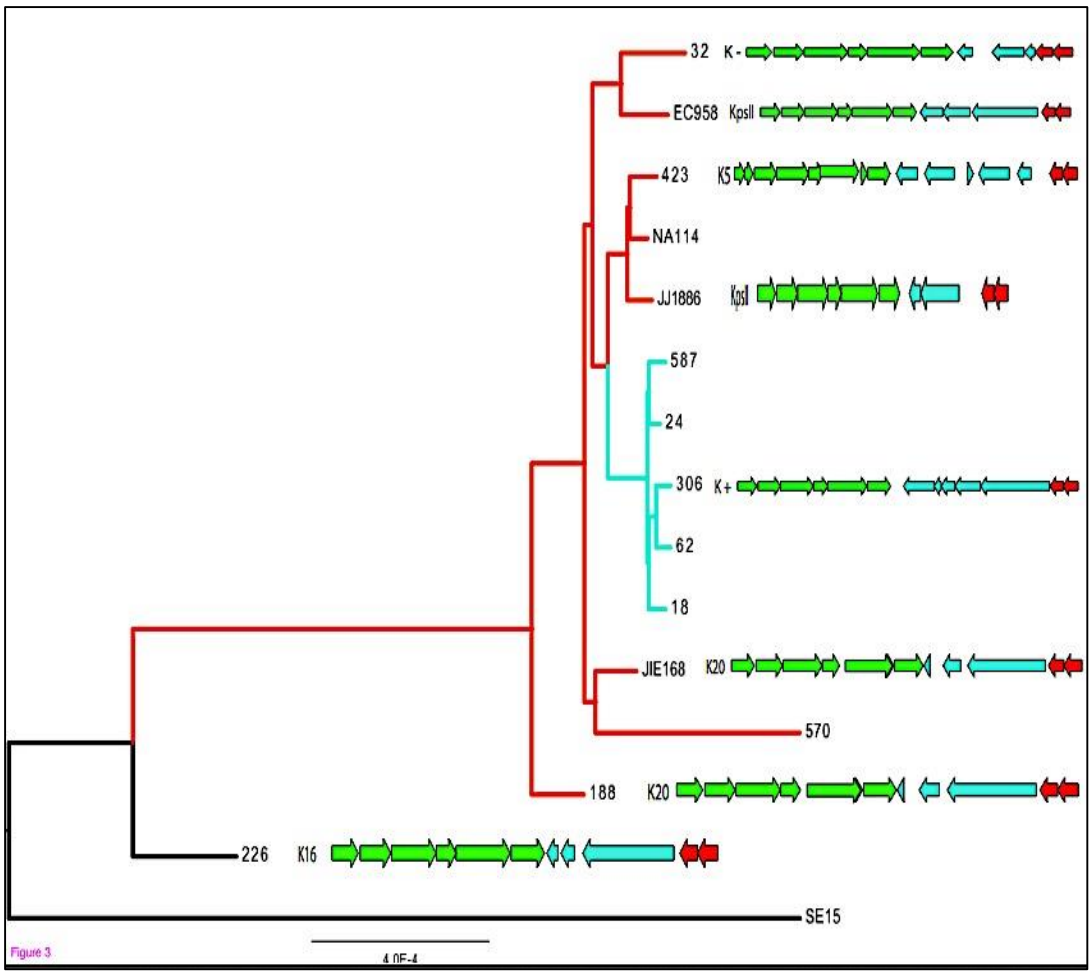


Figure 5.7: Phylogenetic distribution of K-antigen types in *E. coli* ST131 H30Rx core-genome phylogeny of the *E. coli* ST131 genomes analysed in this study, with the SE15 strain included as an outlier. The H30Rx strains are indicated by red colouration of tree branches. The Nottingham "outbreak" strains previously sequenced by our group are indicated by cyan colouration of the tree branches. The K-antigen type and accompanying capsule locus genetics are superimposed to the right of the tree.

5.3.6 The effect of *E. coli* ST131 capsule locus diversity on other virulence-associated phenotypes

On the basis of the extensive recombination at the capsule region of the *E. coli* ST131 H30Rx clade, this study sought to determine any obvious effect of this recombination on the ability of ST131 strains to produce capsule and form biofilms. The ability of ST131 strains to form capsules at 25°C and 37°C on LB and CLED agar plates over a 14-day period was compared. The data showed that all strains had almost similar capsulation morphology on different agar plates at different incubation temperatures, suggesting that

there was no association between the capsule loci present in the different H30Rx strains and levels of capsulation morphology (Table 5.6). Additionally, classical 96 well-plate biofilm formation assays also failed to reveal any significant pattern between different H30Rx capsular variants (Table 5.6).

Table 5.6: Results of capsule-associated phenotype tests for a collection of *E. coli* ST131 strains.

Strain ID	Capsule formation		Biofilm formation	Serotype
	37°C	25°C	37°C	
UTI18	+++	+++	+	O 25: K+: H-
UTI24	+++	+++	++	NT[1]
UTI32	++++	+++	+	O 25: K-: H 4
UTI62	++	++	+	NT
UTI188	++	++	+	O 25:K 20,K 23:H 4
UTI226	+++	++	+++++	O 25: K 16: H-
UTI306	+++	++	+	NT
UTI423	+++	+++	++	O 25: K 5: H 4
UTI587	++	++	++	NT

[1] NT: not tested.

5.4 Discussion

E. coli ST131 is now the dominant causative agent of extra-intestinal infection by *E. coli* in the developed world, and is also heavily responsible for the increase in prevalence in multi-drug resistance in *E. coli* due to extended carriage of the CTX-M-15 ESBL gene (Rogers *et al.* 2011). Although many studies have ascribed the rapid emergence and successful spread of *E. coli* ST131 to the increased antimicrobial resistance of this clone, this phenotype alone is unlikely to provide an explanation for its predominance among other potential MDR uropathogens (Phan *et al.* 2013). Therefore, determining the factors that can drive the success of this clone and give it a fitness advantage over other major ExPEC clones is of great importance.

It has been shown that serum resistance is a major virulence trait for ExPEC strains that are associated with causing bacteraemia, as it allows them to survive in the blood and overcome the host innate defences (Miajlovic and Smith 2014). The ability of *E. coli* strains to resist the bactericidal effect of human serum was tested in this study, and the data showed that all tested *E. coli* ST131 strains were associated with high serum resistance phenotype. Similarly, all ST73 strains were highly serum-resistant, while there was a variation in the levels of serum resistance in other strains belonging to other ExPEC STs ST69, ST95 and ST10. The high serum resistance phenotype observed in ST131 strains concurs with the previous findings showing a high serum resistance phenotype in two matched *E. coli* ST131 strains obtained from urine and blood of a patient identified with urosepsis (McNally *et al.* 2013a), and in the ST131 strain EC958 (Phan *et al.* 2013). According to a previous report by our research group demonstrating the increased association of ST131 and ST73 strains with causing bacteraemia in comparison to other ExPEC STs such as ST69 and ST95 (Alhashash *et al.* 2013), the data of this study suggests that serum resistance might be an important factor in driving the current success of *E. coli* ST131 clone as a major etiological agent of bloodstream infections.

With regard to the difference in the levels of serum resistance between *E. coli* ST131 and non-ST131 strains, our data found a significant difference in the levels of serum resistance at T3 between ST131 and non-ST131 strains, although all the ST73 strains were highly serum resistant. Despite previous studies identifying a high serum resistance phenotype in some ST131 strains (McNally *et al.* 2013a, Phan *et al.* 2013), our comparative study was the first to report the significant difference in the serum

resistance between ST131 and non-ST131 strains. Additionally, although all tested *E. coli* ST131 strains were highly serum-resistant, the data showed no significant difference in the levels of serum resistance between the *E. coli* ST131 strains isolated from blood and those isolated from urine at either T0 or T3. This observation corroborates with the recent finding showing that the levels of serum resistance of the *E. coli* ST131 bacteraemia isolate were not significantly higher than that of the *E. coli* ST131 UTI isolate (McNally *et al.* 2013a). Our data seems to be in disagreement with some previous reports suggesting that *E. coli* bacteraemia isolates are more serum-resistant than *E. coli* UTI isolates (Johnson 1991, Jacobson *et al.* 1992).

Looking at the genetic determinants of serum resistance in *E. coli* ST131, two previous studies by our research group tested the VAG carriage of a collection of UTI and bacteraemia *E. coli* isolates using the multiplex PCR screening assay and found that 83% of ST131 bacteraemia isolates and 80% of ST131 UTI isolates were associated with carrying the plasmid-encoded protein (*traT*) (Croxall *et al.* 2011a, Alhashash *et al.* 2013). The *traT* gene has been identified as an important factor that can contribute to serum resistance in *E. coli* (Moll *et al.* 1980, Montenegro *et al.* 1985), and the high *traT* carriage among the *E. coli* ST131 strains may be one factor accounting for the high serum resistance phenotype observed in all *E. coli* ST131 strains in this study.

A previous study defined the essential genes required for *in vitro* growth and serum resistance of the *E. coli* ST131 strain EC958 using TraDIS (Phan *et al.* 2013), and identified fifty-six genes essential for the serum resistance

of the EC958 strain. The majority of these genes encode membrane proteins or factors involved in the LPS biosynthesis (including O25-antigen biosynthesis and lipid A-core biosynthesis) and enterobacterial common antigen (ECA) biosynthesis. This corroborates with the previous report showing the major role of LPS and its components in conferring the *E. coli* resistance to human serum (Sarkar *et al.* 2014).

The TraDIS screen also identified three genes encoding for colanic acid biosynthesis confirming the previously described role of colanic acid in the serum resistance of *E. coli* (Miajlovic *et al.* 2014). However the K capsular biosynthesis genes were not present in the essential gene set responsible for the serum resistance of EC958 strain suggesting that K antigen does not contribute to the serum resistance of this strain (Phan *et al.* 2013). This result concurs with a previous study on the serum resistance of O75:K5 UPEC strain GR-12, which found that alterations in O75 LPS had more effect on serum resistance than K5 null mutation (Burns and Hull 1998). Additionally, it adds to the previous conflicting findings that showed a variation in the role of capsules in the serum resistance of ExPEC (Opal *et al.* 1982, Leying *et al.* 1990, Stawski *et al.* 1990, Schneider *et al.* 2004, Buckles *et al.* 2009, Sarkar *et al.* 2014).

Given the proposed role of group II capsule polysaccharides of some ExPEC strains in conferring resistance the human serum, we examined the capsule region of twenty ExPEC strains, including ST131 and non-ST131 (Table 5.2), at a genomic level to identify and compare genetic loci present in the capsule region of these strains. Our data identified the presence of twenty-seven capsule loci, either with known or putative functions, in the capsule

region of all ExPEC ST131 and non-ST131 strains. Our data also identified a number of capsule loci that were encoding for hypothetical proteins of unknown functions. Comparing the distribution of these twenty-seven capsule loci between ST131 and non-ST131 strains showed the common presence of the eight highly conserved region 1 (*kpsFEDUCS*) and region 3 (*kpsTM*) loci in all ExPEC strains (Figure 5.4). However the other nineteen region 2 loci were variably distributed among ExPEC strains (Figure 5.4), the vast majority of which were found to be only present in the ST131 strains belonging to H30Rx clade. The observed difference in the distribution of capsule loci between ST131 and non-ST131 strains was interesting and drove us to examine the capsule region of ST131 H30Rx strains in more detail.

On the basis of previous reports showing a genetic (Petty *et al.* 2014) and biochemical (Olesen *et al.* 2013) diversity of the capsule region within *E. coli* ST131 strains belonging to H30Rx clade, the genetic architecture of capsule locus of *E. coli* ST131 was investigated in this study. Surprisingly, a high degree of diversity was observed between the conserved *kpsS* and *kpsTM* regions of the capsule locus, with no noticeable similarity between strains in the variable central genes. Blast analysis of each of the variable central genes in each genetic capsule type also showed no significant hits with any reference *E. coli* sequences expect for the *TagD-2* gene that was found in ST131 strain UTI32 and in the non-ST131 *E. coli* CFT073 strain. The diversity in the capsule locus of ST131 strains concurs the previous finding demonstrating the presence of several variant region 2 gene clusters between the conserved region 1 (*kpsFEDUCS*) and region 3 (*kpsTM*) of ST131 capsule locus (Petty *et al.* 2014).

To determine how this genetic architecture reflected upon biochemical typing, classical capsule typing was used to identify the K antigen type of each of the genetic capsule types for strains. The data revealed the presence a high degree of diversity in the biochemical profile of K antigens that was previously described in members belonging to ST131 H30Rx clade (Olesen *et al.* 2013). Although PCR-based methods have identified K2 antigen type in *E. coli* ST131 via using *kpsMII* primers (Johnson and O'Bryan 2004, Croxall *et al.* 2011a), the use of classical capsule typing was useful in the identification of diverse K antigen types in ST131 strains belonging to H30Rx clade. This finding seems surprising given the monomorphic nature of the H30Rx clade described previously by our research group (Clark *et al.* 2012). In general, our data showed a correlation between the genetic capsular type and the biochemical typing data, and provides a framework for which to contextualise *E. coli* capsule types from genomic data. More importantly for this study, the data clearly showed significant diversity within the capsule locus in *E. coli* ST131 H30Rx strains suggestive of frequent and targeted recombination in this region (Petty *et al.* 2014).

To investigate the genetic diversity of capsule locus in more details, a core genome phylogeny for *E. coli* ST131 was created using all the ST131 genome sequences and previously published methodology (Sahl *et al.* 2011). The resulting phylogenetic tree confirms that all of the strains in the analysis, including those previously sequenced by our group prior to the discovery of H30Rx (Clark *et al.* 2012), belonged to the H30Rx clade. It also showed that the capsule loci were randomly distributed across the phylogeny with the exception of the small cluster of strains containing

UTI18 that have been previously shown to be essentially a single clone (Clark *et al.* 2012). Such a random dispersal of the capsule loci across the phylogenetic tree can only be explained by extensive and targeted recombination events at this discreet location on the genome, suggesting there is some pressure acting on the capsule locus resulting in constant switching of capsule genes as the H30Rx clade evolves. Such extensive recombination has been well characterised in *Streptococcus pneumonia* where capsule locus switching has been shown to play a significant role in vaccine escape (Croucher *et al.* 2013) and in the evolutionary dynamics of densely populated infection foci (Chewapreecha *et al.* 2014). However such dynamism in capsular recombination in *E. coli* is hitherto uncharacterised particularly in such a genetically monomorphic clade as ST131 H30Rx.

Finally, to determine the effect of the extensive recombination at the capsule region of the ST131 H30Rx clade on other virulence-associated phenotypes *in-vitro*, this study tested the ability of members of ST131 strains belonging to H30Rx clade to form capsules and biofilm. The data showed no association between the capsule loci present in the different H30Rx strains and levels of capsulation morphology on agar plates. Similarly, classical 96-well plate biofilm formation assays also failed to identify any significant pattern between different H30Rx capsular variants. These observations, combined with the high serum resistance phenotype identified previously by this study for all ST131 strains, suggest that the presence of different capsular variants in ST131 strains belonging to H30Rx clade had no effect on the capability of these strains to form capsules, biofilm, and to resist the bactericidal activity of human serum. As mentioned earlier, although several reports have described the important role of

capsules in the serum resistance of ExPEC strains (Leying *et al.* 1990, Buckles *et al.* 2009), our data suggests that capsule type may be less important, and that the extensive capsular recombination demonstrated in the ST131 H30Rx clade has no effect on the ability of these pathogens to survive exposure to human serum.

5.5 Conclusion

E. coli ST131 has recently emerged as a globally disseminated, multi-drug-resistant clone responsible for a high proportion of urinary tract and bloodstream infections (Rogers *et al.* 2011). It has been shown that serum resistance is a major virulence trait for ExPEC that cause bloodstream infections (Phan *et al.* 2013), and investigation into serum resistance of *E. coli* has identified several factors that contribute to this phenotype (Miajlovic and Smith 2014). This study showed that all ST131 strains were associated with high serum resistance phenotype, and this might suggest serum resistance as an important factor in driving the current success of this clone.

Based on many reports showing the genetic and biochemical diversity in the capsule region of ST131 strains, testing the genetic architecture and K antigen biochemical profiles of the capsule region of these strains shows a surprising level of diversity within the capsule locus of the H30Rx clade with a phylogenetic distribution highly suggestive of frequent recombination at the locus. This recombination has no obvious detectable effect on virulence-associated phenotypes *in-vitro*. Given the level of diversity observed at the capsule locus it is tempting to speculate that there is significant selective pressure occurring at this site during the life cycle of the H30Rx clade, and

that frequent recombination allows the clade to subvert that pressure. This has been documented to occur in other capsulated pathogens (Chewapreecha *et al.* 2014) and also ties in with previous data from our group showing that ST131 strains did not exhibit inter-species recombination across the *E. coli* species but that rather recombination events were focussed within the ST131 lineage (McNally *et al.* 2013b). Temporal studies of ST131 populations from patients and environmental reservoirs may allow us to determine if capsular switching does occur *in vivo* and if it is an important mechanism in the successful and prolonged dissemination of this important human pathogen.

Chapter six

Final conclusions and future directions

6. Final conclusions and future directions

The current success of *E. coli* ST131 as a globally disseminated pathogen and its association with causing a wide range of difficult-to-treat infections (Rogers *et al.* 2011) highlight the importance of studying the factors that lead to this success and make it such a unique ExPEC clone. The experimental work of this PhD project started with performing a range of comparative assays in order to characterise and compare the traits of ST131 and non-ST131 at a phenotypic-level.

Firstly, the growth rates of a collection of *E. coli* strains, including *E. coli* ST131 and non-ST131 strains, were determined using two different growth assays, turbidity measurement assay and viable cell count assay, in order to examine the association between *E. coli* ST131 and differential growth profile *in vitro*. Our data obtained from measuring the OD value of 49 *E. coli* strains using 4 different growth conditions showed that all ST131 and non-ST131 strains had almost similar growth rates. This was subsequently confirmed by the growth curves obtained from performing viable cell counting assay on five ST131 and five non-ST131 strains, which also showed undetectable difference in the growth rates of the tested ST131 and non-ST131 strains.

Afterwards, on the basis of a previous study by our research group that identified the presence a *fimB* transposon insertion mutation in ten ST131 isolates (Clark *et al.* 2012), PCR screening of the *fimB* insertion in twenty-five *E. coli* ST131 strains showed that 21 of 25 ST131 strains (83.33%) had this insertion. Yeast cell agglutination assay was then performed to test the

effect of this insertion on the ability of *E. coli* ST131 strains to produce type 1 fimbriae, and our data showed that all strains were able to express functional type 1 fimbriae after two rounds of static growth, suggesting no effect of the *fimB* transposon insertion mutation on the capacity of *E. coli* ST131 strains to produce functional type 1 fimbriae.

Previous work by our research group that used gentamicin protection assay to compare the ability of a collection of *E. coli* strains to associate with and to invade T24 bladder epithelial cells (Croxall *et al.* 2011a), and identified the presence of a hyper-invasive phenotype in 56% of the tested ST131 strains. Based on that, we tested and compared the capacity of ten GFP-tagged ST131 and non-ST131 strains to attach to, invade and form IBCs within T24 cells using confocal microscopy, and identified the presence of strain-specific differences between strains, with no difference at ST-level between strains. Afterwards, determining the ability of ten ST131 and non-ST131 strains to persist within U937 macrophages showed also strain-specific difference between strains, with no specific ST131 persistence profile. Taken together, our comprehensive phenotypic characterisation of a collection of ExPEC strains belonging to different STs suggested no difference between ST131 and non-ST131 at a phenotypic-level, and highlighted the presence of other factors that can contribute to ST131 fitness.

Subsequently, given many previous studies demonstrating the association between high metabolic capacity and enhanced bacterial fitness (Le Bouguénec and Schouler 2011), and on the basis of a previous study showing that *E. coli* ST131 isolates had higher metabolic potential and

biochemical profiles compared to other ExPEC ST isolates, which may contribute to the fitness of *E. coli* ST131 (Gibreel *et al.* 2012a), this PhD project tested and compared the metabolic activity of a collection of ST131 and non-ST131 strains using two different testing methodologies: API strips and PM technology. Our API data showed that ST131 strains had a lower metabolic activity for 5 substrates, which is in contrary to a recent finding by Gibreel *et al.* (2012a) showing that ST131 strains had an increased metabolic potential compared to other ExPEC STs. Further testing of the metabolic activity of *E. coli* using phenotypic microarray demonstrated the absence of specific metabolic profile for ST131 strains suggesting that ST131 is not a metabolically distinct lineage of ExPEC and thus metabolism might not contribute to the fitness of this clone. In addition, comparative genome analysis also suggested very little difference in the repertoire of metabolic gene loci between ST131 and no-ST131. Our previous work describing the genetic homogeneity of ST131 (Clark *et al.* 2012) combined with the data presented here provide yet another confounding suggestion that *E. coli* ST131 do not display altered metabolic fitness to other closely related ExPEC from a global metabolic and genomic perspective. Since our findings and those of Gibreel *et al.* (2012a) were obtained from testing different strain sets using different methodologies, we propose that a coordinated international effort to fully understand the evolutionary mechanisms behind the emergence of *E. coli* ST131 is now crucial in order to combat this most serious of bacterial pandemics, and of future episodes of novel *E. coli* lineage emergence.

Based on a recent comparative genome study that only focussed on virulence associated genes and large mobile genetic elements unique to the

ST131 H30Rx clade (Petty *et al.* 2014), we investigated the gene content of a group of *E. coli* including ST131 and non-ST131 strains in order to identify the presence of other loci that are uniquely associated with ST131 H30Rx clade. Our gene content analysis of *E. coli* strains performed using Gegenees (Ågren *et al.* 2012) showed the presence of few genetic loci unique to ST131 H30Rx clade. Further investigation to identify these loci and the functional categories to which these loci belong to was performed using LS-BSR (Sahl *et al.* 2014), and our data identified the presence of 150 loci, the majority of which were hypothetical proteins, uniquely associated with this clade. In the future, it would be interesting to investigate the potential role of the hypothetical proteins identified as unique to ST131 H30Rx strains.

Generally, our data demonstrated that this clade is rather unremarkable in terms of its gene content compared to other ExPEC. Additionally, it identified the presence of a small set of metabolic loci that confirms our previous finding demonstrating that ST131 H30Rx is not a metabolically distinct lineage of ExPEC (Alqasim *et al.* 2014a). However the most striking finding at a genomic level was the identification of the secondary flagellar locus Flag 2 as a region uniquely associated with ST131 H30Rx strains (Totsika *et al.* 2011, Clark *et al.* 2012, Petty *et al.* 2014). Given the major role of flagella in the pathogenesis of *E. coli* strains (Haiko and Westerlund-Wikström 2013), and that the possession of Flag-2 locus may be a key factor for the survival of these strains in specific niches and/or hosts (Kirov 2003), this region merits further detailed bacterial genetics analysis to uncover its true importance to the emergence and success of the H30Rx clade.

It has been known that serum resistance is a major virulence trait for ExPEC that cause bloodstream infections (Phan *et al.* 2013), and investigation into serum resistance of *E. coli* has identified several factors that contribute to this phenotype (Miajlovic and Smith 2014). In this study, we tested the ability of a collection of ST131 and non-ST131 strains to resist human serum. Our data showed that all ST131 and ST73 strains were associated with high serum resistance phenotype, and this might suggest serum resistance as an important factor in driving the current success of this ST131 as a major cause of bloodstream infections worldwide.

The high serum resistance phenotype identified for ST131 strains drove us to explore the mechanisms involved in the serum resistance of *E. coli*. Given many reports showing that polysaccharide capsules might be a major factor allowing *E. coli* to resist the human serum (Leying *et al.* 1990, Buckles *et al.* 2009), and based on many studies demonstrating the genetic (Petty *et al.* 2014) and biochemical (Olesen *et al.* 2013) diversity in the capsule region of ST131 strains, the capsule region of a collection of ExPEC belonging to ST131 H30Rx clade and non-ST131 was tested in more detail at a genomic and biochemical levels. Our capsule genetics data showed a surprising level of diversity within the capsule locus of the H30Rx clade with a phylogenetic distribution highly suggestive of frequent recombination at the locus. Subsequent analysis demonstrated that this recombination has no obvious detectable effect on virulence-associated phenotypes *in-vitro*. Given the level of diversity observed at the capsule locus of ST131 H30Rx strains, it is tempting to speculate that there is significant selective pressure occurring at this site during the life cycle of the H30Rx clade, and that

frequent recombination allows the clade to subvert that pressure. This has recently been reported to occur in other capsulated pathogens (Chewapreecha *et al.* 2014) and also ties in with previous data from our group showing that ST131 strains did not exhibit inter-species recombination across the *E. coli* species but that rather recombination events were focussed within the ST131 lineage (McNally *et al.* 2013b). Temporal studies of ST131 populations from patients and environmental reservoirs may allow us to determine if capsular switching does occur *in vivo* and if it is an important mechanism in the successful and prolonged dissemination of this important human pathogen.

In conclusion, this PhD project characterised the traits that might give ST131 a fitness advantage over other successful ExPEC STs, and generally found no detectable difference at a phenotypic-level between ST131 and non-ST131. The only exception was the increased ability of ST131 strains to resist the bactericidal effect of human serum, although this finding was also observed for strains belonging to other successful ExPEC clone ST73. Our capsule genetics data showed a surprising level of diversity within the capsule locus of the H30Rx clade proposing the presence of frequent and extensive recombination at the locus. This recombination suggests that there is significant selective pressure occurring at this site during the life cycle of the ST131 H30Rx clade, which might provide a fitness advantage to ST131. Finally, our findings obtained from pan-genome analysis of a collection of ExPEC genomes identified the presence of the secondary flagellar locus Flag 2 in the functional category of genes uniquely associated with ST131 H30Rx strains, which might be a key factor for the survival of

these strains in specific niches and/or hosts and therefore in the pathogenesis.

Chapter seven

Appendix

7. Appendix

7.1 Metabolism work supplementary information

Supplementary table 7.1: List of the 120 API biochemical substrates used in the metabolic profiling assay.

Test	Substrate	Reaction
ONPG	ONPG	β-galactosidase production
ADH	Arginine	Arginine dihydrolase production
LDC	Lysine	Lysine decarboxylase production
ODC	Ornithine	Ornithine decarboxylase production
CIT	Citrate	Citrate utilisation
H_2S	Sodium thiosulfate	H2S production
URE	Urea	Urea hydrolysis
TDA	Tryptophan	Deaminase production
Indole	Tryptophan	Indole production
VP	Sodium pyruvate	Acetoin production
GEL	Charcoal gelatine	Gelatinase production
GLU	Glucose	Fermentation/oxidation
MAN	Mannitol	Fermentation/oxidation
INO	Inositol	Fermentation/oxidation
SOR	Sorbitol	Fermentation/oxidation
RHA	Rhamnose	Fermentation/oxidation
SAC	Sucrose	Fermentation/oxidation
MEL	Melibiose	Fermentation/oxidation
AMY	Amygdalin	Fermentation/oxidation
ARA	Arabinose	Fermentation/oxidation
LARL	L-Arabitol	Acidification
GAT	Galacturonate	Acidification
5KG	Potassium 5-keto-gluconate	Acidification
LIP	Lipase	Lipase production
RP	Phenol red	Acidification
βGLU	β-glucose	β-glucosidase
MAL	Maltose	Acidification
ADO	Adonitol	Acidification
PLE	Palatinose	Acidification
βGUR	β-glucuronidase	β-glucuronidase production
MNT	Malonate	Acidification
βNAG	N-acetyl-β-glucosaminidase	N-acetyl-β-glucosaminidase production
βGAL	β-galactosidase	β-galactosidase production
LARA	L-Arabinose	Acidification

DARL	D-Arabitol	Acidification
αGLU	α-glucosidase	α-glucosidase production
αGAL	α-galactosidase	α-galactosidase production
TRE	Trehalose	Acidification
CEL	Cellobiose	Acidification
αMAL	α-Maltosidase	α-Maltosidase production
AspA	L-Aspartic acid arylamidase	L-Aspartic acid arylamidase
GLY	Glycerol	Fermentation
ERY	Erythritol	Fermentation
DARA	D-arabinose	Fermentation
RIB	D-ribose	Fermentation
DXYL	D-xylose	Fermentation
LXYL	L-xylose	Fermentation
MDX	Methyl-β-D-xylopyranoside	Fermentation
GAL	D-galactose	Fermentation
FRU	D-fructose	Fermentation
MNE	D-mannose	Fermentation
SBE	L-sorbose	Fermentation
DUL	Dulcitol	Fermentation
MDM	Methyl-α-D-mannopyranoside	Fermentation
MDG	Methyl-α-D-glucopyranoside	Fermentation
NAG	N-acetyl-glucosamine	Fermentation
AMY	Amygdalin	Fermentation
ARB	Arbutin	Fermentation
ESC	Esculin	Fermentation
SAL	Salicin	Fermentation
LAC	D-lactose	Fermentation
INU	Inulin	Fermentation
MLZ	D-melezitose	Fermentation
RAF	D-raffinose	Fermentation
AMD	Amidon	Fermentation
GLYG	Glycogen	Fermentation
XLT	Xylitol	Fermentation
GEN	Gentiobiose	Fermentation
TUR	D-turanose	Fermentation
LYX	D-lyxose	Fermentation
TAG	D-tagatose	Fermentation
DFUC	D-fucose	Fermentation
LFUC	L-fucose	Fermentation
GNT	Potassium gluconate	Fermentation
2KG	Potassium 2-keto-gluconate	Fermentation
ALP	Alkaline phosphatase	Enzyme production
ESTC4	Esterase C 4	Enzyme production
ESTLipC8	Esterase Lipase C8	Enzyme production
LipC14	Lipase C 14	Enzyme production
Leu arylamidase	Leucine arylamidase	Enzyme production
Val arylamidase	Valine arylamidase	Enzyme production
Cys arylamidase	Cysteine arylamidase	Enzyme production
Trypsin	Trypsin	Enzyme production
α -chymotrypsin	α -chymotrypsin	Enzyme production

ACP	Acid phosphatase	Enzyme production
Naphthol-AS-BI-phosphohydrolyse	Naphthol-AS-BI-phosphohydrolyse	Enzyme production
α - galactosidase	α - galactosidase	Enzyme production
β - galactosidase	β - galactosidase	Enzyme production
β – glucuronidase	β – glucuronidase	Enzyme production
α – glucosidase	α – glucosidase	Enzyme production
β – glucosidase	β – glucosidase	Enzyme production
A-acetyl-β–glucosaminidase	N-acetyl-β–glucosaminidase	Enzyme production
α – mannosidase	α – mannosidase	Enzyme production
α - fucosidase	α - fucosidase	Enzyme production

Supplementary table 7.2: List of the 190 carbon sources used in the Biolog phenotypic microarray assay.

Well no.	Test	Well no.	Test
A01	Negative control	E01	L-Glutamine
A02	L-Arabinose	E02	m-Tartaric Acid
A03	N-Acetyl-D-glucosamine	E03	D-Glucose-1-Phosphate
A04	D-Saccharic Acid	E04	D-Fructose-6-Phosphate
A05	Succinic Acid	E05	Tween 80
A06	D-Galactose	E06	a-Hydroxy Glutaric Acid-g-Lactone
A07	L-Aspartic Acid	E07	a-Hydroxy-Butyric Acid
A08	L-Proline	E08	b-Methyl-D-Glucoside
A09	D-Alanine	E09	Adonitol
A10	D-Trehalose	E10	Maltotriose
A11	D-Mannose	E11	2`-Deoxy-Adenosine
A12	Dulcitol	E12	Adenosine
B01	D-serine	F01	Glycyl-L-Aspartic Acid
B02	D-sorbitol	F02	Citric Acid
B03	Glycerol	F03	m-Inositol
B04	L-Fucose	F04	D-Threonine
B05	D-Glucuronic Acid	F05	Fumaric Acid
B06	D-Gluconic Acid	F06	Bromo-Succinic Acid
B07	D, L-a-Glycerol-Phosphate	F07	Propionic Acid
B08	D-xylose	F08	Mucic Acid
B09	L-Lactic Acid	F09	Glycolic Acid
B10	Formic Acid	F10	Glyoxylic Acid
B11	D-mannitol	F11	D-cellobiose
B12	L-Glutamic Acid	F12	Inosine
C01	D-glucose-6-phosphate	G01	Glycyl-L-Glutamic Acid
C02	D-Galactonic Acid-g-Lactone	G02	Tricarballylic Acid
C03	D, L-Malic Acid	G03	L-Serine
C04	D-Ribose	G04	L-Threonine
C05	Tween 20	G05	L-Alanine
C06	L-rhamnose	G06	L-Alanyl-Glycine
C07	D-Fructose	G07	Acetoacetic Acid

C08	Acetic Acid	G08	N-Acetyl-b-D-Mannosamine
C09	a-D-glucose	G09	Mono Methyl Succinate
C10	Maltose	G10	Methyl Pyruvate
C11	D-Melibiose	G11	D-Malic Acid
C12	Thymidine	G12	L-Malic Acid
D01	L-Asparagine	H01	Glycyl-L-Proline
D02	D-Aspartic Acid	H02	p-Hydroxy-Phenylacetic Acid
D03	D-Glucosaminic Acid	H03	m-Hydroxy-Phenylacetic Acid
D04	1,2-Propanediol	H04	Tyramine
D05	Tween 40	H05	D-Psicose
D06	a-Keto-Glutaric Acid	H06	L-Lyxose
D07	a-Keto-Butyric Acid	H07	Glucuronamide
D08	a-Methyl-D-Galactoside	H08	Pyruvic Acid
D09	a-D-Lactose	H09	L-Galactonic Acid-g-Lactone
D10	Lactulose	H10	D-Galacturonic Acid
D11	Sucrose	H11	b-Phenylethylamine
D12	Uridine	H12	Ethanolamine

Supplementary table 7.3: Metabolic profiling results for 50 *E. coli* ST131 and non-ST131 strains obtained from using 120 API biochemical substrates.

Test	**ST131 (n=25)**	**ST69 (n=7)**	**ST73 (n=7)**	**ST95 (n=7)**	**ST10 (n=4)**
ONPG +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
ADH +	13/13 52%/52%	2/2 28%/28%	4/4 57%/57%	1/1 14%/14%	1/1 25%/25%
LDC +	22/22 88%/88%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
ODC +	25/25 100%/100%	1/1 14%/14%	7/7 100%/100%	7/7 100%/100%	2/2 50%/50%
Citrate +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
H_2S +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
Urease +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
TDA+	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
Indole +	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
VP +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GEL +	0/0	0/0	0/0	0/0	0/0

	0%/0%	0%/0%	0%/0%	0%/0%	0%/0%
Glucose +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
MAN +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
INO +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
SOR +	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	3/3 75%/75%
RHA +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
SAC +	24/24 96%/96%	7/7 100%/100%	3/3 42%/42%	1/1 14%/14%	1/1 25%/25%
MEL +	23/23 92%/92%	7/7 100%/100%	5/5 71%/71%	7/7 100%/100%	4/4 100%/100%
AMY +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
ARA +	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
Nitrate → nitrites	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
LARL +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GAT +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
5KG +	2/2 8%/8%	7/7 100%/100%	6/6 85%/85%	4/4 57%/57%	3/3 75%/75%
Lipase +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
RP +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
βglu +	2/2 8%/8%	1/1 14%/14%	1/1 14%/14%	1/1 14%/14%	1/1 25%/25%
MAL +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
ADO +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
PLE +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
βGUR +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%

MNT +	1/1 4%/4%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
βNAG +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
βGAL +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
LARA +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
DARL +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
αGLU +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
αGAL +	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
TRE +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
RHA +	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
INO +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
CEL +	0/0 0%/0%	3/3 43%/43%	3/3 43%/43%	2/2 28%/28%	2/2 50%/50%
αMAL +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
ASPA +	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GLY	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
ERY	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
DARA	3/5 12%/20%	0/7 0%/100%	0/6 0%/85%	0/7 0%/100%	1/3 25%/75%
LARA	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
RIB	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
DXYL	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	5/5 71%/71%	4/4 100%/100%
LXYL	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
ADO	0/0	1/1	1/2	1/1	0/0

	0%/0%	14%/14%	14%/28%	14%/14%	0%/0%
MDX	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GAL	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
GLU	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
FRU	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
MNE	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
SBE	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	5/5 71%/71%	1/1 25%/25%
RHA	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
DUL	4/9 16%/36%	6/7 85%/100%	5/6 71%/85%	4/5 57%/71%	3/3 75%/75%
INO	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
MAN	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
SOR	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	3/3 75%/75%
MDM	0/0 0%/0%	0/1 0%/14%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
MDG	1/2 4%/8%	0/1 0%/14%	0/0 0%/0%	1/1 14%/14%	0/0 0%/0%
NAG	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
AMY	0/2 0%/8%	2/3 28%/42%	1/1 14%/14%	3/3 43%/43%	1/2 25%/50%
ARB	2/5 8%/20%	3/3 42%/42%	2/4 28%/57%	2/2 28%/28%	2/3 50%/75%
ESC	4/6 16%/24%	3/3 42%/42%	3/7 42%/100%	4/4 57%/57%	4/4 100%/100%
SAL	2/7 8%/28%	3/3 42%/42%	2/7 28%/100%	2/3 28%/43%	2/4 50%/100%
CEL	0/0 0%/0%	3/3 43%/43%	2/3 28%/43%	2/2 28%/28%	1/2 25%/50%
MAL	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%

LAC	23/23 92%/92%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
MEL	23/23 92%/92%	7/7 100%/100%	5/7 71%/100%	5/5 71%/71%	4/4 100%/100%
SAC	25/25 100%/100%	7/7 100%/100%	5/6 71%/85%	3/3 43%/43%	3/3 75%/75%
TRE	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
INU	0/0 0%/0%	1/2 14%/28%	1/1 14%/14%	3/3 43%/43%	1/1 25%/25%
MLZ	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
RAF	15/23 60%/92%	7/7 100%/100%	5/6 71%/85%	4/4 57%/57%	3/3 75%/75%
AMD	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GLYG	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
XLT	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GEN	4/4 16%/16%	3/3 42%/42%	2/5 28%/71%	2/2 28%/28%	1/1 25%/25%
TUR	1/3 4%/12%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
LYX	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
TAG	24/24 96%/96%	1/2 14%/28%	7/7 100%/100%	4/5 57%/71%	1/1 25%/25%
DFUC	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/1 0%/14%	0/0 0%/0%
LFUC	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
DARL	0/0 0%/0%	1/1 14%/14%	1/2 14%/28%	1/1 14%/14%	0/0 0%/0%
LARL	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
GNT	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
2KG	1/3 4%/12%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
5KG	2/2	7/7	6/6	4/4	3/3

	8%/8%	100%/100%	85%/85%	57%/57%	75%/75%
ALP	24/24 96%/96%	7/7 100%/100%	6/6 85%/85%	7/7 100%/100%	4/4 100%/100%
EST C4	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
EST LIP C8	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
Lipase C14	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
LAR	24/24 96%/96%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
VAR	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
CAR	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
Trypsin	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
α–c-trypsin	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
ACP	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
Naph-AS-BI-phosph	1/1 2%/2%	2/2 28%/28%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
α- galactosidase	5/5 20%/20%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	1/1 25%/25%
β- galactosidase	25/25 100%/100%	7/7 100%/100%	7/7 100%/100%	7/7 100%/100%	4/4 100%/100%
β–glucuronidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	2/2 50%/50%
α– glucosidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
β – glucosidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
N-acetyl-β–glucosaminidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
α – mannosidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%
α - fucosidase	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%	0/0 0%/0%

7.2 Bioinformatics software scripts and command lines

1. *De novo* Velvet assembly (Zerbino and Birney 2008)

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

2. PAGIT assembly script

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)
ls
source sourceme.pagit
./dorun.sh

3. Prokka annotation (Seemann 2014)

Prokka annotation was performed using the command line:

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

4. Mugsy alignment command line (Angiuoli and Salzberg 2011)

Mugsy alignment was performed using the command line:

source mugsyenv.sh
./mugsy --directory /home/alan/Desktop sample1.fasta sample2.fasta sample3.fa sample4.fa sample5.fa sample6.fa sample7.fa sample8.fa sample9.fa sample10.fa sample11.fa

5. RaxML maximum likelihood phylogeny command line (Stamatakis *et al.* 2005)

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. Compare_bsr.py script in LS-BSR software (Sahl *et al.* 2014)

This script is used to looks for CDS differences between two user-defined populations. The "names.txt" file contains the names, as they should be listed in your separate groups file. The compare_bsr.py script requirements include: BSR matrix, two new-line delimited group files, taken from "names.txt" and FASTA file of all CDS sequences. The command line used was:

$phyton compare_BSR .py -1 group1.txt -2 group2.txt –f consensus.fasta –b_bsr_matrix_values.txt

References

ABRAHAM, J. M., FREITAG, C. S., CLEMENTS, J. R. & EISENSTEIN, B. I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proceedings of the National Academy of Sciences U S A,* **82,** 5724-5727.

ADAMS-SAPPER, S., DIEP, B. A., PERDREAU-REMINGTON, F. & RILEY, L. W. (2012). Clonal composition and community clustering of drug-susceptible and resistant *Escherichia coli* isolates from blood stream infections. *Antimicrobial Agents and Chemotherapy***,** AAC. 01025-12.

ÅGREN, J., SUNDSTRÖM, A., HÅFSTRÖM, T. & SEGERMAN, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One,* **7,** e39107.

AKRAM, M., SHAHID, M. & KHAN, A. U. (2007). Etiology and antibiotic resistance patterns of community-acquired urinary tract infections in J N M C Hospital Aligarh, India. *Annals of Clinical Microbiology and Antimicrobials,* **6,** 1-7.

ALHASHASH, F., WESTON, V., DIGGLE, M. & MCNALLY, A. (2013). Multidrug-Resistant *Escherichia coli* Bacteremia. *Emerging Infectious Diseases,* **19,** 1699.

ALQASIM, A., EMES, R., CLARK, G., NEWCOMBE, J., LA RAGIONE, R. & MCNALLY, A. (2014a). Phenotypic microarrays suggest *Escherichia coli* ST131 Is not a metabolically distinct lineage of extra-intestinal pathogenic *E. coli*. *PloS One,* **9,** e88374.

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

ALTERI, C. J., SMITH, S. N. & MOBLEY, H. L. (2009). Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathogens,* **5,** e1000448.

ANDERSEN, P. S., STEGGER, M., AZIZ, M., CONTENTE-CUOMO, T., GIBBONS, H. S., KEIM, P., SOKURENKO, E. V., JOHNSON, J. R. & PRICE, L. B. (2013). Complete genome sequence of the epidemic and highly virulent CTX-M-15-producing H30-Rx subclone of *Escherichia coli* ST131. *Genome Announcements,* **1,** e00988-13.

ANDERSON, G. G., PALERMO, J. J., SCHILLING, J. D., ROTH, R., HEUSER, J. & HULTGREN, S. J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. *Science,* **301,** 105-107.

ANDERSON, J., EFTEKHAR, F., AIRD, M. & HAMMOND, J. (1979). Role of bacterial growth rates in the epidemiology and pathogenesis of urinary infections in women. *Journal of Clinical Microbiology,* **10,** 766-771.

ANFORA, A. T., HAUGEN, B. J., ROESCH, P., REDFORD, P. & WELCH, R. A. (2007). Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by *Escherichia coli* CFT073. *Infection and Immunity,* **75,** 5298-5304.

ANGIUOLI, S. V. & SALZBERG, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics,* **27,** 334-342.

ARTHUR, M., JOHNSON, C. E., RUBIN, R. H., ARBEIT, R. D., CAMPANELLI, C., KIM, C., STEINBACH, S., AGARWAL, M., WILKINSON, R. & GOLDSTEIN, R. (1989). Molecular epidemiology of adhesin and hemolysin virulence factors among uropathogenic *Escherichia coli*. *Infection and Immunity,* **57,** 303-313.

ASHKAR, A. A., MOSSMAN, K. L., COOMBES, B. K., GYLES, C. L. & MACKENZIE, R. (2008). FimH adhesin of type 1 fimbriae is a potent inducer of innate antimicrobial responses which requires TLR4 and type 1 interferon signalling. *PLoS Pathogens,* **4,** e1000233.

ASSIMACOPOULOS, A., JOHNSTON, B., CLABOTS, C. & JOHNSON, J. R. (2012). Post-prostate biopsy infection with *Escherichia coli* ST131 leading to epididymo-orchitis and meningitis caused by Gram-negative bacilli. *Journal of Clinical Microbiology,* **50,** 4157-4159.

AVASTHI, T. S., KUMAR, N., BADDAM, R., HUSSAIN, A., NANDANWAR, N., JADHAV, S. & AHMED, N. (2011). Genome of multidrug-resistant uropathogenic *Escherichia coli* strain NA114 from India. *Journal of Bacteriology,* **193,** 4272-4273.

BARER, M. (2001). Bacterial growth, culturability and viability. *In:* TANG, Y.-W., LIU, D., POXTON, I., SCHWARTZMAN, J. & SUSSMAN, M. (eds.) *Molecular Medical Microbiology, Three-Volume Set.* London: Academic Press.

BARONDESS, J. J. & BECKWITH, J. (1995). *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *Journal of Bacteriology,* **177,** 1247-1253.

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

BIEN, J., SOKOLOVA, O. & BOZKO, P. (2012). Role of uropathogenic *Escherichia coli* virulence factors in development of urinary tract infection and kidney damage. *International Journal of Nephrology,* **2012**.

BISERCIĆ, M., FEUTRIER, J. Y. & REEVES, P. R. (1991). Nucleotide sequences of the gnd genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic gnd locus. *Journal of Bacteriology,* **173,** 3894-3900.

BLANC, V., LEFLON-GUIBOUT, V., BLANCO, J., HAENNI, M., MADEC, J.-Y., RAFIGNON, G., BRUNO, P., MORA, A., LOPEZ, C. & DAHBI, G. (2014). Prevalence of day-care centre children (France) with faecal CTX-M-producing *Escherichia coli* comprising O25b: H4 and O16: H5 ST131 strains. *Journal of Antimicrobial Chemotherapy***,** dkt519.

BLANGO, M. G. & MULVEY, M. A. (2010). Persistence of uropathogenic *Escherichia coli* in the face of multiple antibiotics. *Antimicrobial Agents and Chemotherapy,* **54,** 1855-1863.

BOCHNER, B. R. (2003). New technologies to assess genotype-phenotype relationships. *Nature Reviews Genetics,* **4,** 309-314.

BOCHNER, B. R., GADZINSKI, P. & PANOMITROS, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research,* **11,** 1246-1255.

BOKIL, N. J., TOTSIKA, M., CAREY, A. J., STACEY, K. J., HANCOCK, V., SAUNDERS, B. M., RAVASI, T., ULETT, G. C., SCHEMBRI, M. A. & SWEET, M. J. (2011). Intramacrophage survival of uropathogenic *Escherichia coli*: Differences between diverse clinical isolates and between mouse and human macrophages. *Immunobiology,* **216,** 1164-1171.

BOYD, E. F. & HARTL, D. L. (1998). Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *Journal of Bacteriology,* **180,** 1159-1165.

BRESOLIN, G., TRČEK, J., SCHERER, S. & FUCHS, T. M. (2008). Presence of a functional flagellar cluster Flag-2 and low-temperature expression of flagellar genes in *Yersinia enterocolitica* W22703. *Microbiology,* **154,** 196-206.

BRZUSZKIEWICZ, E., THÜRMER, A., SCHULDES, J., LEIMBACH, A., LIESEGANG, H., MEYER, F.-D., BOELTER, J., PETERSEN, H., GOTTSCHALK, G. & 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). *Archives of Microbiology,* **193,** 883-891.

BUCKLES, E. L., WANG, X., LANE, M. C., LOCKATELL, C. V., JOHNSON, D. E., RASKO, D. A., MOBLEY, H. L. & DONNENBERG, M. S. (2009). Role of the K2 capsule in *Escherichia coli* urinary tract infection and serum resistance. *Journal of Infectious Diseases,* **199,** 1689-1697.

BURNS, S. M. & HULL, S. I. (1998). Comparison of Loss of Serum Resistance by Defined Lipopolysaccharide Mutants and an Acapsular Mutant of Uropathogenic *Escherichia coli* O75: K5. *Infection and Immunity,* **66,** 4244-4253.

CAGNACCI, S., GUALCO, L., DEBBIA, E., SCHITO, G. C. & 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,** 2605-12.

CANALS, R., ALTARRIBA, M., VILCHES, S., HORSBURGH, G., SHAW, J. G., TOMÁS, J. M. & MERINO, S. (2006). Analysis of the lateral flagellar gene system of *Aeromonas hydrophila* AH-3. *Journal of Bacteriology,* **188,** 852-862.

CARVER, T. J., RUTHERFORD, K. M., BERRIMAN, M., RAJANDREAM, M.-A., BARRELL, B. G. & PARKHILL, J. (2005). ACT: the Artemis comparison tool. *Bioinformatics,* **21,** 3422-3423.

CHAIN, P., GRAFHAM, D., FULTON, R., FITZGERALD, M., HOSTETLER, J., MUZNY, D., ALI, J., BIRREN, B., BRUCE, D. & BUHAY, C. (2009). Genome project standards in a new era of sequencing. *Science (New York, NY),* **326**.

CHASSAING, B. & DARFEUILLE-MICHAUD, A. (2013). Adherent-invasive *Escherichia coli* biofilm formation assays. *Bio-protocol,* **13**.

CHAUDHURI, R. R. & HENDERSON, I. R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infection, Genetics and Evolution,* **12,** 214-226.

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

CHEN, S. L., HUNG, C.-S., XU, J., REIGSTAD, C. S., MAGRINI, V., SABO, A., BLASIAR, D., BIERI, T., MEYER, R. R. & OZERSKY, P. (2006b). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia col*i: a comparative genomics approach. *Proceedings of the National Academy of Sciences,* **103,** 5977-5982.

CHEWAPREECHA, C., HARRIS, S. R., CROUCHER, N. J., TURNER, C., MARTTINEN, P., CHENG, L., PESSIA, A., AANENSEN, D. M., MATHER, A. E. & PAGE, A. J. (2014). Dense genomic sampling identifies highways of pneumococcal recombination. *Nature Genetics,* **46,** 305-309.

CHRISTEN, B., ABELIUK, E., COLLIER, J. M., KALOGERAKI, V. S., PASSARELLI, B., COLLER, J. A., FERO, M. J., MCADAMS, H. H. & SHAPIRO, L. (2011). The essential genome of a bacterium. *Molecular Systems Biology,* **7**.

CIRL, C., WIESER, A., YADAV, M., DUERR, S., SCHUBERT, S., FISCHER, H., STAPPERT, D., WANTIA, N., RODRIGUEZ, N. & WAGNER, H. (2008). Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain–containing proteins. *Nature Medicine,* **14,** 399-406.

CLARK, G., PASZKIEWICZ, K., HALE, J., WESTON, V., CONSTANTINIDOU, C., PENN, C., ACHTMAN, M. & 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,** 868-877.

CLARKE, D. J. (2010). The Rcs phosphorelay: more than just a two-component pathway. *Future Microbiology,* **5,** 1173-1184.

COELHO, A., MORA, A., MAMANI, R., LOPEZ, C., GONZALEZ-LOPEZ, J. J., LARROSA, M. N., QUINTERO-ZARATE, J. N., DAHBI, G., HERRERA, A., BLANCO, J. E., BLANCO, M., ALONSO, M. P., PRATS, G. & BLANCO, J. (2010). Spread of *Escherichia coli* O25b:H4-B2-ST131 producing CTX-M-15 and SHV-12 with high virulence gene content in Barcelona (Spain). *Journal of Antimicrobial Chemotherapy,* **66,** 1-10.

COQUE, T. M., NOVAIS, Â., CARATTOLI, A., POIREL, L., PITOUT, J., PEIXE, L., BAQUERO, F., CANTÓN, R. & NORDMANN, P. (2008). Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum -lactamase CTX-M-15. *Emerging Infectious Diseases,* **14,** 195-200.

COVE SMITH, A. & ALMOND, M. (2007). Management of urinary tract infections in the elderly. *Trends in Urology, Gynaecology & Sexual Health,* **12,** 31-34.

CROSA, J. H. (1989). Genetics and molecular biology of siderophore-mediated iron transport in bacteria. *Microbiological Reviews,* **53,** 517-530.

CROSS, A. S., KIM, K. S., WRIGHT, D. C., SADOFF, J. C. & GEMSKI, P. (1986). Role of lipopolysaccharide and capsule in the serum resistance of bacteremic strains of *Escherichia coli*. *Journal of Infectious Diseases,* **154,** 497-503.

CROUCHER, N. J., FINKELSTEIN, J. A., PELTON, S. I., MITCHELL, P. K., LEE, G. M., PARKHILL, J., BENTLEY, S. D., HANAGE, W. P. & LIPSITCH, M. (2013). Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nature Genetics,* **45,** 656-663.

CROXALL, G., HALE, J., WESTON, V., MANNING, G., CHEETHAM, P., ACHTMAN, M. & MCNALLY, A. (2011a). 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,** 2501-2508.

CROXALL, G., WESTON, V., JOSEPH, S., MANNING, G., CHEETHAM, P. & MCNALLY, A. (2011b). Increased human pathogenic potential of *Escherichia coli* from polymicrobial urinary tract infections in comparison to isolates from monomicrobial culture samples. *Journal of Medical Microbiology,* **60,** 102-109.

CROXEN, M. A., LAW, R. J., SCHOLZ, R., KEENEY, K. M., WLODARSKA, M. & FINLAY, B. B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clinical Microbiology Reviews,* **26,** 822-880.

DA SILVA, G. J. & MENDONÇA, N. (2012). Association between antimicrobial resistance and virulence in *Escherichia coli*. *Virulence,* **3,** 18-28.

DAHBI, G., MORA, A., LÓPEZ, C., ALONSO, M. P., MAMANI, R., MARZOA, J., COIRA, A., GARCÍA-GARROTE, F., PITA, J. M. & VELASCO, D. (2013). Emergence of new variants of ST131 clonal group among extraintestinal pathogenic *Escherichia coli* producing extended-spectrum β-lactamases. *International Journal of Antimicrobial Agents,* **42,** 347-351.

DAVIES, B. (1977). Biochemical typing of urinary *Escherichia coli* strains by means of the API 20 E enterobacteriaceae system. *Journal of Medical Microbiology,* **10,** 293-298.

DAVIS, J. M., CARVALHO, H. M., RASMUSSEN, S. B. & O'BRIEN, A. D. (2006). Cytotoxic necrotizing factor type 1 delivered by outer membrane vesicles of uropathogenic *Escherichia coli* attenuates polymorphonuclear leukocyte antimicrobial activity and chemotaxis. *Infection and Immunity,* **74,** 4401-4408.

DE LOUVOIS, J. (1994). Acute bacterial meningitis in the newborn. *Journal of Antimicrobial Chemotherapy,* **34,** 61-73.

DHANJI, H., DOUMITH, M., ROONEY, P. J., O'LEARY, M. C., LOUGHREY, A. C., HOPE, R., WOODFORD, N. & LIVERMORE, D. M. (2011). Molecular epidemiology of fluoroquinolone-resistant ST131 *Escherichia coli* producing CTX-M extended-spectrum β-lactamases in nursing homes in Belfast, UK. *Journal of Antimicrobial Chemotherapy,* **66,** 297-303.

DÍAZ, M. A., HERNÁNDEZ-BELLO, J. R., RODRÍGUEZ-BAÑO, J., MARTÍNEZ-MARTÍNEZ, L., CALVO, J., BLANCO, J. & PASCUAL, A. (2010). Diversity of *Escherichia coli* strains producing extended-spectrum β-lactamases in Spain: second nationwide study. *Journal of Clinical Microbiology,* **48,** 2840-2845.

DOBRINDT, U. (2005). (Patho-) Genomics of *Escherichia coli*. *International Journal of Medical Microbiology,* **295,** 357-371.

DURSO, L. M., SMITH, D. & HUTKINS, R. W. (2004). Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157: H7 strains. *Applied and Environmental Microbiology,* **70,** 6466-6472.

EBERLE, K. & KIESS, A. (2012). Phenotypic and genotypic methods for typing *Campylobacter jejuni* and *Campylobacter coli* in poultry. *Poultry Science,* **91,** 255-264.

EDÉN, C. S. (1986). Bacterial adherence in urinary tract infections caused by *Escherichia coli*. *Scandinavian Journal of Urology and Nephrology,* **20,** 81-88.

EDWARDS, J. S., IBARRA, R. U. & PALSSON, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology,* **19,** 125-130.

ELSINGHORST, E. A. (1994). Measurement of invasion by gentamicin resistance. *Methods in Enzymology,* **236,** 405-420.

ENDER, P. T., GAJANANA, D., JOHNSTON, B., CLABOTS, C., TAMARKIN, F. J. & JOHNSON, J. R. (2009). Transmission of an extended-spectrum-beta-lactamase-producing *Escherichia coli* (sequence type ST131) strain between a father and daughter resulting in septic shock and emphysematous pyelonephritis. *Journal of Clinical Microbiology,* **47,** 3780-3782.

ENGEL, D. R., MAURER, J., TITTEL, A. P., WEISHEIT, C., CAVLAR, T., SCHUMAK, B., LIMMER, A., VAN ROOIJEN, N., TRAUTWEIN, C. & TACKE, F. (2008). CCR2 mediates homeostatic and inflammatory release of Gr1high monocytes from the bone marrow, but is dispensable for bladder infiltration in bacterial urinary tract infection. *The Journal of Immunology,* **181,** 5579-5586.

ESCHERICH, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. *Review of Infectious diseases,* **10,** 1220-1225.

ESCOBAR-PÁRAMO, P., CLERMONT, O., BLANC-POTARD, A.-B., BUI, H., LE BOUGUÉNEC, C. & DENAMUR, E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Molecular Biology and Evolution,* **21,** 1085-1094.

FEIL, E. J. (2004). Small change: keeping pace with microevolution. *Nature Reviews Microbiology,* **2,** 483-495.

FEINGOLD, D. S. (1969). The serum bactericidal reaction. IV. Phenotypic conversion of *Escherichia coli* from serum-resistance to serum-sensitivity by diphenylamine. *The Journal of infectious diseases***,** 437-444.

FELDMANN, F., SORSA, L. J., HILDINGER, K. & SCHUBERT, S. (2007). The salmochelin siderophore receptor IroN contributes to invasion of urothelial cells by extraintestinal pathogenic *Escherichia coli in vitro*. *Infection and Immunity,* **75,** 3183-3187.

FOXMAN, B. (2002). Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *The American Journal of Medicine,* **113,** 5-13.

FOXMAN, B. (2010). The epidemiology of urinary tract infection. *Nature Reviews Urology,* **7,** 653-660.

GARCIA, E. C., BRUMBAUGH, A. R. & MOBLEY, H. L. (2011). Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infection and Immunity,* **79,** 1225-1235.

GAUGER, E. J., LEATHAM, M. P., MERCADO-LUBO, R., LAUX, D. C., CONWAY, T. & COHEN, P. S. (2007). Role of motility and the flhDC operon in

Escherichia coli MG1655 colonization of the mouse intestine. *Infection and Immunity,* **75,** 3315-3324.

GEMSKI, P., CROSS, A. & SADOFF, J. (1980). K1 antigen - associated resistance to the bactericidal activity of serum. *FEMS Microbiology Letters,* **9,** 193-197.

GENEVAUX, P., BAUDA, P., DUBOW, M. S. & OUDEGA, B. (1999). Identification of Tn10 insertions in the *rfaG*, *rfaP*, and *galU* genes involved in lipopolysaccharide core biosynthesis that affect *Escherichia coli* adhesion. *Archives of Microbiology,* **172,** 1-8.

GIBREEL, T. M., DODGSON, A. R., CHEESBROUGH, J., BOLTON, F. J., FOX, A. J. & UPTON, M. (2012a). High metabolic potential may contribute to the success of ST131 uropathogenic *Escherichia coli*. *Journal of Clinical Microbiology,* **50,** 3202-3207.

GIBREEL, T. M., DODGSON, A. R., CHEESBROUGH, J., FOX, A. J., BOLTON, F. J. & UPTON, M. (2012b). Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *Journal of Antimicrobial Chemotherapy,* **67,** 346-356.

GOBERT, A. P., CHENG, Y., WANG, J.-Y., BOUCHER, J.-L., IYER, R. K., CEDERBAUM, S. D., CASERO, R. A., NEWTON, J. C. & WILSON, K. T. (2002). *Helicobacter pylori* induces macrophage apoptosis by activation of arginase II. *The Journal of Immunology,* **168,** 4692-4700.

GOERING, R. V. (2010). Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. *Infection, Genetics and Evolution,* **10,** 866-875.

GOLDING, G. R., PERSAUD, N., LEVETT, P. N., MCDONALD, R. R., IRVINE, J., NSUNGU, M., WOODS, S., KHAN, M., MATASEJE, L. F. & MULVEY, M. R. (2012). Characterization of *Escherichia coli* urinary tract infection isolates in remote northern Saskatchewan communities: the Northern Antibiotic Resistance Partnership. *Diagnostic Microbiology and Infectious Disease,* **74,** 242-247.

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

GROISMAN, E. A. & OCHMAN, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell,* **87,** 791-794.

GROZDANOV, L., RAASCH, C., SCHULZE, J., SONNENBORN, U., GOTTSCHALK, G., HACKER, J. & DOBRINDT, U. (2004). Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *Journal of Bacteriology,* **186,** 5432-5441.

GUPTA, K., SCHOLES, D. & STAMM, W. E. (1999). Increasing prevalence of antimicrobial resistance among uropathogens causing acute uncomplicated cystitis in women. *The Journal of the American Medical Association,* **281,** 736-738.

GUYER, D. M., HENDERSON, I. R., NATARO, J. P. & MOBLEY, H. L. (2000). Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. *Molecular Microbiology,* **38,** 53-66.

GUYER, D. M., KAO, J.-S. & MOBLEY, H. L. (1998). Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheterassociated bacteriuria and from fecal samples. *Infection and Immunity,* **66,** 4411-4417.

GUYER, D. M., RADULOVIC, S., JONES, F.-E. & MOBLEY, H. L. (2002). Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. *Infection and Immunity,* **70,** 4539-4546.

HABEEB, M., HAQUE, A., IVERSEN, A. & GISKE, C. (2014). Occurrence of virulence genes, 16S rRNA methylases, and plasmid-mediated quinolone resistance genes in CTX-M-producing *Escherichia coli* from Pakistan.

European Journal of Clinical Microbiology & Infectious Diseases, **33,** 399-409.

HACKER, J., BENDER, L., OTT, M., WINGENDER, J., LUND, B., MARRE, R. & GOEBEL, W. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extra intestinal *Escherichia coli* isolates. *Microbial Pathogenesis,* **8,** 213-225.

HAIKO, J. & WESTERLUND-WIKSTRÖM, B. (2013). The role of the bacterial flagellum in adhesion and virulence. *Biology,* **2,** 1242-1267.

HALLIN, M., DEPLANO, A. & STRUELENS, M. J. (2012). Molecular typing of bacterial pathogens: A tool for the epidemiological study and control of infectious diseases. *New Frontiers of Molecular Epidemiology of Infectious Diseases.* Springer.

HANCOCK, V., FERRIERES, L. & 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,** 167-175.

HANNAN, T. J., TOTSIKA, M., MANSFIELD, K. J., MOORE, K. H., SCHEMBRI, M. A. & HULTGREN, S. J. (2012). Host–pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic *Escherichia coli* bladder infection. *FEMS Microbiology Reviews,* **36,** 616-648.

HEJNOVA, J., DOBRINDT, U., NEMCOVA, R., RUSNIOK, C., BOMBA, A., FRANGEUL, L., HACKER, J., GLASER, P., SEBO, P. & BUCHRIESER, C. (2005). Characterization of the flexible genome complement of the commensal *Escherichia coli* strain A0 34/86 (O83: K24: H31). *Microbiology,* **151,** 385-398.

HERIAS, M. V., MIDTVEDT, T., HANSON, L. A. & WOLD, A. E. (1995). Role of *Escherichia coli* P fimbriae in intestinal colonization in gnotobiotic rats. *Infection and Immunity,* **63,** 4781-4789.

HERIAS, M. V., MIDTVEDT, T., HANSON, L. A. & WOLD, A. E. (1997). *Escherichia coli* K5 capsule expression enhances colonization of the large intestine in the gnotobiotic rat. *Infection and Immunity,* **65,** 531-536.

HERZER, P. J., INOUYE, S., INOUYE, M. & WHITTAM, T. S. (1990). Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *Journal of Bacteriology,* **172,** 6175-6181.

HILBERT, D. W., PASCAL, K. E., LIBBY, E. K., MORDECHAI, E., ADELSON, M. E. & TRAMA, J. P. (2008). Uropathogenic *Escherichia coli* dominantly suppress the innate immune response of bladder epithelial cells by a lipopolysaccharide-and Toll-like receptor 4-independent pathway. *Microbes and Infection,* **10,** 114-121.

HOMANN, O. R., CAI, H., BECKER, J. M. & LINDQUIST, S. L. (2005). Harnessing natural diversity to probe metabolic pathways. *PLoS Genetics,* **1,** e80.

HOOTON, T. M. & STAMM, W. E. (1997). Diagnosis and treatment of uncomplicated urinary tract infection. *Infectious Disease Clinics of North America,* **11,** 551-581.

HOWARD, C. & GLYNN, A. (1971). The virulence for mice of strains of *Escherichia coli* related to the effects of K antigens on their resistance to phagocytosis and killing by complement. *Immunology,* **20,** 767.

HPA (2010). Polymicrobial bacteraemia and fungaemia in England, Wales, and Northern Ireland: 2010 [Online]. HPA. Available: http://webarchive.nationalarchives.gov.uk/20140714084352/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317132423063 [Accessed 22-12-2014 2014].

HUNSTAD, D. A., JUSTICE, S. S., HUNG, C. S., LAUER, S. R. & HULTGREN, S. J. (2005). Suppression of bladder epithelial cytokine responses by uropathogenic *Escherichia coli*. *Infection and Immunity,* **73,** 3999-4006.

IYODA, S., TAMURA, K., ITOH, K., IZUMIYA, H., UENO, N., NAGATA, K., TOGO, M., TERAJIMA, J. & WATANABE, H. (2000). Inducible stx2 phages are lysogenized in the enteroaggregative and other phenotypic *Escherichia coli* O86: HNM isolated from patients. *FEMS Microbiology Letters,* **191,** 7-10.

JACOBSON, S. H., ÖSTENSON, C. G., TULLUS, K. & BRAUNER, A. (1992). Serum resistance in *Escherichia coli* strains causing acute pyelonephritis and bacteraemia. *acta pathologica, microbiologica, et immunologica Scandinavica,* **100,** 147-153.

JOHNSON, J. R. (1991). Virulence factors in *Escherichia coli* urinary tract infection. *Clinical Microbiology Reviews,* **4,** 80.

JOHNSON, J. R., ANDERSON, J. T., CLABOTS, C., JOHNSTON, B. & COOPERSTOCK, M. (2010a). Within-household sharing of a fluoroquinolone-resistant *Escherichia coli* sequence type ST131 strain causing pediatric osteoarticular infection. *The Pediatric Infectious Disease Journal,* **29,** 473-475.

JOHNSON, J. R., DELAVARI, P., KUSKOWSKI, M. & STELL, A. L. (2001a). Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *Journal of Infectious Diseases,* **183,** 78-88.

JOHNSON, J. R., JELACIC, S., SCHOENING, L. M., CLABOTS, C., SHAIKH, N., MOBLEY, H. L. & TARR, P. I. (2005). The IrgA homologue adhesin Iha is an *Escherichia coli* virulence factor in murine urinary tract infection. *Infection and Immunity,* **73,** 965-971.

JOHNSON, J. R., JOHNSTON, B., CLABOTS, C., KUSKOWSKI, M. A. & CASTANHEIRA, M. (2010b). *Escherichia coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. *Clinical Infectious Diseases,* **51,** 286-294.

JOHNSON, J. R., JOHNSTON, B., CLABOTS, C., KUSKOWSKI, M. A., PENDYALA, S., DEBROY, C., NOWICKI, B. & RICE, J. (2010c). *Escherichia coli* sequence type ST131 as an emerging fluoroquinolone-resistant uropathogen among renal transplant recipients. *Antimicrobial Agents and Chemotherapy,* **54,** 546-50.

JOHNSON, J. R., JOHNSTON, B., KUSKOWSKI, M. A., NOUGAYREDE, J.-P. & OSWALD, E. (2008). Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. *Journal of Clinical Microbiology,* **46,** 3906-3911.

JOHNSON, J. R. & KUSKOWSKI, M. (2000). Clonal origin, virulence factors, and virulence. *Infection and Immunity,* **68,** 424-425.

JOHNSON, J. R., MENARD, M., JOHNSTON, B., KUSKOWSKI, M. A., NICHOL, K. & ZHANEL, G. G. (2009a). Epidemic clonal groups of *Escherichia coli* as a cause of antimicrobial-resistant urinary tract infections in Canada (2002-2004). *Antimicrobial Agents and Chemotherapy,* **53,** 2733-2739.

JOHNSON, J. R., MILLER, S., JOHNSTON, B., CLABOTS, C. & DEBROY, C. (2009b). Sharing of *Escherichia coli* sequence type ST131 and other multidrug-resistant and urovirulent *E. coli* strains among dogs and cats within a household. *Journal of Clinical Microbiology,* **47,** 3721-3725.

JOHNSON, J. R. & O'BRYAN, T. T. (2004). Detection of the *Escherichia coli* group 2 polysaccharide capsule synthesis gene *kpsM* by a rapid and specific PCR-based assay. *Journal of Clinical Microbiology,* **42,** 1773-1776.

JOHNSON, J. R., O'BRYAN, T. T., DELAVARI, P., KUSKOWSKI, M., STAPLETON, A., CARLINO, U. & RUSSO, T. A. (2001b). Clonal relationships and extended virulence genotypes among *Escherichia coli* isolates from women with a first or recurrent episode of cystitis. *Journal of Infectious Diseases,* **183,** 1508-1517.

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

JOHNSON, J. R. & RUSSO, T. A. (2002b). Uropathogenic *Escherichia coli* as agents of diverse non–urinary tract extraintestinal infections. *Journal of Infectious Diseases,* **186,** 859-864.

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

JOHNSON, J. R., RUSSO, T. A., TARR, P. I., CARLINO, U., BILGE, S. S., VARY, J. C. & STELL, A. L. (2000). Molecular epidemiological and phylogenetic associations of two novel putative virulence genes, *iha* and *iroNE*. coli, among *Escherichia coli* isolates from patients with urosepsis. *Infection and Immunity,* **68,** 3040-3047.

JOHNSON, J. R., TCHESNOKOVA, V., JOHNSTON, B., CLABOTS, C., ROBERTS, P. L., BILLIG, M., RIDDELL, K., ROGERS, P., QIN, X. & BUTLER-WU, S. (2013). Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *Journal of Infectious Diseases,* **207,** 919-928.

JOHNSON, J. R., URBAN, C., WEISSMAN, S. J., JORGENSEN, J. H., LEWIS, J. S., HANSEN, G., EDELSTEIN, P., ROBICSEK, A., CLEARY, T. & ADACHI, J. (2012). Molecular epidemiological analysis of *Escherichia coli* sequence type ST131 (O25: H4) and *blaCTX-M-15* among extended-spectrum β-Lactamase-producing *E. coli* from the United States (2000-2009). *Antimicrobial Agents and Chemotherapy***,** AAC. 05824-11.

JOHNSON, T. J., KARIYAWASAM, S., WANNEMUEHLER, Y., MANGIAMELE, P., JOHNSON, S. J., DOETKOTT, C., SKYBERG, J. A., LYNNE, A. M., JOHNSON, J. R. & 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,** 3228-3236.

JONES, G. L., WARREN, R., SKIDMORE, S., DAVIES, V., GIBREEL, T. & UPTON, M. (2008). Prevalence and distribution of plasmid-mediated quinolone resistance genes in clinical isolates of *Escherichia coli* lacking extended-spectrum -lactamases. *Journal of Antimicrobial Chemotherapy,* **62,** 1245-51.

KAPER, J. B., NATARO, J. P. & MOBLEY, H. L. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology,* **2,** 123-140.

KARISIK, E., ELLINGTON, M., LIVERMORE, D. & WOODFORD, N. (2008). Virulence factors in *Escherichia coli* with CTX-M-15 and other extended-spectrum β-lactamases in the UK. *Journal of Antimicrobial Chemotherapy,* **61,** 54-58.

KATOULI, M. (2010). Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. *Iranian Journal of Microbiology,* **2,** 59.

KATOULI, M., KÜHN, I. & MÖLLBY, R. (1990). Evaluation of the stability of biochemical phenotypes of *Escherichia coli* upon subculturing and storage. *Journal of General Microbiology,* **136,** 1681-1688.

KIROV, S. M. (2003). Bacteria that express lateral flagella enable dissection of the multifunctional roles of flagella in pathogenesis. *FEMS Microbiology Letters,* **224,** 151-159.

KLEMM, P., JØRGENSEN, B. J., VAN DIE, I., DE REE, H. & BERGMANS, H. (1985). The *fim* genes responsible for synthesis of type 1 fimbriae in *Escherichia coli*, cloning and genetic organization. *Molecular and General Genetics,* **199,** 410-414.

KLEMM, P. & SCHEMBRI, M. A. (2000). Bacterial adhesins: function and structure. *International Journal of Medical Microbiology,* **290,** 27-35.

KLUMPP, S. & HWA, T. (2014). Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Current Opinion in Biotechnology,* **28,** 96-102.

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

KORHONEN, T., VÄISÄNEN, V., SAXÉN, H., HULTBERG, H. & SVENSON, S. (1982). P-antigen-recognizing fimbriae from human uropathogenic *Escherichia coli* strains. *Infection and Immunity,* **37,** 286-291.

KRAMER, A., SCHWEBKE, I. & KAMPF, G. (2006). How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases,* **6,** 130.

KROGFELT, K. A., BERGMANS, H. & KLEMM, P. (1990). Direct evidence that the FimH protein is the mannose-specific adhesin of *Escherichia coli* type 1 fimbriae. *Infection and Immunity,* **58,** 1995-1998.

KUDINHA, T., JOHNSON, J. R., ANDREW, S. D., KONG, F., ANDERSON, P. & GILBERT, G. L. (2013). *Escherichia coli* sequence type 131 (ST131) as a prominent cause of antibiotic resistance among urinary *Escherichia coli* isolates from reproductive-age women. *Journal of Clinical Microbiology***,** JCM. 01315-13.

KUHNERT, P., BOERLIN, P. & FREY, J. (2006). Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. *FEMS Microbiology Reviews,* **24,** 107-117.

KUHNERT, P., NICOLET, J. & FREY, J. (1995). Rapid and accurate identification of *Escherichia coli* K-12 strains. *Applied and Environmental Microbiology,* **61,** 4135-4139.

KUMAZAWA, J. & MATSUMOTO, T. (1997). Complicated urinary tract infections. *In:* T, B. (ed.). Basel: Karger.

KUNIN, C. M. (1962). Antibody distribution against nonenteropathic *E. coli*: relation to age, sex, and breast feeding. *Archives of Internal Medicine,* **110,** 676-686.

KUSECEK, B., WLOCH, H., MERCER, A., VAISÄNEN, V., PLUSCHKE, G., KORHONEN, T. & ACHTMAN, M. (1984). Lipopolysaccharide, capsule, and fimbriae as virulence factors among O1, O7, O16, O18, or O75 and K1, K5, or K100 *Escherichia coli*. *Infection and Immunity,* **43,** 368-379.

LANGRIDGE, G. C., PHAN, M.-D., TURNER, D. J., PERKINS, T. T., PARTS, L., HAASE, J., CHARLES, I., MASKELL, D. J., PETERS, S. E. & DOUGAN, G. (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Research,* **19,** 2308-2316.

LAPIERRE, P. & GOGARTEN, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in Genetics,* **25,** 107-110.

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. (2008a). UK epidemic *Escherichia coli* strains A–E, with CTX-M-15 β-lactamase, all belong to the international O25: H4-ST131 clone. *Journal of Antimicrobial Chemotherapy,* **62,** 1241-1244.

LAU, S. H., REDDY, S., CHEESBROUGH, J., BOLTON, F. J., WILLSHAW, G., CHEASTY, T., FOX, A. J. & UPTON, M. (2008b). Major uropathogenic *Escherichia coli* strain isolated in the northwest of England identified by multilocus sequence typing. *Journal of Clinical Microbiology,* **46,** 1076-80.

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

LAVIGNE, J.-P., VERGUNST, A. C., GORET, L., SOTTO, A., COMBESCURE, C., BLANCO, J., O'CALLAGHAN, D. & NICOLAS-CHANOINE, M.-H. (2012). Virulence potential and genomic mapping of the worldwide clone *Escherichia coli* ST131. *PLoS One,* **7,** e34294.

LAY JR, J. O., LIYANAGE, R., BORGMANN, S. & WILKINS, C. L. (2006). Problems with the "omics". *Trends in Analytical Chemistry,* **25,** 1046-1056.

LE BOUGUÉNEC, C. & SCHOULER, C. (2011). Sugar metabolism, an additional virulence factor in enterobacteria. *International Journal of Medical Microbiology,* **301,** 1-6.

LEE, M. Y., CHOI, H. J., CHOI, J. Y., SONG, M., SONG, Y., KIM, S. W., CHANG, H. H., JUNG, S. I., KIM, Y. S. & KI, H. K. (2010). Dissemination of ST131 and ST393 community-onset, ciprofloxacin-resistant *Escherichia coli* clones causing urinary tract infections in Korea. *Journal of Infection,* **60,** 146-153.

LEFLON-GUIBOUT, V., BLANCO, J., AMAQDOUF, K., MORA, A., GUIZE, L. & NICOLAS-CHANOINE, M. H. (2008). Absence of CTX-M enzymes but high prevalence of clones, including clone ST131, among fecal *Escherichia coli* isolates from healthy subjects living in the area of Paris, France. *Journal of Clinical Microbiology,* **46,** 3900-3905.

LESAGE, A. (1897). Contribution a l'étude des entérites infantiles—sérodiagnostic des races de *Bacterium coli*. *CR Seances Soc. Biol. Fil,* **49,** 900-901.

LEWIS, J. P. (2010). Metal uptake in host–pathogen interactions: role of iron in *Porphyromonas gingivalis* interactions with host organisms. *Periodontology 2000,* **52,** 94-116.

LEYING, H., SUERBAUM, S., KROLL, H., STAHL, D. & OPFERKUCH, W. (1990). The capsular polysaccharide is a major determinant of serum resistance in K-1-positive blood culture isolates of *Escherichia coli*. *Infection and Immunity,* **58,** 222-227.

LITWIN, C. M. & CALDERWOOD, S. (1993). Role of iron in regulation of virulence genes. *Clinical Microbiology Reviews,* **6,** 137-149.

LITWIN, M. S., SAIGAL, C. S., YANO, E. M., AVILA, C., GESCHWIND, S. A., HANLEY, J. M., JOYCE, G. F., MADISON, R., PACE, J. & POLICH, S. M. (2005). Urologic diseases in America Project: analytical methods and principal findings. *The Journal of Urology,* **173,** 933-937.

LIU, R. & OCHMAN, H. (2007). Origins of flagellar gene operons and secondary flagellar systems. *Journal of Bacteriology,* **189,** 7098-7104.

LOCKHART, W. R. & LISTON, J. (**1970)**. Methods for Numerical Taxonomy, Bethesda, American Society for Microbiology.

LÓPEZ-CERERO, L., NAVARRO, M. D., BELLIDO, M., MARTÍN-PEÑA, A., VIÑAS, L., CISNEROS, J. M., GÓMEZ-LANGLEY, S. L., SÁNCHEZ-MONTESEIRÍN, H., MORALES, I. & PASCUAL, A. (2013). *Escherichia coli* belonging to the worldwide emerging epidemic clonal group O25b/ST131: risk factors and clinical implications. *Journal of Antimicrobial Chemotherapy***,** dkt405.

MADIGAN, M. T., MARTINKO, J. M., DUNLAP, P. V. & CLARK, D. P. (**2009)**. Brock Biology of Microorganisms, San Francisco, Pearson Benjamin Cummings.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K. & CAUGANT, D. A. (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,** 3140-3145.

MANGES, A. R., TABOR, H., TELLIS, P., VINCENT, C. & TELLIER, P.-P. (2008). Endemic and epidemic lineages of *Escherichia coli* that cause urinary tract infections. *Emerging Infectious Diseases,* **14,** 1575.

MARTINEZ, J., CERCENADO, E., PÉREZ-DÍAZ, J. & BAQUERO, F. (1986). Multifactorial determination of systemic invasivity in *Escherichia coli*. *FEMS Microbiology Letters,* **37,** 259-261.

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

MARTINEZ-MEDINA, M., MORA, A., BLANCO, M., LÓPEZ, C., ALONSO, M. P., BONACORSI, S., NICOLAS-CHANOINE, M.-H., DARFEUILLE-MICHAUD, A., GARCIA-GIL, J. & BLANCO, J. (2009). Similarity and divergence among adherent-invasive *Escherichia coli* and extraintestinal pathogenic *E. coli* strains. *Journal of Clinical Microbiology,* **47,** 3968-3979.

MATSUMURA, Y., YAMAMOTO, M., NAGAO, M., HOTTA, G., MATSUSHIMA, A., ITO, Y., TAKAKURA, S. & 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,** 2612-2620.

MCCABE, W., KAIJSER, B., OLLING, S., UWAYDAH, M. & HANSON, L. (1978). *Escherichia coli* in bacteremia: K and O antigens and serum sensitivity of strains from adults and neonates. *Journal of Infectious Diseases,* **138,** 33-41.

MCCORMICK, A., FLEMING, D. & CHARLTON, J. (**1995)**. Morbidity Statistics from General Practice: Fourth National Study 1991-1992*,* London, HMSO.

MCNALLY, A., ALHASHASH, F., COLLINS, M., ALQASIM, A., PASZCKIEWICZ, K., WESTON, V. & DIGGLE, M. (2013a). Genomic analysis of extra‐intestinal pathogenic *Escherichia coli* urosepsis. *Clinical Microbiology and Infection,* **19,** E328-E334.

MCNALLY, A., CHENG, L., HARRIS, S. R. & CORANDER, J. (2013b). 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,** 699-710.

MCNALLY, A., DALTON, T., LA RAGIONE, R. M., STAPLETON, K., MANNING, G. & NEWELL, D. G. (2006). *Yersinia enterocolitica* isolates of differing biotypes from humans and animals are adherent, invasive and persist in macrophages, but differ in cytokine secretion profiles *in vitro*. *Journal of Medical Microbiology,* **55,** 1725-1734.

MEDINI, D., DONATI, C., TETTELIN, H., MASIGNANI, V. & RAPPUOLI, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development,* **15,** 589-594.

MELEKOS, M. D. & NABER, K. G. (2000). Complicated urinary tract infections. *International Journal of Antimicrobial Agents,* **15,** 247.

MIAJLOVIC, H., COOKE, N. M., MORAN, G. P., ROGERS, T. R. & 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,** 298-305.

MIAJLOVIC, H. & SMITH, S. G. (2014). Bacterial self‐defence: how *Escherichia coli* evades serum killing. *FEMS Microbiology Letters,* **354,** 1-9.

MILES, A., MISRA, S. & IRWIN, J. (1938). The estimation of the bactericidal power of the blood. *Journal of Hygiene,* **38,** 732-749.

MILKMAN, R. & CRAWFORD, I. P. (1983). Clustered third-base substitutions among wild strains of *Escherichia coli*. *Science,* **221,** 378-380.

MILLS, M., MEYSICK, K. C. & 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,** 5869-5880.

MIRA, A., MARTÍN-CUADRADO, A. B., D'AURIA, G. & RODRÍGUEZ-VALERA, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology,* **13,** 45-57.

MOBLEY, H., CHIPPENDALE, G., TENNEY, J., HULL, R. & WARREN, J. (1987). Expression of type 1 fimbriae may be required for persistence of *Escherichia coli* in the catheterized urinary tract. *Journal of Clinical Microbiology,* **25,** 2253-2257.

MOLL, A., MANNING, P. & TIMMIS, K. (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,** 359-367.

MONTENEGRO, M., BITTER-SUERMANN, D., TIMMIS, J., AGÜERO, M., CABELLO, F., SANYAL, S. & TIMMIS, K. (1985). *traT* gene sequences, serum resistance and pathogenicity-related factors in clinical isolates of *Escherichia coli* and other gram-negative bacteria. *Microbiology,* **131,** 1511-1521.

MORA, A., BLANCO, M., LÓPEZ, C., MAMANI, R., BLANCO, J. E., ALONSO, M. P., GARCÍA-GARROTE, F., DAHBI, G., HERRERA, A. & FERNÁNDEZ, A. (2011). Emergence of clonal groups O1: HNM-D-ST59, O15: H1-D-ST393, O20: H34/HNM-D-ST354, O25b: H4-B2-ST131 and ONT: H21, 42-B1-ST101 among CTX-M-14-producing *Escherichia coli* clinical isolates in Galicia, northwest Spain. *International Journal of Antimicrobial Agents,* **37,** 16-21.

MORABITO, S., KARCH, H., MARIANI-KURKDJIAN, P., SCHMIDT, H., MINELLI, F., BINGEN, E. & CAPRIOLI, A. (1998). *Enteroaggregative, Shiga toxin-producing Escherichia coli* O111: H2 *associated with an outbreak of hemolytic-uremic syndrome*. *Journal of Clinical Microbiology,* **36,** 840-842.

MOSSMAN, K. L., MIAN, M. F., LAUZON, N. M., GYLES, C. L., LICHTY, B., MACKENZIE, R., GILL, N. & ASHKAR, A. A. (2008). Cutting edge: FimH adhesin of type 1 fimbriae is a novel TLR4 ligand. *The Journal of Immunology,* **181,** 6702-6706.

MU, X.-Q. & BULLITT, E. (2006). Structure and assembly of P-pili: a protruding hinge region used for assembly of a bacterial adhesion filament. *Proceedings of the National Academy of Sciences,* **103,** 9861-9866.

MULVEY, M. A. (2002). Adhesion and entry of uropathogenic *Escherichia coli*. *Cellular Microbiology,* **4,** 257-271.

MULVEY, M. A., LOPEZ-BOADO, Y. S., WILSON, C. L., ROTH, R., PARKS, W. C., HEUSER, J. & HULTGREN, S. J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science,* **282,** 1494-1497.

MURPHY, S. L. (2000). Deaths: final data for 1998. Journal, 48, 11.

MYLOTTE, J. M., TAYARA, A. & GOODNOUGH, S. (2002). Epidemiology of bloodstream infection in nursing home residents: evaluation in a large cohort from multiple homes. *Clinical Infectious Diseases,* **35,** 1484-1490.

NATARO, J. P. & KAPER, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews,* **11,** 142-201.

NÈGRE, V. L., BONACORSI, S., SCHUBERT, S., BIDET, P., NASSIF, X. & BINGEN, E. (2004). The siderophore receptor IroN, but not the high-pathogenicity island or the hemin receptor ChuA, contributes to the bacteremic step of *Escherichia coli* neonatal meningitis. *Infection and Immunity,* **72,** 1216-1220.

NEIDHARDT, F. C. (1999). Bacterial growth: constant obsession with dN/dt. *Journal of Bacteriology,* **181,** 7405-7408.

NEILANDS, J. Siderophores: diverse roles in microbial and human physiology. Ciba Foundation Symposium 51-Iron Metabolism, 1976. Wiley Online Library, 107-124.

NICOLAS-CHANOINE, M.-H., BERTRAND, X. & MADEC, J.-Y. (2014). *Escherichia coli* ST131, an intriguing clonal group. *Clinical Microbiology Reviews,* **27,** 543-574.

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

NIELUBOWICZ, G. R. & MOBLEY, H. L. (2010). Host–pathogen interactions in urinary tract infection. *Nature Reviews Urology,* **7,** 430-441.

OLESEN, B., HANSEN, D. S., NILSSON, F., FRIMODT-MØLLER, J., LEIHOF, R. F., STRUVE, C., SCHEUTZ, F., JOHNSTON, B., KROGFELT, K. A. & 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,** 1779-1785.

OPAL, S., CROSS, A. & GEMSKI, P. (1982). K antigen and serum sensitivity of rough *Escherichia coli*. *Infection and Immunity,* **37,** 956-960.

ORSKOV, F. & ORSKOV, I. (1983). Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the *Enterobacteriaceae* and other bacteria. *The Journal of Infectious Diseases.*

ORSKOV, F. & ORSKOV, I. (1984). Serotyping of *Escherichia coli*. *In:* T, I. B. (ed.) *Methods in microbiology.* London: Academic Press.

ORSKOV, F. & ORSKOV, I. (1992). *Escherichia coli* serotyping and disease in man and animals. *Canadian Journal of Microbiology,* **38,** 699-704.

ORSKOV, I., ORSKOV, F., JANN, B. & JANN, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriological Reviews,* **41,** 667.

OTEO, J., DIESTRA, K., JUAN, C., BAUTISTA, V., NOVAIS, Â., PÉREZ-VÁZQUEZ, M., MOYÁ, B., MIRÓ, E., COQUE, T. M. & OLIVER, A. (2009). Extended-spectrum [beta]-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,** 173-176.

OWENS, R. C., JOHNSON, J. R., STOGSDILL, P., YARMUS, L., LOLANS, K. & QUINN, J. (2011). Community transmission in the United States of a CTX-M-15-producing sequence type ST131 *Escherichia coli* strain resulting in death. *Journal of Clinical Microbiology,* **49,** 3406-3408.

PACHECO, A. R., CURTIS, M. M., RITCHIE, J. M., MUNERA, D., WALDOR, M. K., MOREIRA, C. G. & SPERANDIO, V. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature,* **492,** 113-117.

PANCHOLI, V. & CHHATWAL, G. S. (2003). Housekeeping enzymes as virulence factors for pathogens. *International Journal of Medical Microbiology,* **293,** 391-401.

PARTHASARATHY, G., YAO, Y. & KIM, K. (2007). Flagella promote *Escherichia coli* K1 association with and invasion of human brain microvascular endothelial cells. *Infection and Immunity,* **75,** 2937-2945.

PAUL, S., LINARDOPOULOU, E. V., BILLIG, M., TCHESNOKOVA, V., PRICE, L. B., JOHNSON, J. R., CHATTOPADHYAY, S. & SOKURENKO, E. V. (2013). Role of homologous recombination in adaptive diversification of extraintestinal *Escherichia coli*. *Journal of Bacteriology,* **195,** 231-242.

PEIGNE, C., BIDET, P., MAHJOUB-MESSAI, F., PLAINVERT, C., BARBE, V., MÉDIGUE, C., FRAPY, E., NASSIF, X., DENAMUR, E. & BINGEN, E. (2009). The plasmid of *Escherichia coli* strain S88 (O45: K1: H7) that causes neonatal meningitis is closely related to avian pathogenic *E. coli* plasmids and is associated with high-level bacteremia in a neonatal rat meningitis model. *Infection and Immunity,* **77,** 2272-2284.

PEIRANO, G., MULVEY, G. L., ARMSTRONG, G. D. & PITOUT, J. D. (2013). Virulence potential and adherence properties of *Escherichia coli* that produce CTX-M and NDM β-lactamases. *Journal of Medical Microbiology,* **62,** 525-530.

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

PELEG, A., SHIFRIN, Y., ILAN, O., NADLER-YONA, C., NOV, S., KOBY, S., BARUCH, K., ALTUVIA, S., ELGRABLY-WEISS, M. & ABE, C. M. (2005). Identification of an *Escherichia coli* operon required for formation of the O-antigen capsule. *Journal of Bacteriology,* **187,** 5259-5266.

PETTY, N. K., BULGIN, R., CREPIN, V. F., CERDEÑO-TÁRRAGA, A. M., SCHROEDER, G. N., QUAIL, M. A., LENNARD, N., CORTON, C., BARRON, A. & CLARK, L. (2010). The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *Journal of Bacteriology,* **192,** 525-538.

PETTY, N. K., ZAKOUR, N. L. B., STANTON-COOK, M., SKIPPINGTON, E., TOTSIKA, M., FORDE, B. M., PHAN, M.-D., MORIEL, D. G., PETERS, K. M. & DAVIES, M. (2014). Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proceedings of the National Academy of Sciences,* **111,** 5694-5699.

PHAN, M.-D., PETERS, K. M., SARKAR, S., LUKOWSKI, S. W., ALLSOPP, L. P., MORIEL, D. G., ACHARD, M. E., TOTSIKA, M., MARSHALL, V. M. & UPTON, M. (2013). The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. *PLoS Genetics,* **9,** e1003834.

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

PITOUT, J. D. (2012a). Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Frontiers in Microbiology,* **3,** 1-7.

PITOUT, J. D. (2012b). Extraintestinal pathogenic *Escherichia coli*: an update on antimicrobial resistance, laboratory diagnosis and treatment. *Expert Review of Anti-infective Therapy,* **10,** 1165-1176.

PLATELL, J. L., COBBOLD, R. N., JOHNSON, J. R., HEISIG, A., HEISIG, P., CLABOTS, C., KUSKOWSKI, M. A. & TROTT, D. J. (2011). Commonality among fluoroquinolone-resistant sequence type ST131 extraintestinal *Escherichia coli* isolates from humans and companion animals in Australia. *Antimicrobial Agents and Chemotherapy,* **55,** 3782-3787.

PRICE, L. B., JOHNSON, J. R., AZIZ, M., CLABOTS, C., JOHNSTON, B., TCHESNOKOVA, V., NORDSTROM, L., BILLIG, M., CHATTOPADHYAY, S. & STEGGER, M. (2013). The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 Is driven by a single highly pathogenic subclone, H30-Rx. *mBio,* **4,** e00377-13.

PUPO, G. M., KARAOLIS, D., LAN, R. & REEVES, P. R. (1997). Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infection and Immunity* **65,** 2685-2692.

RASKO, D. A., ROSOVITZ, M., MYERS, G. S., MONGODIN, E. F., FRICKE, W. F., GAJER, P., CRABTREE, J., SEBAIHIA, M., THOMSON, N. R. & CHAUDHURI, R. (2008). The pangenome structure of *Escherichia coli:* comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology,* **190,** 6881-6893.

RATNAM, S., MARCH, S. B., AHMED, R., BEZANSON, G. & KASATIYA, S. (1988). Characterization of *Escherichia coli* serotype O157: H7. *Journal of Clinical Microbiology,* **26,** 2006-2012.

REIGSTAD, C. S., HULTGREN, S. J. & GORDON, J. I. (2007). Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled. *Journal of Biological Chemistry,* **282,** 21259-21267.

REN, C.-P., BEATSON, S. A., PARKHILL, J. & PALLEN, M. J. (2005). The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel

flagellar system from *Escherichia coli*. *Journal of Bacteriology,* **187,** 1430-1440.

REUTER, S. (2011). Evolutionary Dynamics of the *Yersinia enterocolitica* complex. PhD, Nottingham Trent University.

RHEN, M., KLEMM, P. & KORHONEN, T. (1986). Identification of two new hemagglutinins of *Escherichia coli*, N-acetyl-D-glucosamine-specific fimbriae and a blood group M-specific agglutinin, by cloning the corresponding genes in *Escherichia coli* K-12. *Journal of Bacteriology,* **168,** 1234-1242.

RICHARD, H. T. & FOSTER, J. W. (2003). Acid resistance in *Escherichia coli*. *Advances in Applied Microbiology,* **52,** 167-186.

RICHARDSON, M. (2009). Principal component analysis. *Mathematical Modelling and Scientific Computing, University of Oxford, Oxford, UK*.

ROBSON, W. P. & DANIELS, R. (2008). The Sepsis Six: helping patients to survive sepsis. *British Journal of Nursing,* **17,** 16-21.

ROESCH, P. L., REDFORD, P., BATCHELET, S., MORITZ, R. L., PELLETT, S., HAUGEN, B. J., BLATTNER, F. R. & WELCH, R. A. (2003). Uropathogenic *Escherichia coli* use d - serine deaminase to modulate infection of the murine urinary tract. *Molecular Microbiology,* **49,** 55-67.

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

ROHMER, L., HOCQUET, D. & MILLER, S. I. (2011). Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends in Microbiology,* **19,** 341-348.

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

ROSEN, D. A., HOOTON, T. M., STAMM, W. E., HUMPHREY, P. A. & HULTGREN, S. J. (2007). Detection of intracellular bacterial communities in human urinary tract infection. *PLoS Medicine,* **4,** e329.

RUSSO, T., BROWN, J. J., JODUSH, S. T. & JOHNSON, J. R. (1996). The O4 specific antigen moiety of lipopolysaccharide but not the K54 group 2 capsule is important for urovirulence of an extraintestinal isolate of *Escherichia coli*. *Infection and Immunity,* **64,** 2343-2348.

RUSSO, T. A., DAVIDSON, B. A., GENAGON, S. A., WARHOLIC, N. M., MACDONALD, U., PAWLICKI, P. D., BEANAN, J. M., OLSON, R., HOLM, B. A. & KNIGHT III, P. R. (2005). *E. coli* virulence factor hemolysin induces neutrophil apoptosis and necrosis/lysis in vitro and necrosis/lysis and lung injury in a rat pneumonia model. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* **289,** L207-L216.

RUSSO, T. A. & JOHNSON, J. R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli:* ExPEC. *Journal of Infectious Diseases,* **181,** 1753-1754.

RUSSO, T. A. & 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,** 449-456.

RUSSO, T. A., MCFADDEN, C. D., CARLINO-MACDONALD, U. B., BEANAN, J. M., BARNARD, T. J. & JOHNSON, J. R. (2002). IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. *Infection and Immunity,* **70,** 7156-7160.

SAARELA, S., TAIRA, S., NURMIAHO-LASSILA, E.-L., MAKKONEN, A. & RHEN, M. (1995). The *Escherichia coli* G-fimbrial lectin protein participates both in fimbrial biogenesis and in recognition of the receptor N-acetyl-D-glucosamine. *Journal of Bacteriology,* **177,** 1477-1484.

SAHL, J. W., CAPORASO, G., RASKO, D. A. & KEIM, P. S. (2014). The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ,* **332,** 1-12.

SAHL, J. W., JOHNSON, J. K., HARRIS, A. D., PHILLIPPY, A. M., HSIAO, W. W., THOM, K. A. & RASKO, D. A. (2011). Genomic comparison of multi-drug resistant invasive and colonizing *Acinetobacter baumannii* isolated from diverse human body sites reveals genomic plasticity. *BMC Genomics,* **12,** 291.

SAINI, L. S., GALSWORTHY, S. B., JOHN, M. A. & VALVANO, M. A. (1999). Intracellular survival of *Burkholderia* cepacia complex isolates in the presence of macrophage cell activation. *Microbiology,* **145,** 3465-3475.

SAKARYA, S., ERTEM, G. T., ONCU, S., KOCAK, I., EROL, N. & ONCU, S. (2003). *Escherichia coli* bind to urinary bladder epithelium through nonspecific sialic acid mediated adherence. *FEMS Immunology & Medical Microbiology,* **39,** 45-50.

SANSANO, M., REYNARD, A. & CUNNINGHAM, R. (1985). Inhibition of serum bactericidal reaction by lipopolysaccharide. *Infection and Immunity,* **48,** 759-762.

SARKAR, S., ULETT, G. C., TOTSIKA, M., PHAN, M.-D. & SCHEMBRI, M. A. (2014). Role of Capsule and O Antigen in the Virulence of Uropathogenic *Escherichia coli*. *PloS One,* **9,** e94786.

SCHAECHTER, M. (2010). From growth physiology to systems biology. *International Microbiology,* **9,** 157-161.

SCHILLING, J. D., MARTIN, S. M., HUNG, C. S., LORENZ, R. G. & HULTGREN, S. J. (2003). Toll-like receptor 4 on stromal and hematopoietic cells mediates innate resistance to uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences,* **100,** 4203-4208.

SCHLOSS, P. D., WESTCOTT, S. L., RYABIN, T., HALL, J. R., HARTMANN, M., HOLLISTER, E. B., LESNIEWSKI, R. A., OAKLEY, B. B., PARKS, D. H. & ROBINSON, C. J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology,* **75,** 7537-7541.

SCHNEIDER, G., DOBRINDT, U., BRÜGGEMANN, H., NAGY, G., JANKE, B., BLUM-OEHLER, G., BUCHRIESER, C., GOTTSCHALK, G., EMÖDY, L. & HACKER, J. (2004). The pathogenicity island-associated K15 capsule determinant exhibits a novel genetic structure and correlates with virulence in uropathogenic *Escherichia coli* strain 536. *Infection and Immunity,* **72,** 5993-6001.

SEEMANN, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* **30,** 1-2.

SELANDER, R., CAUGANT, D. & WHITTAM, T. (1987). Genetic structure and variation in natural populations of *Escherichia coli. Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology,* **2,** 1625-1648.

SELANDER, R. K. & LEVIN, B. R. (1980). Genetic diversity and structure in *Escherichia coli* populations. *Science,* **210,** 545-547.

SHEHATA, T. E. & MARR, A. G. (1971). Effect of nutrient concentration on the growth of *Escherichia coli*. *Journal of Bacteriology,* **107,** 210-216.

SHEPPARD, M., WEBB, C., HEATH, F., MALLOWS, V., EMILIANUS, R., MASKELL, D. & MASTROENI, P. (2003). Dynamics of bacterial growth and distribution within the liver during *Salmonella* infection. *Cellular Microbiology,* **5,** 593-600.

SHEPPARD, S. K., DIDELOT, X., MERIC, G., TORRALBO, A., JOLLEY, K. A., KELLY, D. J., BENTLEY, S. D., MAIDEN, M. C., PARKHILL, J. & FALUSH, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences,* **110,** 11923-11927.

SKAAR, E. P. (2010). The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS Pathogens,* **6,** e1000949.

SLAVCHEV, G., PISAREVA, E. & MARKOVA, N. (2013). Virulence of uropathogenic *Escherichia coli*. *Journal of Culture Collections,* **6,** 3-9.

SMITH, H. W. (1974). A search for transmissible pathogenic characters in invasive strains of *Escherichia coli:* the discovery of a plasmid-controlled toxin and a plasmid-controlled lethal character closely associated, or identical, with colicine V. *Journal of General Microbiology,* **83,** 95-111.

SMITH, J. M., SMITH, N. H., O'ROURKE, M. & SPRATT, B. G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences,* **90,** 4384-4388.

SMITH, Y. C., GRANDE, K. K., RASMUSSEN, S. B. & 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,** 750-757.

SNIPEN, L. & USSERY, D. W. (2010). Standard operating procedure for computing pangenome trees. *Standards in Genomic Sciences,* **2,** 135.

SNYDER, J. A., HAUGEN, B. J., BUCKLES, E. L., LOCKATELL, C. V., JOHNSON, D. E., DONNENBERG, M. S., WELCH, R. A. & MOBLEY, H. L. (2004). Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infection and Immunity,* **72,** 6373-6381.

SOLL, D., PUJOL, C., LOCKHART, S., MURRAY, P., BARON, E., JORGENSEN, J., LANDRY, M. & PFALLER, M. (2006). Laboratory procedures for the epidemiological analysis of microorganisms. *Manual of Clinical Microbiology* **1,** 129-151.

SRINIVASAN, U., FOXMAN, B. & MARRS, C. F. (2003). Identification of a gene encoding heat-resistant agglutinin in *Escherichia coli* as a putative virulence factor in urinary tract infection. *Journal of Clinical Microbiology,* **41,** 285-289.

STAMATAKIS, A., LUDWIG, T. & MEIER, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics,* **21,** 456-463.

STAMM, W. E. & HOOTON, T. M. (1993). Management of urinary tract infections in adults. *New England Journal of Medicine,* **329,** 1328-1334.

STAWSKI, G., NIELSEN, L., ØRSKOV, F. & ØRSKOV, I. (1990). Serum sensitivity of a diversity of *Escherichia coli* antigenic reference strains. *acta pathologica, microbiologica, et immunologica Scandinavica,* **98,** 828-838.

STENTEBJERG-OLESEN, B., CHAKRABORTY, T. & KLEMM, P. (1999). Type 1 Fimbriation and Phase Switching in a Natural *Escherichia coli fimB* Null Strain, Nissle 1917. *Journal of Bacteriology,* **181,** 7470-7478.

STENUTZ, R., WEINTRAUB, A. & WIDMALM, G. (2006). The structures of *Escherichia coli* O - polysaccharide antigens. *FEMS Microbiology Reviews,* **30,** 382-403.

STEWART, B. J. & MCCARTER, L. L. (2003). Lateral flagellar gene system of *Vibrio parahaemolyticus*. *Journal of Bacteriology,* **185,** 4508-4518.

STOLL, B. J., HANSEN, N., FANAROFF, A. A., WRIGHT, L. L., CARLO, W. A., EHRENKRANZ, R. A., LEMONS, J. A., DONOVAN, E. F., STARK, A. R. & TYSON, J. E. (2002). Changes in pathogens causing early-onset sepsis in very-low-birth-weight infants. *New England Journal of Medicine,* **347,** 240-247.

SUBASHCHANDRABOSE, S., HAZEN, T. H., BRUMBAUGH, A. R., HIMPSL, S. D., SMITH, S. N., ERNST, R. D., RASKO, D. A. & MOBLEY, H. L. (2014). Host-specific induction of *Escherichia coli* fitness genes during human urinary tract infection. *Proceedings of the National Academy of Sciences*, 201415959.

SUBASHCHANDRABOSE, S., SMITH, S. N., SPURBECK, R. R., KOLE, M. M. & MOBLEY, H. L. (2013). Genome-wide detection of fitness genes in

uropathogenic *Escherichia coli* during systemic infection. *PLoS Pathogens,* **9,** e1003788.

SULLIVAN, M. J., PETTY, N. K. & BEATSON, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics,* **27,** 1009-1010.

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

SVANBORG, C., BERGSTEN, G., FISCHER, H., FRENDÉUS, B., GODALY, G., GUSTAFSSON, E., HANG, L., HEDLUND, M., KARPMAN, D. & LUNDSTEDT, A.-C. (2001). The'innate'host response protects and damages the infected urinary tract. *Annals of Medicine,* **33,** 563-570.

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

TAL, S., GULLER, V., LEVI, S., BARDENSTEIN, R., BERGER, D., GUREVICH, I. & GUREVICH, A. (2005). Profile and prognosis of febrile elderly patients with bacteremic urinary tract infection. *Journal of Infection,* **50,** 296-305.

TALAN, D. A., STAMM, W. E., HOOTON, T. M., MORAN, G. J., BURKE, T., IRAVANI, A., REUNING-SCHERER, J. & CHURCH, D. A. (2000). Comparison of ciprofloxacin (7 days) and trimethoprim-sulfamethoxazole (14 days) for acute uncomplicated pyelonephritis in women. *The Journal of the American Medical Association,* **283,** 1583-1590.

TARTOF, S. Y., SOLBERG, O. D., MANGES, A. R. & RILEY, L. W. (2005). Analysis of a uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. *Journal of Clinical Microbiology,* **43,** 5860-5864.

TARTOF, S. Y., SOLBERG, O. D. & RILEY, L. W. (2007). Genotypic analyses of uropathogenic *Escherichia coli* based on fimH single nucleotide polymorphisms (SNPs). *Journal of Medical Microbiology,* **56,** 1363-1369.

TENAILLON, O., SKURNIK, D., PICARD, B. & DENAMUR, E. (2010). The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology,* **8,** 207-217.

TERASHIMA, H., KOJIMA, S. & HOMMA, M. (2008). Flagellar motility in bacteria: structure and function of flagellar motor. *International Review of Cell and Molecular Biology,* **270,** 39-85.

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. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America,* **102,** 13950-13955.

TOH, H., OSHIMA, K., TOYODA, A., OGURA, Y., OOKA, T., SASAMOTO, H., PARK, S.-H., IYODA, S., KUROKAWA, K. & MORITA, H. (2010). Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. *Journal of Bacteriology,* **192,** 1165-1166.

TOTSIKA, M., BEATSON, S. A., SARKAR, S., PHAN, M.-D., PETTY, N. K., BACHMANN, N., SZUBERT, M., SIDJABAT, H. E., PATERSON, D. L. & UPTON, M. (2011). Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One,* **6,** e26578.

TOTSIKA, M., GOMES MORIEL, D., IDRIS, A., A ROGERS, B., J WURPEL, D., PHAN, M.-D., L PATERSON, D. & A SCHEMBRI, M. (2012). Uropathogenic *Escherichia coli* mediated urinary tract infection. *Current Drug Targets,* **13,** 1386-1399.

TOUCHON, M., HOEDE, C., TENAILLON, O., BARBE, V., BAERISWYL, S., BIDET, P., BINGEN, E., BONACORSI, S., BOUCHIER, C. & BOUVET, O. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics,* **5,** e1000344.

TOWNSEND, S. M., HURRELL, E., GONZALEZ-GOMEZ, I., LOWE, J., FRYE, J. G., FORSYTHE, S. & BADGER, J. L. (2007). *Enterobacter sakazakii* invades brain capillary endothelial cells, persists in human macrophages influencing cytokine secretion and induces severe brain pathology in the neonatal rat. *Microbiology,* **153,** 3538-3547.

TOWNSEND, S. M., POLLACK, H. A., GONZALEZ-GOMEZ, I., SHIMADA, H. & BADGER, J. L. (2003). *Citrobacter koseri* brain abscess in the neonatal rat: survival and replication within human and rat macrophages. *Infection and Immunity,* **71,** 5871-5880.

TRABULSI, L. R., KELLER, R. & TARDELLI GOMES, T. (2002). Typical and atypical enteropathogenic *Escherichia coli*. *Emerging Infectious Diseases,* **8,** 508-513.

UCHIDA, Y., MOCHIMARU, T., MOROKUMA, Y., KIYOSUKE, M., FUJISE, M., ETO, F., ERIGUCHI, Y., NAGASAKI, Y., SHIMONO, N. & KANG, D. (2010). Clonal spread in Eastern Asia of ciprofloxacin-resistant *Escherichia coli* serogroup O25 strains, and associated virulence factors. *International Journal of Antimicrobial Agents,* **35,** 444-450.

ULETT, G. C., TOTSIKA, M., SCHAALE, K., CAREY, A. J., SWEET, M. J. & SCHEMBRI, M. A. (2013). Uropathogenic *Escherichia coli* virulence and innate immune responses during urinary tract infection. *Current Opinion in Microbiology,* **16,** 100-107.

VAAS, L. A., SIKORSKI, J., MICHAEL, V., GÖKER, M. & KLENK, H.-P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS One,* **7,** e34846.

VAN BELKUM, A., STRUELENS, M., DE VISSER, A., VERBRUGH, H. & TIBAYRENC, M. (2001). Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clinical Microbiology Reviews,* **14,** 547-560.

VAN BELKUM, A., TASSIOS, P., DIJKSHOORN, L., HAEGGMAN, S., COOKSON, B., FRY, N., FUSSING, V., GREEN, J., FEIL, E. & GERNER - SMIDT, P. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and Infection,* **13,** 1-46.

VAN DER BIJ, A. K., PEIRANO, G., PITONDO-SILVA, A. & PITOUT, J. D. (2012). The presence of genes encoding for different virulence factors in clonally related *Escherichia coli* that produce CTX-Ms. *Diagnostic Microbiology and Infectious Disease,* **72,** 297-302.

VIGIL, K. J., JOHNSON, J. R., JOHNSTON, B. D., KONTOYIANNIS, D. P., MULANOVICH, V. E., RAAD, I. I., DUPONT, H. L. & ADACHI, J. A. (2010). *Escherichia coli* pyomyositis: an emerging infectious disease among patients with hematologic malignancies. *Clinical Infectious Diseases,* **50,** 374-80.

VIMONT, S., BOYD, A., BLEIBTREU, A., BENS, M., GOUJON, J.-M., GARRY, L., CLERMONT, O., DENAMUR, E., ARLET, G. & VANDEWALLE, A. (2012). The CTX-M-15-producing *Escherichia coli* clone O25b: H4-ST131 has high intestine colonization and urinary tract infection abilities. *PloS One,* **7,** e46547.

WEINTRAUB, A. (2007). Enteroaggregative *Escherichia coli*: epidemiology, virulence and detection. *Journal of Medical Microbiology,* **56,** 4-8.

WEISER, J. N. & GOTSCHLICH, E. C. (1991). Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of *Escherichia coli* K-1. *Infection and Immunity,* **59,** 2252-2258.

WEISS, J., VICTOR, M., CROSS, A. S. & ELSBACH, P. (1982). Sensitivity of K1-encapsulated *Escherichia coli* to killing by the bactericidal/permeability-increasing protein of rabbit and human neutrophils. *Infection and Immunity,* **38,** 1149-1153.

WEISSMAN, S. J., CHATTOPADHYAY, S., APRIKIAN, P., OBATA - YASUOKA, M., YAROVA - YAROVAYA, Y., STAPLETON, A., BA - THEIN, W., DYKHUIZEN, D., JOHNSON, J. R. & SOKURENKO, E. V. (2006). Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Molecular Microbiology,* **59,** 975-988.

WEISSMAN, S. J., JOHNSON, J. R., TCHESNOKOVA, V., BILLIG, M., DYKHUIZEN, D., RIDDELL, K., ROGERS, P., QIN, X., BUTLER-WU, S. & COOKSON, B. T. (2012). High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. *Applied and Environmental Microbiology,* **78,** 1353-1360.

WELCH, R. (1991). Pore - forming cytolysins of gram - negative bacteria. *Molecular Microbiology,* **5,** 521-528.

WELCH, R., BURLAND, V., PLUNKETT, G., REDFORD, P., ROESCH, P., RASKO, D., BUCKLES, E., LIOU, S.-R., BOUTIN, A. & HACKETT, J. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences,* **99,** 17020-17024.

WHITFIELD, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annual Review of Biochemistry,* **75,** 39-68.

WHITFIELD, C. & ROBERTS, I. S. (1999). Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Molecular Microbiology,* **31,** 1307-1319.

WHITT, D. D. & SALYERS, A. A. (**2002)**. Bacterial pathogenesis: A molecular approach. USA: ASM Press.

WHITTAM, T. S., WOLFE, M. L., WACHSMUTH, I. K., ORSKOV, F., ORSKOV, I. & WILSON, R. (1993). Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infection and Immunity,* **61,** 1619-1629.

WILLIAMSON, D. A., FREEMAN, J. T., PORTER, S., ROBERTS, S., WILES, S., PATERSON, D. L. & JOHNSON, J. R. (2013). Clinical and molecular correlates of virulence in *Escherichia coli* causing bloodstream infection following transrectal ultrasound-guided (TRUS) prostate biopsy. *Journal of Antimicrobial Chemotherapy,* **68,** 2898-2906.

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

WOLD, A. E., CAUGANT, D. A., LIDIN-JANSON, G., DE MAN, P. & SVANBORG, C. (1992). Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics. *Journal of Infectious Diseases,* **165,** 46-52.

WOODFORD, N., CARATTOLI, A., KARISIK, E., UNDERWOOD, A., ELLINGTON, M. J. & LIVERMORE, D. M. (2009). Complete nucleotide sequences of plasmids *pEK204*, *pEK499* and *pEK516*, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25: H4-ST131 clone. *Antimicrobial Agents and Chemotherapy,* **53,** 4472-4482.

WU, X.-R., SUN, T.-T. & MEDINA, J. J. (1996). *In vitro* binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. *Proceedings of the National Academy of Sciences,* **93,** 9630-9635.

YARDENA, S.-I., BOAZ, F., RUTH, O.-W., YOAV, G., ALIZA, N., DAVID, S. & MICHAEL, G. (2002). Reappraisal of community-acquired bacteremia: a

proposal of a new classification for the spectrum of acquisition of bacteremia. *Clinical Infectious Diseases,* **34,** 1431-1439.

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

ZHAO, Y., WU, J., YANG, J., SUN, S., XIAO, J. & YU, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics,* **28,** 416-418.