Molecular Epidemiology of Uropathogenic *Escherichia coli* in North West England and characterisation of the ST131 clone in the region

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Science

2011

Tarek M. Gibreel

School of Medicine

List of Contents

Chapter One Introduction

Chapter Two Materials and Methods

Chapter Three Results

Chapter Four Discussion

Chapter Five References

Appendices (presented electronically in the attached CD)

Appendix I: Summary of Isolates demographic data.

Appendix II: Distribution of the phylogenetic groups among the 300 UPEC isolates.

Appendix III: Metabolic profiles for the 300 UPEC isolates with corresponding Bio-score and Bio-code.

Appendix IV: Antibiotic susceptibility profiles of the 300 UPEC isolates.

Appendix V: Distribution of different mechanisms of resistance as it is predicted by the Advanced Expert System (AES) software with corresponding resistance-score.

Appendix VI: Virulence factors profiles of the 300 UPEC isolates with corresponding virulence-score.

Total words count=45.390

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AST	Antibiotic Susceptibility Test
Bp	Base pair
BURST	Based Upon Related Sequence Types
CC	Clonal Complex
CFU	Colony Forming Unit
CGA	Clonal Group A
CIs	Confidence Intervals
DAEC	Diffusely Adherant *E. coli*
DLV	Double Locus Variant
dNTPs	Deoxyribonucleotide Triphosphate
EAEC	Enteroaggregative *E. coli*
EC	Enteric *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. Coli*
ECOR	*E. Coli* Reference collection
EPEC	Enteropathogenic *E. coli*
ERIC	Enterobacterial Repetitive Intergenic Consensus Sequence PCR
ESBL	Extended Spectrum β-Lactamase
ETEC	Enterotoxogenic *E .coli*
I^S_A	Standardized Index of Association
Kb	Kilobases
MLEE	Multilocus Enzyme Electrophoresis
MLST	Multilocus Sequence Typing

MRI	Manchester Royal Infirmary
NJ	Neighbor-Joining Cluster method
NMEC	Neonatal Meningitis associated *E. coli*
PAIs	Pathogenicity Islands
PCR	Polymerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis
PP	Plasmid Profile
QRDR	Quinolone Resistance Determining Region
RAPD	Random Amplification Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SEPEC	Sepsis-associated *E. coli*
SID	Simpson's Index of Diversity
SLV	Single Locus Variant
SNP	Single nucleotide Polymorphism
ST	Sequence Type
START	Sequence Type Analysis and Recombinational Tests
TBE	Tris Borate EDTA
TLV	Triple Locus Variant
UPEC	Uropathogenic *E. coli*
UPGMA	Unweighted Pair Group Method with Arithmetic Mean Cluster Analysis
UTI	Urinary Tract Infection
VF	Virulence Factor

ABSTRACT

Multilocus Sequence-Typing (MLST) is a phylogenetic technique based on the detection of differences in multiple conserved housekeeping genes. Together with powerful evaluation software, MLST provides an extensive classification scheme for highly diverse species. However, despite the increasing use of MLST as a trusted epidemiological tool, the population structure of UPEC has been poorly studied using this technique, as most of the previous studies conducted have been limited either by bias towards certain characteristics, such as antimicrobial resistance and serogroup, or included a limited number of strains. Such studies can give a false impression of the population structure due to overrepresentation of certain Sequence types (STs).

In this thesis, MLST was applied to 300 *E. coli* isolates collected from in the North West of England between June 2007 and June 2009. Firstly, the prevalence, diversity, epidemiological relationships and phylogenetic origins of the identified STs were determined. Secondly, possible associations of key UPEC STs with other genotypic and phenotypic profiles were assessed. Thirdly, as ST131 was recently reported as one of the most successful UPEC clones, an extensive examination of isolates of this clone was carried out involving identification of multiple drug resistant subclones and attempts were made to recognise putative predictor markers for identification of the ST131 clone.

MLST analysis of the studied population revealed a consistent profile of STs that occurred repeatedly in the collection. It consisted primarily of ST73 (16%) followed by ST131 (13.3%), ST69 (9%), ST95 (6.3%), ST10 (4.3%), ST127 (3.6%), ST14 (2.6%) and ST405 (1.6%) some of the STs (ST127 and ST80) in the panel have never been reported as remarkable uropathogens.

The broad range of virulence factor (VF) genes screened here allowed the recognition of VF patterns significantly associated with different STs. Most notably, ST127, which, based on phylogenetic analysis, appears to be a newly evolved clone, gave the highest virulence score. This virulent genotype may permit survival of ST127 isolates in the population long enough for them to gain antibiotic resistance. In contrast, multidrug resistant isolates of the ST131 clone were defined by a low virulence score and distinctive VF profiles.

Metabolic reactions have been conventionally used for the classification of bacteria into families and species. Interestingly, in the assessment of the metabolic activity of different STs, members of the ST131 clone showed a high metabolic capacity compared to those of other STs, which may compensate for the low virulence capacity and explain the virulence reported for members of this ST. In contrast, ST127 showed the lowest metabolic capacity, even though it held the highest VF-score among the commonly detected STs.

Multivariate logistic regression analysis demonstrated that ST131 is best described by its fluoroquinolone resistance and possession of PAI, the *ibeA* gene and expression of DR antigen-specific adhesins, whereas the O25b-CTX-M-15 ST131 sub-clone was only differentiated from the rest of the ST131 clone members by the production of Extended spectrum Beta-lactamase (ESBL) enzymes.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

1) The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

2) Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

3) The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

4) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisors, Dr. Mathew Upton and Dr. Andrew Dodgson, their understanding, encouraging and personal guidance have provided a good basis for the present thesis.

I am deeply grateful to Prof. Eric Bolton, Acting Director **o**f Regional Microbiology Network**,** HPA Central Office, for his advice and constructive comments, and for his important support throughout this work.

I wish to express my warm and sincere thanks to Prof. Andrew Fox in Health Protection Agency North West Laboratory and Manchester Medical Microbiology Partnership who gave me the opportunity to work with his team in the HPA laboratories.

I would like to extend my thanks to Dr. John Cheesbrough in Preston Microbiology Services, Royal Preston Hospital for providing me with uropathogenic *E. coli* isolates.

I owe my most sincere gratitude to Miss Audrey Coke and Mr. John Moat for technical assistance and support they have provided during the years.

During this work I had assistance and support from many colleagues especially Dr. Elloulu Ben-darif and Dr. Sally Lau, for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work in the Microbiology department, Manchester Royal Infirmary and the HPA regional laboratories.

I owe my loving thanks to my wife and children, without their encouragement and understanding it would have been impossible for me to finish this work and to whom this dissertation is dedicated. My special gratitude is due to my extended family for their loving support.

The administrative and financial support of the Libyan Cultural Affair's staff is gratefully acknowledged.

Publications

This project gave rise to the following published work, which includes conference presentations or proceedings, manuscripts to be submitted for publication.

Tarek M. Gibreel, Frederick Bolton, John Cheesbrough, Andrew Fox, Andrew Dodgson and Mathew Upton. (2010) Identification of CTX-M-15 producing O25b-ST131 *Escherichia coli* isolates on the basis of distinct antibiotic susceptibility and metabolic profiles. 20th European congress of Clinical Microbiology and Infectious Diseases (ECCMID), Vienna, Austria. (Poster presentation).

Tarek M. Gibreel, Frederick Bolton, John Cheesbrough, Mathew Upton , Andrew Fox and Andrew Dodgson. (2010) Analysis of the mechanisms of fluoroquinolone resistance in clonal group ST131 uropathogenic *Escherichia coli.* 20th European congress of Clinical Microbiology and Infectious Diseases (ECCMID), Vienna, Austria. (Poster presentation).

Tarek M. Gibreel, Frederick Bolton, John Cheesbrough, Mathew Upton , Andrew Fox and Andrew Dodgson(2008) Genetic diversity and population genetic structure of uropathogenic *Escherichia coli* isolated in Manchester, UK. *Escherichia coli* national meeting, British Society for Antimicrobial Chemotherapy, London, UK. (Poster presentation).

Tarek Gibreel, Andrew R. Dodgson, John Cheesborough, Andrew J. Fox, Frederick R. Bolton and Mathew Upton. Molecular analysis of UPEC from the Northwest region of England reveals recent local sub-clone diversification and low virulence in ST131 clone. (Under preparation).

Tarek M. Gibreel, Frederick Bolton, John Cheesbrough, Mathew Upton , Andrew Fox and Andrew Dodgson. Analysis of the Mechanisms of Fluoroquinolone Resistance in Clonal group 131 Uropathogenic *Escherichia coli* (under preparation).

Introduction

1. Introduction

Since it was discovered in 1885, *Escherichia coli* has been subjected to comprehensive studies and become one of the best understood and characterized organisms (Kuhnert, Boerlin & Frey 2000; Weintraub 2007).

Despite the vast knowledge accumulated throughout the last decades, *E. coli* is still one of the major causes of infection in humans. In addition to urinary tract infection (UTI), where it is responsible for most of the reported cases and gastrointestinal disease, *E. coli* has been associated with a number of other diseases including pneumonia and meningitis (Kaper, Nataro & Mobley 2004; Kuhnert *et al.* 2000).

Emergence of pathogenic *E. coli* strains has frequently been reported worldwide. Sporadic outbreaks caused by *E. coli* O157:H7 are increasingly being reported worldwide (Duffell *et al.* 2003) and in the United States it causes 73,000 enteric related illnesses with 2,168 hospitalizations and 61 deaths annually (Rangel *et al.* 2005).

Recently, a new class of Extended Spectrum Beta-Lactamase producing strain (called CTX-M producing *E. coli*) has emerged and spread globally and is most often seen in urinary tract infections (Lavigne *et al.* 2007; Mugnaioli *et al.* 2006; Woodford *et al.* 2007).

Certain multidrug-resistant, uropathogenic lineages of *E. coli* have been associated with outbreaks such as the one reported between 1987-1988 in south London where *E. coli* O15:K52:H1 caused community acquired cystitis, pyelonephritis and septicaemia (Manges *et al.* 2001).

1.1. Escherichia coli

Escherichia coli is a very diverse species of bacteria that forms a major part of the normal intestinal flora of humans and other mammals. Because of its wide distribution in the environment *E. coli* is able to colonise the intestines shortly after birth following the ingestion of contaminated food or water and is known to have a principle role in maintaining healthy guts (Kaper *et al.* 2004; Weintraub 2007).

1.2. Growth characteristics

E. coli is gram negative facultative anaerobic, motile bacteria, a taxonomically well defined member of the family *Enterobacteriaceae*. Its capability to grow in temperatures ranging from 7°C to 50°C and tolerate acidic condition down to pH 4.4 is reflected in its ability to survive various environmental conditions and it can be detected contaminating different surfaces from hands and clothes to soil and underground water (Kaper *et al.* 2004; Kramer, Schwebke & Kampf 2006; Kuhnert *et al.* 2000). It can be recovered easily from clinical specimens using simple culture media incubated at 37^{o}C either in the presence or absence of oxygen (Garrity 2005; Weintraub 2007).

1.3. Pathotypes of *Escherichia coli*

Strains constituting the species *E. coli* are of a broad variety ranging between non-pathogenic commensals that commonly inhabit the gastrointestinal tract of humans and other mammals and major pathogenic strains that may cause serious diseases.

According to its capability of causing disease, pathogenic *E. coli* are broadly classified into two major categories, enteric *E. coli* (EC) that mainly cause infections limited to

the mucous lining of the intestines and extra intestinal pathogens (ExPEC), which have the capability to spread from the intestine and to cause infections in other parts of the body (Kuhnert *et al.* 2000; Marrs, Zhang & Foxman 2005; Xie *et al.* 2006).

On the basis of their pathogenic features the enteric pathogenic *E. coli* are further differentiated into six pathotypes: enteropathogenic *E. coli* (EPEC); enterohemorrhagic *E. coli* (EHEC); enterotoxigenic *E. coli* (ETEC); enteroinvassive *E. coli* (EIEC); enteroaggregative *E. coli* (EAEC); and diffusely adherent *E. coli* (DAEC) (Kaper *et al.* 2004; Weintraub 2007).

1.3.1. Extra intestinal pathogenic *Escherichia coli*

Compared to enteric *E. coli*, ExPEC has not yet been well categorized. According to the type of disease they cause, the ExPEC are differentiated into: Uropathogenic *E. coli* (UPEC); Sepsis-associated *E. coli* (SEPEC); and Neonatal meningitis associated *E .coli* (NMEC) (Johnson *et al.* 2003a; Johnson & Russo 2005; Kuhnert *et al.* 2000).

However, *E. coli* is known to cause infections in other anatomical sites. It has been recently suggested that the use of an inclusive term, such as ExPEC, rather than the restrictive terms such as UPEC , SEPEC and NMEC, would be useful to reflect their broad infectious abilities (Johnson *et al.* 2003a; Johnson & Russo 2002).

1.4. *Escherichia coli* as an uropathogen

E. coli is the commonest pathogen to cause UTI, accounting for as much as 90% of all UTIs seen among non-hospitalized patients and up to 50% of all nosocomial UTI (Srinivasan, Foxman & Marrs 2003; Tartof *et al.* 2005). Due to anatomical differences (i.e. the shorter female urethra), UTIs are more common among women than men with

almost all women experiencing at least one UTI in their lifetime (Moore, Day & Albers 2002; Tartof *et al.* 2005). UPEC are responsible for 85-95% of uncomplicated cystitis in premenopausal women, the estimated number of cases being 130-175 million/year worldwide. In addition, UPEC cause 90% of uncomplicated pyelonephritis (5.4 million cases/year worldwide) (Russo & Johnson 2003).

1.4.1. Clinical aspects and pathogenesis

UPEC are associated with asymptomatic bacteriuria and uncomplicated cystitis as well as severe pyelonephritis. Infection usually occurs following the movement of UPEC from the intestinal tract and, in some cases, the vagina where it persists as a component of the normal flora, to the periurethral area (Kuhnert *et al.* 2000; Xie *et al.* 2006). These movements occur as a result of a complex combination of host behaviours, host susceptibility and certain abilities possessed by UPEC that have yet to be fully understood (Xie *et al.* 2006). From the periurethral area the bacteria ascend the urethra and colonize the bladder resulting in cystitis. Some strains may be able to invade the lining epithelial cells and multiply intracellularly, forming bacterial inclusions that may play an important role in recurrent UTIs (Schilling & Hultgren 2002). However from the bladder UPEC may spread to the kidneys and cause pyelonephritis and other serious complications may occur when the bacteria invade the blood stream resulting in systemic infection (Kaper *et al.* 2004).

Clinically, a total count of $>10^5$ CFU/ml of UPEC in midstream urine is considered as indicative of bacteriuria, however, in women a count of less than 10^5 CFU/ml appear to be associated with UTIs (Lloyd, Rasko & Mobley 2007). The clinical features vary considerably and depend in part on the age and the gender of the patient and other underlying clinical conditions. Disease ranges from painful urination in uncomplicated

urethritis or cystitis to severe systemic illness associated with abdominal or back pain, fever, sepsis and decreased kidney function in some cases of pyelonephritis (Naber *et al.* 2001).

Generally the ability of UPEC to move from the intestinal tract, colonize and cause infection in the urinary tract has resulted from a cumulative action of several specific virulence factors. The virulence factors include determinants that aid resistance to host defences, mediate adhesion, increase iron acquisition or paralyse ureteric peristalsis (Johnson 1991). Table 1.1 shows different types of virulence factors associated with different types of UTIs.

Simultaneous expression of multiple virulence factors appears to be more common among UPEC than faecal isolates and it is more common among UPEC isolated from upper UTI than isolates from lower UTI, suggesting synergistic activity of the virulence factors to overcome host defence and establish an infection (Johnson 1991; Yamamoto 2007).

Moreover, in a comparative genomic hybridization analysis of a set of UPEC and faecal / commensal *E. coli* isolates against a pyelonephritogenic *E. coli* strain (CFT073) using microarray, Lloyd and his colleagues were able to delineate 13 genomic islands of which 10 were previously unrecognized islands. The authors also identified 131 genes exclusively found in UPEC isolates, many of which are of hypothetical function and may play important role in pathogenesis of UTI (Lloyd *et al.* 2007).

Table 1.1 Summary of extra intestinal virulence factors and their association with UPEC

Factor	activity	Epidemiology	Remarks
P fimbriae	Adhesion	70%, 36%, 24% and 19% of isolates from pyelonephritis, cystitis, asymptomatic bacteriuria and faeces of healthy individuals, respectively.	Prevalence of P fimbriae among isolates from bacteraemia arising from UTI is as high as 71%, suggesting strong correlation of P fimbriae with the ability of UPEC to cause severe infections.
X adhesins Dr. family	Adhesion	Up to 50%, 26%, 6% and 18% associated with isolates isolated from cystitis, pyelonephritis, asymptomatic bacteriuria and faecal isolates respectively.	
S fimbriae	Adhesion	Mainly associated with meningitis and bacteraemia	Binding sites for S fimbriae are found on the epithelial cells of the proximal and distal tubules, collecting ducts and glomerulae and in the renal interstitium.
F1C	Adhesion	-	Binding site found on some renal tissue. Its role in facilitating UTI still unknown.
Type 1 fimbriae	Adhesion	Present in almost all *E. coli* isolates.	Binding sites are found on several cells including buccal, intestinal, vaginal epithelial cells suggesting a role in the colonization of mouth, gut and vagina.

Continued.

Factor	activity	Epidemiology	remarks
Siderophores			
Aerobactin	Iron acquisition	73%, 49%, 58%, 41% and 38% of isolates from pyelonephritis, cystitis, bacteraemia, asymptomatic bacteriuria and faecal isolates, respectively.	Proposed association with *E. coli* causing serious and complicated UTI, as it promotes bacterial growth in conditions limited in iron concentration.
Toxins			
Haemolysin (HlyA)	Lysis of erythrocytes and leukocytes	Present in 49%, 40%, 20% and 12% of isolates from pyelonephritis, cystitis, asymptomatic bacteriuria and faecal isolates.	Mainly associated with invasive UTI. It plays a complex role from aid in iron acquisition to disruption of phagocyte function and direct toxicity to host cells.
Cytotoxic necrotizing factor 1 (CNF1)	Inhibit phagocytic activity	Frequently associated with UTI isolates and significantly related to prostatitis.	
Secreted autotransporter toxin (Sat)	Vacuolating cytotoxin	More frequently associated with pyelonephritis causing isolates.	
Capsular polysaccharide			
K antigen	Antiphagocytic	Certain types of K antigen commonly detected among isolates from pts UTI in comparison to faecal isolates. K1 & K5 were detected in 63% of isolates from women with pyelonephritis.	Together K & O antigens serve as an important tool to differentiate UPEC from other *E. coli* strains.
O antigen	Serum resistance	Certain types of O antigen exhibit anticomplementary activity are associated with *E. coli* isolated from UTIs, (such as O1, O2, O4, O6, O16, O18, O22, O25 and O75).	

Summarized from: (Guyer *et al.* 2002; Johnson 1991; Kaper *et al.* 2004; Mills, Meysick & O'Brien 2000)

1.5. Antimicrobial resistance in UPEC

As one of the most frequently isolated pathogens in clinical practice, UPEC are considered to be a major reservoir for genes encoding antimicrobial resistance. Several factors may contribute to the development and spread of antibiotic resistance including volume of antibiotic use, poor hygienic conditions, use of antibiotics in animal feeds and overcrowded living conditions together with bacterial virulent characteristics (Erb *et al.* 2007; Lau *et al.* 2008b; Moreno *et al.* 2009). In addition to pathogenic *E. coli*, antibiotic resistance may involve commensal *E. coli* in the bowel, which may become a major reservoir of resistant strains (Erb *et al.* 2007).

Results from the Sentry antimicrobial surveillance programme, which is a longitudinal surveillance program designed to track global antimicrobial resistance trends, show remarkable variations in antibiotic resistance involving most of the known antibiotic groups with high resistance rates reported from South America and Asia and the lowest in Europe and North America (Fluit *et al.* 2000; Gordon & Jones 2003).

1.5.1. Beta-lactam antibiotics

These are the most widely used antibiotics in clinical medicine (Bush & Macielag 2010). In the treatment of UTI, amoxicillin with or without clavulanic acid and the first generation cephalosporins have been used in the treatment of uncomplicated UTIs, whereas third generation cephalosporins are recommended for treatment of complicated upper UTI (Anonymous 2006).

Resistance to beta-lactam antibiotics is mostly associated with bacterial production of different beta-lactamase enzymes that break the beta-lactam ring and inactivate the antibiotics (Bush & Macielag 2010).

1.5.1.1. Beta-lactamases

TEM1 and TEM2 are the most common plasmid-mediated beta-lactamases in Gram-negative bacteria, including *E. coli*. In addition to a less common enzyme termed SHV, all are able to hydrolyze penicillins and narrow spectrum cephalosporins, such as cephalothin but they are not effective against higher generation cephalosporins such as cefotaxime, ceftazidime, ceftriaxone, or cefepime. However, their action can be overcome with beta-lactamase inhibitors like clavulanic acid (Canton & Coque 2006).

1.5.1.2. Extended Spectrum Beta-lactamases

Shortly after the introduction of cefotaxime early in 1980s, new derivatives of the common TEM-1, -2 and SHV-1 enzymes, named SHV2 and TEM3, were detected with transferable resistance to the oxyimino-cephalosporins (eg, cefotaxime, ceftazidime, and ceftriaxone). They were named extended spectrum beta-lactamases (ESBLs) in 1989 by Philippon and colleagues (Philippon, Labia & Jacoby 1989). At the same time, a new family of ESBLs was detected and named CTX-M. These were characterized by higher levels of resistance to cefotaxime than to ceftazidime and the first description was in an isolate from Munich (Canton & Coque 2006).

Although TEM and SHV ESBLs were dominant among ESBLs during the 1990s, occasional nosocomial outbreaks, mostly of CTX-M-2-producing Enterobacteriaceae, were reported (Pitout 2009). This situation changed during the 2000s with the worldwide emergence of CTX-M-15 producing *E. coli* as an important cause of community-onset UTIs (Canton & Coque 2006; Coque *et al.* 2008; Ellington *et al.* 2006; Lavigne *et al.* 2007; Pitout 2009; Woodford *et al.* 2007). According to the www.lahey.org database accessed on January 2011, 137 different CTX-M enzymes

have been identified (www.lahey.org/studies/other.asp#table1) and can be clustered into five different groups (CTX-M-1, -2, -8, -9 and -25) (Canton & Coque 2006).

CTX-M-15 beta-lactamase belongs to the CTX-M-1 group and is characterized by an increased catalytic activity against ceftazidime, and bacteria producing these enzymes often test resistant to this agent. It has often been associated with carriage of other beta-lactamases such as TEM-1 and OXA-1 as well as the aminoglycoside-modifying enzyme AAC(6')-Ib-cr that has an additional ability to acetylate fluoroquinolones offering resistance to norfloxacin and ciprofloxacin (Peirano & Pitout 2010).

1.5.2. Fluoroquinolone

Quinolones are family of synthetic antibacterial agents with broad spectrum activity. They bind DNA gyrase and topoisomerase IV when they are in a complex with DNA and inhibit chromosome re-ligation after enzyme-mediated cleavage (Walker 1999). Quinolones have long been used in the treatment of UTIs with the first developed quinolone (nalidixic acid) in 1962, which had marked antibacterial activity against gram-negative bacteria making it suitable for the treatment of UTI. This was followed by introduction of the fluoroquinolones with broad spectrum activity and better systemic distribution allowing use in upper and complicated UTIs (Ball 2000). Ciprofloxacin is the fluoroquinolone currently recommended for the treatment of upper and complicated UTI (Anonymous 2006). Due to the increased resistance against most commonly used antibiotics, fluoroquinolones have been used with increasing frequency in both complicated and uncomplicated UTI leading to a rapid increase in fluoroquinolone resistance (Jacoby 2005).

Resistance to fluoroquinolones typically arises as a result of a series of mutations in the target enzymes (DNA gyrase and topoisomerase IV) and of changes in drug entry and efflux. Mutations are selected first in the more susceptible target genes coding DNA gyrase and topoisomerase IV, *gyrA* resulting in resistance to nalidixic acid, whereas additional mutations in the next most susceptible target *parC* augment resistance further, to involve fluoroquinolones. Resistance to quinolones can also be mediated by plasmids that produce the Qnr protein, which protects the quinolone targets from inhibition (Hooper 2001).The resistance to fluoroquinolones in *E. coli* has increased significantly over the last decade. In the annual reports of the European Antimicrobial Resistance Surveillance Network (EARS-Net), 28 of 29 countries in Europe reported a significant increase in fluoroquinolone resistance between 2001–2007 and 8 of 28 for the period 2006- 2009 with the majority reporting resistance around 20% (with a range between 7% and 43%). In the UK, resistant isolates of *E. coli* increased from 7–18% between 2001 and 2009(Anonymous 2007; Anonymous 2009).

1.5.3. Trimethoprim

Trimethoprim is a bacteriostatic antibiotic that has been widely used alone or in combination with sulphonamides (SXT) in the treatment and prophylaxis of UTI (Huovinen *et al.* 1986). Trimethoprim acts by interfering with the action of bacterial dihydrofolate reductase, inhibiting synthesis of tetrahydrofolic acid leading to the inhibition of DNA replication and resistance in *E. coli* is mainly mediated by plasmid mediated gene (*dfr*) encoded modified dihydrofolate reductases whose configuration is not susceptible to trimethoprim (Skold 2001).

Although resistance to trimethoprim alone or as a combination with sulphonamides has increased during recent decades (Steinke *et al.* 2001), it is still recommended for

treatment of uncomplicated UTI(Anonymous 2006). However, in 2001, Manges *et al*. described a clonal group of SXT resistant *E. coli*, named clonal group A (CGA), which was found among women with acute uncomplicated UTI in one university community (Manges *et al.* 2001). Since then CGA isolates have been recovered from several populations across the United States (Johnson *et al.* 2002; Manges, Dietrich & Riley 2004; Manges *et al.* 2001) and has also been described in other parts of the world (Johnson *et al.* 2009; Johnson *et al.* 2005c; Manges *et al.* 2008) suggesting clonal dispersion leading to a remarkable increase in SXT resistance (Johnson *et al.* 2002).

1.5.4. Aminoglycosides

Aminoglycosides are a group of broad spectrum bactericidal drugs, which includes many drugs such as gentamicin, amikacin, streptomycin, tobramycin, kanamicin and neomycin. They share chemical, antimicrobial, pharmacologic and toxicity characteristics and because of their poor pharmacokinetics and considerable toxicity their use been limited to serious infections (Mingeot-Leclercq, Glupczynski & Tulkens 1999).

Aminoglycosides act by inhibiting protein synthesis in bacteria through binding to the ribosomes. In gram negative enteric bacteria, resistance to aminoglycosides is mainly mediated by plasmid encoded aminoglycoside modifying enzymes (Mingeot-Leclercq *et al.* 1999). To date, over 85 aminoglycoside modifying enzymes have been identified, however only few such as AAC(3)-I, AAC(3)-II, AAC(3)-III, AAC(3)-IV, AAC(6')-I and ANT(2'') appear to cause the majority of aminoglycoside resistance among *E. coli* isolates (Ho *et al.* 2010)

Although aminoglycosides are not recommended for the treatment of UTI in UK(Anonymous 2006), aminoglycoside resistant *E. coli* have previously been reported (Johnson *et al.* 1995; Woodford *et al.* 2007), Recent studies have suggested food producing animals are an important reservoir for the aminoglycoside-resistance gene found in human isolates (Chaslus-Dancla *et al.* 1991; Johnson *et al.* 1994).

1.5.5. Overview of antimicrobial drug resistance in UPEC

During recent decades, antimicrobial resistance has significantly increased among UPEC (Erb *et al.* 2007; Moreno *et al.* 2006; Yu *et al.* 2007) and unexpected trends of antimicrobial resistance among *E. coli* have been reported globally. In the UK, concern was raised over the emergence of community acquired gentamicin resistant UPEC (Woodford *et al.* 2007), as well as the increase in ciprofloxacin resistance in *E. coli* isolated from blood (Livermore *et al.* 2003) and the global spread of CTX-M ESBL producing strains (Canton & Coque 2006; Ellington *et al.* 2006; Lavigne *et al.* 2007; Woodford *et al.* 2007). In 2001, Manges and colleagues, reported community spread of one clonal group of UPEC responsible for up to 50% of *E. coli* isolates resistant to co-trimoxazole derived from women with acute cystitis in California, Michigan and Minnesota, USA (Manges *et al.* 2001).

Several biological mechanisms may contribute to development and spread of these distinctive resistance patterns in UPEC. Although mutations responsible for antibiotic resistance are in some cases a result of positive selection pressure (Martinez & Baquero 2000), the main mechanism for the development of antibiotic resistance is horizontal gene transfer, which has been considered an important route for transmission of virulence factors and antimicrobial resistance in *E. coli* (Blahna *et al.* 2006).

The conjugative plasmids responsible for the spread of antibiotic resistance usually contain an integron structure consisting of an *int*-1-gene encoding the integrase that catalyzes the integration and excision of the gene cassettes encoding antibiotic resistance (Hall & Collis 1998). For trimethoprim resistance, where horizontal dissemination is so far considered to be the main route of spread, at least 30 different *dfr* genes are known to be responsible for trimethoprim resistance. These genes are mostly located on plasmids and reside within integrons (Blahna *et al.* 2006).

The role of plasmids in dissemination of antibiotic resistance has long been established and involves almost every class of clinically important antibiotics. The recent spread of ESBL producing strains in the community has been strongly associated with plasmid mediated CTX-M enzymes particularly CTX-M 15 (Nicolas-Chanoine *et al.* 2008). Similarly, the rise in fluoroquinolone resistance has increasingly been associated with plasmid mediated *aac (6)-lb-cr* and *qnr* genes (Coque *et al.* 2008; Sabtcheva *et al.* 2009), which suggests horizontal transfer of these resistance encoding plasmids (Nicolas-Chanoine *et al.* 2008) and the fact that these different resistance genes co-exist on same plasmid, together with other antibiotic resistance genes, facilitates their dissemination through co-selection processes. However, the fact that many of these geographically dispersed strains share considerable genomic and phenotypic characteristics, supports an alternative explanation, that is of clonal expansion (Nicolas-Chanoine *et al.* 2008).

However, the extent that each mechanism contributes to the observed rise in antibiotic resistance is not always clear. Under conditions where horizontal gene transfer was sufficient to drive the dissemination of antibiotic resistance with clonal expansion amplifying the genes within individual hosts, regionally independent gene distributions were expected. Conversely, in cases where horizontal gene transfer is rare, clonal

expansion accounts for most of the increase in the resistance level and region-dependent gene distributions are to be expected (Blahna *et al.* 2006).

The role of mutators (section 1.6.2.1.2) in causing the high prevalence of antimicrobial resistance among UPEC is still under debate. However, a strong correlation between antimicrobial resistance and high frequency mutators is often reported, suggesting that it could be a result of the selection pressure exerted by commonly used antibiotics (Baquero *et al.* 2005; Denamur *et al.* 2002; Ellington *et al.* 2006; Miller, O'Neill & Chopra 2004). In addition, the increasing use of fluoroquinolones, which are known to be a mutagenic drugs, as an alternative therapy for UTIs could explain the high prevalence of hyper mutable ESBL producing strains (Baquero *et al.* 2005).

Furthermore, the ability to form biofilms and to develop intracellular bacterial communities within murine bladder urothelium allows UPEC to establish reservoirs protected from the immune system and antibiotic treatment and serve as a persistent source of bacteria. This may contribute to the emergence of antibiotic resistant UPEC strains (Parsek & Singh 2003; Rosen *et al.* 2007)

Recently, management of UTIs has become increasingly challenging as a result of emerging resistance to most first-line antimicrobial agents, necessitating revised empirical treatment approaches (Cagnacci *et al.* 2008).

Understanding the rules governing the interplay among all these discrete factors involved in the development, transfer and spread of antibiotic resistance could help explain the emergence and dissemination of successful strains and increase an understanding the evolution and population structure of UPEC may eventually facilitate development of better management strategies.

1.6. Epidemiology of UPEC

The epidemiology of UTIs caused by UPEC has always been difficult to assess as UTI in general is not a reportable disease in many countries (Foxman *et al.* 2002; Stamm 2001). However, recent community spread of certain lineages of UPEC with abnormal characteristics, such as resistance to antibiotics or high level of virulence, provide evidence of sporadic epidemics or outbreaks, which otherwise would have remained unrecognised (Johnson *et al.* 2009; Leflon-Guibout *et al.* 2004; Manges *et al.* 2001; Woodford *et al.* 2007). In addition, with a lack of ongoing surveillance, conventional epidemiology fails to appreciate fully the specificity of UPEC and the fact that its pathogenicity is distinct from other *E. coli*. In contrast, molecular epidemiological tools provide a closer look at microbial traits explaining virulence behaviour and host predilections of the pathogen and offer insights into the origins and spread of bacterial diseases (Johnson & Russo 2005).

1.6.1. Molecular epidemiology of UPEC

The completion of the genome sequence of *E. coli* K-12 created a great opportunity to determine the precise functions of all encoded genes. This provided data for complete genome comparisons, to identify genes coding for unique properties in related pathogens, as well as giving an insight into the evolutionary relations of different pathogens (Blattner *et al.* 1997). However, the available genetic data on pathogenic *E. coli* indicates that up to 10-20% of the genomic information found in highly pathogenic *E. coli* is not present in *E. coli* K-12. Most of the pathogen specific DNA appears to encode various virulence factors responsible for the pathogenicity of *E. coli* (Kuhnert *et al.* 2000).

1.6.2. Evolution of UPEC

Based on phenotypic studies carried out on a defined standard reference set of natural isolates from a variety of hosts and geographical locations (the ECOR collection), the natural populations of *E. coli* appear to exist as a collection of lineages with limited exchange of chromosomal genes (Ochman & Selander 1984; Selander & Levin 1980). On the other hand, the rapid development of molecular techniques such as polymerase chain reaction (PCR), DNA sequencing and molecular typing has had a great impact on the study of bacterial evolution and, thus, genetic studies of UPEC show evidence of a much more diverse population structure (Foxman 2007).

1.6.2.1. Sources of genetic variation

Generally, three basic mechanisms can generate genetic variation leading to bacterial diversity: mutation; hypermutation; genetic recombination. The impact of each mechanism on the evolution and the diversity of UPEC will be considered below.

1.6.2.1.1. Mutation

Referring to any inheritable change in genetic material, mutations are basically caused by DNA replication errors and include substitutions and insertion or deletion of nucleotides. In some cases this is promoted by the activity of different transposable elements.

1.6.2.1.1.1. Substitutions

Substitutions or point mutations can be synonymous (changes in DNA sequence do not affect the translated amino acid) or non-synonymous (changes in DNA sequence result in altered amino acid sequence). Such mutations occur randomly but continuously in most bacteria during DNA replication. Most synonymous substitutions are free from

natural selection and therefore provide a good measuring index for time of divergence between two species. On the other hand, non-synonymous substitutions are mostly deleterious in nature due to positive selection. Detection of both types may therefore provide information about the degree of selection exerted on a defined gene (Kimura 1983).

1.6.2.1.1.2. Insertions and deletions

Insertions and deletions is another type of mutation, which could be the result of errors in DNA replication in which a few bases are inserted or deleted, such as seen with slipped-strand mispairing (Levinson & Gutman 1987). Insertion or deletion of large blocks of DNA is usually caused by DNA crossover or DNA transposition. Generally if it occurs in a coding region, it causes frame shift mutations, which almost always result in altered amino acid sequence (Snyder & Champness 2007a).

1.6.2.1.1.3. Inversions

Inversion mutations are usually caused by homologous recombination between inverted repeats in the same DNA fragment in which a DNA sequence is flipped over and lies in the reverse orientation. Inversion mutations often cause no phenotypic alteration and rarely occur in the evolution of *E. coli* (Mahan & Roth 1991).

1.6.2.1.2. Hypermutation

Hypermutators, which are strains that show a higher mutation rate than wild type organism due to deficient mismatch repair systems, are able to exchange genetic material at a higher rate than the wild type organism, which eventually increases the rate of genetic variation (LeClerc *et al.* 1996; van Belkum *et al.* 2001). Hypermutators have been detected among different bacterial populations including *Neisseria meningitidis*, *Salmonella enterica* and *E. coli*. Although their evolutionary value is

limited, UPEC generally express a high frequency of mutators (Denamur *et al.* 2002; Hall & Henderson-Begg 2006; van Belkum *et al.* 2001). Furthermore, they could be associated with the development of chronic UTI through the development of persistent intracellular reservoirs (Labat *et al.* 2005).

1.6.2.1.3. Homologous recombination

This occurs more often and refers to the genetic exchange between homologous sequences of closely related organisms and affects the variation in existing genes rather than introducing new genetic information into the genome. The fact that almost all organisms have some kind of homologous recombination mechanism indicates its important role in bacterial survival allowing rapid adaptation to the environment (Snyder & Champness 2007b).

1.6.2.1.4. Lateral gene transfer

Similar to homologous recombination, lateral gene transfer involves acquisition of new genetic material, but unlike homologous recombination it involves different strains/species and results in introduction of new genetic information into the lineage and, hence, is often referred to as non-homologous recombination (Doolittle 1999). Recent studies on genetic evolution of enteric bacteria reveals an integral role of lateral gene transfer in the diversification and speciation of enteric organisms (Lawrence & Ochman 1998; Ochman, Lawrence & Groisman 2000). Although the level of recombination and its effect on population structure in *E. coli* are still a matter of debate, several reports strongly suggest that horizontal transfer has a great impact on the evolution and the diversity of UPEC through acquisition of new genetic material either by conjugation or transduction (Guttman & Dykhuizen 1994; Oelschlaeger, Dobrindt & Hacker 2002; van Belkum *et al.* 2001).

Comparative studies on genetic structure of *E. coli* and other enteric pathogens were able to identify reliable indicators of lateral gene transfer, which include atypical GC content, mosaic alleles and inconsistency between gene trees (Lawrence & Ochman 1998).

1.6.2.1.5. Pathogenicity islands (PAIs)

These were first described in *E. coli*, but have recently been found in the genomes of various pathogens of humans, animals and plants (Schmidt & Hensel 2004).

Pathogenicity islands, which are unstable regions of chromosomal DNA 10-200 kilobases (kb) in size that encode determinants responsible for extra intestinal pathogenicity, were frequently detected in UPEC isolates but are absent from the genomes of non-pathogenic *E. coli*. The instability of these pathogenic sequences is related to its acquisition of mobility genes (e.g. integrases and insertion sequences) and the association with tRNA genes, which have been suggested to act as integration sites for foreign DNA.

The G+C content of pathogenicity islands frequently differs from the rest of the genome, suggesting that they have been acquired from other related bacterial species by horizontal gene transfer (Lloyd *et al.* 2007). In their study of the genetic variation of the P-associated pilus gene (*pap*) in different UPEC, Plos *et al* (1989) provided evidence supporting the horizontal gene transfer hypothesis for *pap* genes among UPEC (Plos *et al.* 1989) and similar findings were reported by others (Bingen *et al.* 1998; Hacker & Kaper 2000).

1.6.3. Population genetics

Population genetics of bacteria is the study of genetic variation of bacterial populations and is used to explain the evolutionary basis of genetic variation between and within bacterial species (Trevors 1998).

Earlier studies suggested that *E. coli* form a clonal population structure (Achtman *et al.* 1983; Achtman & Pluschke 1986; Ochman & Selander 1984).

In typical clonal populations, clonality arises solely from vertical transmission of genetic information through binary fission whereby new lineages emerge by the accumulation of mutations over successive generations. Consequently, the distribution of chromosomal polymorphisms will be non-random, or in linkage disequilibrium (Spratt & Maiden 1999).

However, bacterial make attempts to conserve genomic integrity and possess a number of mechanisms that, in addition to DNA-damage repair, promote genetic diversity through homologous exchange of sequences between members of the same clone or otherwise different clones (Aguilera *et al.* 2007). The frequency of such recombinational events in bacterial populations, which can vary from very low (clonal populations) as seen with *Enterococcus faecium*, to very high (non-clonal populations) such as *Helicobacter pylori*, determine the population structure. Differences in the ratio of genetic change caused by recombination in relation to that caused by mutation leads to a spectrum of bacterial population structures, with most diversification in the bacterial population involving recombination. However, recombination often will not prevent the emergence of clonal lineages within the population (Spratt & Maiden 1999; Turner *et al.* 2007).

1.6.4. Population structure of UPEC

The strong association of certain serotypes with UTIs (Johnson *et al.* 1997; Kuhnert *et al.* 2000) and with certain subgroups of *E. coli* such as UPEC clonal group A strains, which share similar O antigen groups (O11, O17, O73 & O77) (Kuhnert *et al.* 2000; Lloyd *et al.* 2007; Prats *et al.* 2000; Tartof *et al.* 2005) initially lead to the suggestion that *E. coli* populations are in general clonal in nature. The successful application of techniques such as multilocus enzyme electrophoresis (MLEE), which indexes the allelic variation in multiple chromosomal genes, generated large databases that allowed statistical analysis of bacterial populations with enough data to strongly support the suggestion that populations exhibit strong linkage disequilibrium (non-random association of alleles) and revealed the rarity of recombinational exchanges between bacterial lineages (Feil *et al.* 2001; Smith *et al.* 1993).

Conversely, analysis of closely related *E. coli* clones reveals frequent involvement of recombination in the diversification of *E. coli* (Guttman 1997; Guttman & Dykhuizen 1994).

Phylogenetic analysis of MLEE profiles of the *E.coli* Reference collection (ECOR) strains define four major phylogenetic groups (A, B1, B2, D) and ungrouped strains (UG) (LeCointre *et al.* 1998). Most ExPEC are derived from group B2 and to a lesser extent from group D, while the other pathogenic *E. coli* are consistently distributed in all classes with the commensal strains belong to group A (Clermont, Bonacorsi & Bingen 2000). Results from molecular studies of the recognised extra-intestinal virulence factors are compatible with this classification, where B2 group strains are highly pathogenic with numerous virulence determinants responsible for extra intestinal infections and strains of phylogenetic group D seem to have fewer virulence determinants than B2 group strains, whereas strains of the group A and B1, are most

often devoid of extra intestinal virulence determinants (Figure 1.1) (Johnson & Russo 2005). Recent efforts have been applied to further sub-typing of UPEC based on acquisition of specific pathogenicity islands, which could provide an ideal diagnostic tool for UTI caused by UPEC (Kanamaru *et al.* 2006a).

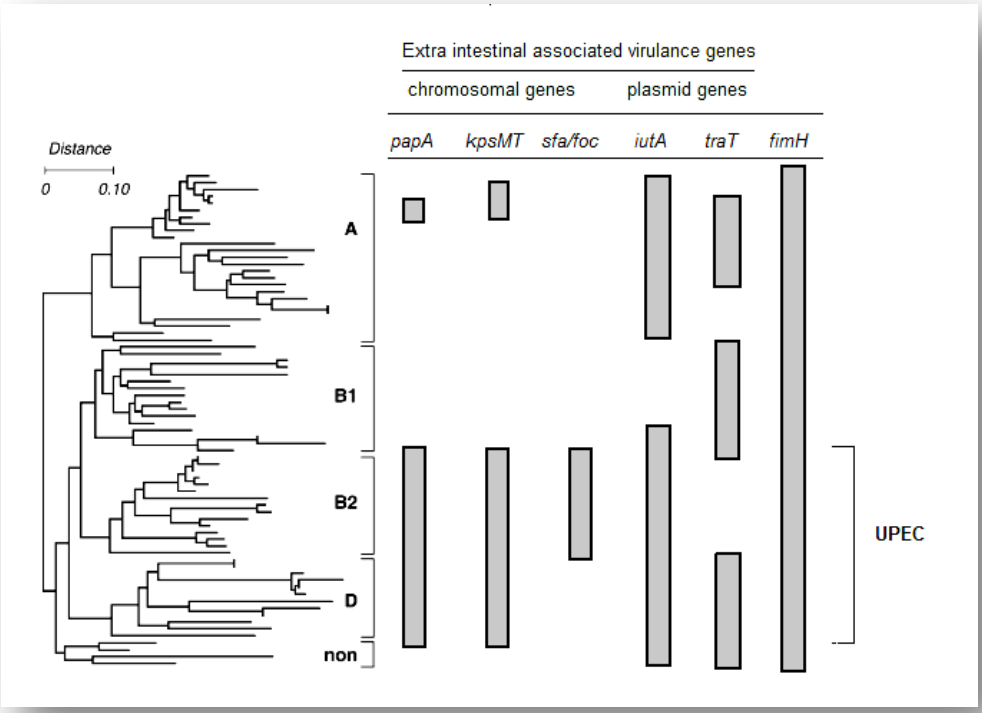


Figure 1.1 Phylogenetic distribution of Extra intestinal virulence associated genes in *E. coli* adapted from (Johnson & Russo 2005) (*papA*, P fimbriae; *kpsMT*, group II capsule synthesis; *sfa/foc*, S and F1C fimbriae; *iutA*, aerobactin; *traT*, serum resistance; *fimH*, type I fimbriae)

1.7. Methods used for typing pathogenic *E. coli*

The ability to carry out epidemiological investigations to determine the population structure and the epidemiological distribution of pathogens is important to increase our understanding of the evolutionary processes of such pathogens, and subsequently improve public health.

Usually, the closeness of strains is a reflection of the phenotypic and genotypic variation seen either in bacterial behaviour toward the environment or the chromosomal

DNA of the examined strains. Therefore, the ability of a typing technique to identify related strains depends greatly on the rate at which these variations occur in the tested parameter (Tenover, Arbeit & Goering 1997).

Several successful approaches have been introduced for typing and classification of *E. coli*. These range from novel methods to modification of existing ones to enable more precise, sensitive, cheap and most of all rapid methods for the identification and differentiation of UPEC strains.

Based on the target parameter, typing methods can be divided into two categories

1. Phenotyping methods 2. Genotyping methods

Phenotyping methods are those that characterize the products of gene expression in order to identify and further classify strains into different subgroups such as biotyping, phage typing, antibiotyping, serotyping, polyacrylamide gel electrophoresis and multilocus enzyme electrophoresis. All of which differ in their characteristics such as sensitivity, precision and availability, but most of them have limited discriminatory power (van Belkum *et al.* 2001).

Genotypic methods are those that are based on direct analysis of the genetic structure of examined strains. The nature of the genetic screen will vary depending on whether it is for the detection of certain genes or plasmids or screening for pathogenicity islands or a certain clonal/phylogenetic group. Several typing methods are commonly used for *E. coli* including , pulsed field gel electrophoresis (PFGE), restriction fragment length polymorphism analysis (RFLP), ribotyping, random amplification of polymorphic DNA (RAPD), enterobacterial repetitive intergenic consensus sequence PCR (ERIC)

and multilocus sequence typing (MLST) (Johnson & Russo 2005; Olive & Bean 1999; Power 1996).

Most of these methods differ in their characteristics such as typeability, discriminatory power, reproducibility, cost and time consumption (Johnson & Russo 2005; van Belkum *et al.* 2001) (Soll, Pujol & Lockhart 2007)

Moreover, using these techniques to examine the relation between strains depends greatly on the rate at which changes occurs in the examined parameter. Phenotyping methods such as biotyping and antibiotyping tend to change at a rapid pace according to the changes in the growth environment, which makes them mostly suitable to detect recent changes among closely related strains, providing sufficient information for primary epidemiological investigation (Hopkins & Hilton 2000).

1.7.1. Phenotyping methods

1.7.1.1. Biotyping

Generally biotyping, like most phenotypic methods, is considered to be an unreliable epidemiological tool because of its modest reproducibility and poor discriminatory power. However, advancement in automated biotyping methods provide more reliable tools that rely on a variety of novel substrates and precise interpretation techniques that increase the reproducibility and discriminatory power (Tenover *et al.* 1997).

The reliability of biotyping as an epidemiological tool depends greatly on the pathogen in question, as bacteria vary in their biochemical reactions, some bacterial species lacking biochemical diversity, whereas others, such as *E. coli*, have considerable biochemical variability (Godbout-DeLasalle & Higgins 1986). Despite this, biotyping methods designed for identification of strains of the family *Enterobacteriaceae* have

been shown to be dependable and a useful method in characterization of *E. coli* (Brauner *et al.* 1987; Leclercq *et al.* 2001).

Moreover, automation of biochemical fingerprinting methods, based on numerical analysis of biochemical reaction kinetics to differentiate bacterial strains, has been proven to be highly discriminatory and reproducible when fermenting bacteria are tested and the resulting classifications were comparable to that of PFGE (Kuhn 1985; Kuhn *et al.* 1995).

1.7.1.2. Antibiogram typing

Similar to biotyping, antimicrobial susceptibility patterns are of poor discriminatory power and of low reproducibility because of the selective pressure exerted by frequent exposure to antibiotics, especially in healthcare facilities. In addition, the fact that many antimicrobial resistance genes are associated with mobile genetic elements significantly decreases the reliability of antibiotyping methods as an epidemiological tool (Tenover *et al.* 1997).

However, detection of unusual antimicrobial susceptibility patterns often indicates spread of certain strains and, thus, can be used in combination with other typing methods to provide an initial screening tool (Manges *et al.* 2001; Phillips *et al.* 1988).

1.7.1.3. Serotyping

Serotyping remains an important epidemiological tool to describe lineages and population composition. It is based on antibodies directed against 173 O antigens, 80 K antigens and 56 H antigens creating up to 100,000 or more possible combinations (Orskov & Orskov 1992). Several studies showed a strong association of different serotypes with different pathotypes. Certain serotypes are commonly associated with UTIs (Johnson *et al.* 1997; Kuhnert *et al.* 2000) and with certain subgroups of *E. coli*

such as UPEC clonal group A strains, which share similar O antigen groups (O11, O17, O73 & O77) (Kuhnert *et al.* 2000; Lloyd *et al.* 2007; Prats *et al.* 2000; Tartof *et al.* 2005).

However, the high diversity of *E. coli* requires strains to be tested against a large number of antisera, which reduces the practicality of using this method and increases the cost further limiting its use.

1.7.1.4. Multilocus enzyme electrophoresis (MLEE)

Until the application of DNA sequencing became more routinely available, multilocus enzyme electrophoresis had a great role in describing the phylogenetic structure of *E. coli* and other organisms. It uses the electrophoretic mobility of enzyme variants to detect polymorphism in housekeeping genes. Variations in electrophoretic mobility are the result of DNA variations at sites leading to amino acid changes (Selander *et al.* 1986). Regardless of its contribution to global epidemiological studies, multilocus enzyme electrophoresis has limitations regarding standardization and comparability with data from other laboratories (Selander *et al.* 1986; Urwin & Maiden 2003).

1.7.2. Genotyping

1.7.2.1. Restriction based methods

1.7.2.1.1. Plasmid analysis

Plasmids are self replicating, extrachromosomal circles of DNA that encode a wide variety of genes, including those mediating antimicrobial resistance, virulence and metabolism of hydrocarbons (Snyder & Champness 2007c).

Upon cell division, copies of plasmid are transferred to daughter cells and subsequently cells of the same clone carry the same plasmids. The number and the size of plasmid in each cell provide a comparable profile that can be used to differentiate strains. Despite the natural ability of plasmids to transfer horizontally, which can interfere resolution of a strain's true relationship, and its low discrimination power, plasmid profiling still provides a reliable initial epidemiological screening tool (Foley, Lynne & Nayak 2009). In addition to antibiotic resistance, other features may be plasmid coded such as serotype or phage susceptibility and, thus, different typing systems may not be totally independent of one another. Plasmid profiles are most useful when combined with other methods of screening or typing (Mayer 1988).

Following the first report of using agarose gel electrophoresis for plasmid profile characterization using *E. coli* K-12 to standardize the method (Meyers *et al,* 1976), a number of widely used procedures for plasmid isolation and profiling have been described, including those by (Birnboim & Doly 1979; Kado & Liu 1981; Samrook & Russell 2001).

Most of these methods rely on alkaline denaturation of chromosomal DNA while supercoiled plasmid DNA remains intact and separation of the nucleic acid is carried out on the basis of physical differences between circular plasmid and linear chromosomal fragments in the purification process (Cloninger *et al.* 2008).

One of the problems with these methods is in the coexistence of three conformations (linear and supercoiled) of the same plasmid, which may affect the migration properties of plasmids during gel electrophoresis (Olsen *et al.* 1993).

To overcome this drawback and improve discrimination between strains, restriction endonucleases, such as *HindIII,* can be used to generate more distinctive profiles (Foley *et al.* 2009).

From the literature, it can be seen that plasmid profiling has been used for a wide range of pathogens such as *E. coli* (Domingue *et al.* 2003);(Bebora *et al.* 1994; Jan, Meshram & Kulkarni 2009), *Shigella* (Dutta *et al.* 2002) and *Salmonella* (Mohan *et al.* 1995; Threlfall *et al.* 1990).

1.7.2.1.2. Pulsed field gel electrophoresis (PFGE)

PFGE is often referred to as the gold standard of molecular typing methods, as it is characterized by a high discriminatory power that is superior to other typing methods (Olive & Bean 1999).

PFGE directly detects variation in the genetic sequence of chromosome by scanning the entire genome. The bacterial genome is digested by specific restriction enzymes followed by electrophoretic separation using a current with variable polarity. The resulting electrophoretic patterns are strain specific (Olive & Bean 1999; Tenover *et al.* 1995).

A few limitations have been associated with PFGE including standardization and time consumption (Olive & Bean 1999; van Belkum *et al.* 2001). In an attempt to improve its reproducibility Tenover *et al.* (1995) suggested interpretation criteria for the analysis of restriction patterns.

From the literature, PFGE appears to have been used in epidemiological studies of a broad range of microorganism (Fukigai *et al.* 2007; Germon *et al.* 2005) including *E. coli*, especially O157:H7, where an association between PFGE profile and isolate source were proposed (Bender *et al.* 1997; Parveen *et al.* 2001).

1.7.2.2. Polymerase chain reaction based methods

Since its development in the mid 80s by Mullis and his colleagues (Saiki *et al.* 1985), numerous PCR based methods have been introduced for epidemiological purposes,

including restriction fragment length polymorphism PCR (RFLP-PCR), ribotyping PCR, arbitrary primed PCR (AP-PCR/RAPD) and enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR) (Power 1996).

However, development of quantitative real time PCR provided a great tool for epidemiological research and routine diagnosis. Recently, Birkett *et al.* (2007) developed a new multiplex real time taqMan PCR assay for rapid detection and genotyping of CTX-M ESBL producing enteric bacteria, which could be useful for local epidemiological investigations (Birkett *et al.* 2007).

1.7.2.2.1. Random amplification of polymorphic DNA (RAPD)

Different terms have been used to describe these methods including arbitrary primer – PCR or random amplification of polymorphic DNA or multiple arbitrary amplification profiling (MAAP). In these techniques, short primers of non specific sequence (around 10 bases) are used to initiate random amplification of the targeted DNA. The number and the location of these priming sites vary for different strains based on the bacterial genome sequence and the resulting electrophoresed DNA profile is strain specific (Olive & Bean 1999; Power 1996).

A number of studies have successfully use RAPD assays in typing *E. coli* (Ensor *et al.* 2006; Vidovic, Germida & Korber 2007). Furthermore, in a study by Garcia-Martinez *et al.* (1996), characterizing a collection of 74 UPEC isolates obtained from three hospitals located in geographically distant towns in Spain and some reference strains.RAPD method showed a high discriminatory power and was able to cluster most of the tested isolates in to two major groups.

The principle advantage of RAPD is the lack of a requirement for prior knowledge of the target DNA sequence, but several limitation have been associated with it including

limited reproducibility and difficulty comparing profiles with others from different laboratories (Power 1996).

1.7.2.2.2. Enterobacterial repetitive intergenic consensus (ERIC-PCR)

Several highly conserved repeated sequence sets were identified in *E. coli*, including repetitive extragenic palindromic (REP) and enterobacterial repetitive intergenic consensus (ERIC) sequences (Hulton, Higgins & Sharp 1991).

The ERIC-PCR method uses these specific intergenic repeated sequences found in the family *Enterobacteriaceae* as primer sites to amplify the regions between them. The number and location of these sequences varies from strain to strain and the electrophoretically resolved amplified fragment will form a distinct DNA fingerprint (Power 1996).

In a comparison study of prevalence of colonization with UPEC, carried out by Johnson *et al.* (1998), ERIC-PCR appeared as discriminative as commonly used typing methods but was less time consuming. In another study, Manges *et al.* (2001) successfully used ERIC-PCR to screen a collection of UPEC for clonal group A strains. Using a similar technique with another set of specific repeated sequences (BOX A1R), Johnson *et al.* (1997) were able to distinguish strains of the J96-like clonal group among O4:H5 UPEC, allowing rapid identification of members of this clone, which is responsible for significant infections in human.

Lack of reproducibility was commonly reported, which could be explained by the low stringency annealing conditions used (Chulain, Morris & Cormican 2006; Hopkins & Hilton 2000). However, an automated format was developed using separation of fluorescently labelled fragments with a DNA sequencer to enhance consistency and allow creation of a portable database to enable intra-laboratories comparison (Olive &

Bean 1999). More recently, the DiversiLab System, which is a rapid reproducible automated fingerprinting system using repetitive based sequence PCR (rep-PCR) was successfully used as an epidemiological tool to investigate different microorganisms including UPEC (Kilic *et al.* 2010; Lau *et al.* 2010; Ratkai *et al.* 2010)

1.7.2.2.3. Restriction fragment length polymorphisms RFLP and PCR-RFLP

In these approaches, the DNA sequence is subjected to restriction by specific endonuclease enzyme resulting in strain specific patterns of DNA fragments of different size and number (Foley *et al.* 2009). RFLP has widely been used in typing food borne pathogens such as *Salmonella* (Jordan *et al.* 2009; Paiva *et al.*), *Campylobacter* species, (Ayling *et al.* 1996) and *E. coli* (Arthur *et al.* 1990).

As a variation of the original RFLP, PCR-RFLP involves amplification of specific sequences in the bacteria and digestion of the PCR amplicons with specific endonucleases to generate strain specific banding patterns (Foley *et al.* 2009). Studies using PCR-RFLP cover a wide range of application from population structure determination (Shima *et al.* 2006) to detection of various characteristics such as antibiotic resistance (Jones *et al.* 2008).

1.7.2.2.4. Ribotyping

Ribotyping relies on differences in the location and number of ribosomal RNA (rRNA) gene sequences present in the bacterial genome. In a similar way to RFLP, ribotyping involves restriction of bacterial DNA with specific endonucleases and the separation of the DNA fragments by gel electrophoresis followed by Southern blotting and probing with rRNA specific probes (Bouchet, Huot & Goldstein 2008).

Despite the low discriminatory power of ribotyping compared to other typing methods such as PFGE (Clermont *et al.* 2001; Hahm *et al.* 2003), examination of the literature

shows that ribotyping has been successfully used to evaluate the relatedness and clinical importance pathogens such as *E. coli* (Parveen *et al.* 1999; Tarkka, Ahman & Siitonen 1994), *Salmonella enterica* (Olsen *et al.* 1992) and *Vibrio cholerae* O1 (Popovic *et al.* 1993).

1.7.2.3. DNA Sequence based methods

1.7.2.3.1. Single nucleotide polymorphism (SNP) analysis

The increased number of fully sequenced bacterial genomes has facilitated identification of single nucleotide polymorphisms (SNPs), which provides a reliable epidemiological tool with high discriminatory power. As bacterial species evolve, changes take place as a result of point mutations or horizontal transfer of genetic information, resulting in strains divergence. SNP analysis takes advantage of these changes at multiple loci to differentiate isolates.

Several methods have been used to detect polymorphisms at defined SNP locations including sequencing of the region, mass spectrometry, which detects mass differences in the target sequence due to difference in the mass of various nucleotides, real-time PCR using hybridization probes where the variation in nucleotides affect the efficiency of probe binding and other methods including microarray, RFLP and flow cytometery (Foley *et al.* 2009).

SNP analysis has been used successfully in characterization and differentiation of several pathogens including *Bacillus anthracis* (Pearson *et al.* 2004), *Mycobacterium tuberculosis* (Filliol *et al.* 2006), *Yersinia pestis* (Achtman *et al.* 2004), *Campylobacter jejuni* (Best *et al.* 2004) and *Salmonella* species including *Salmonella* Typhi (Ben-Darif *et al.*; Octavia & Lan 2007).

Identification of informative SNPs, specific for six major clonal complexes within discriminatory alleles of an MLST database, was used for rapid identification of *C. jejuni* (Best *et al.* 2004) and SNP analysis was used for development of an automated phylogenetic grouping technique for reconstruction of the relationships between 30 *E. coli* reference strains (Hommais *et al.* 2005).

Zhang and colleagues recognized 906 SNPs in 523 chromosomal genes for use in genotyping *E. coli* O157:H7 using comparative genome sequencing microarray techniques (Zhang *et al.* 2006).

1.7.2.3.2. Multilocus sequence typing

Most of the currently used genotyping methods rely on detection of genetic differences between tested populations with no consideration given to the frequency of genetic variation, which may accumulate rapidly. These methods are useful in the investigation of local or sporadic outbreaks.

However for long term or global epidemiological studies, methods that differentiate isolates based on genes that exhibit slowly accumulated genetic variation are required (Enright & Spratt 1999; Urwin & Maiden 2003). As mentioned above, MLEE, which has long been used in population genetic studies, differentiates isolates by the relative electrophoretic mobility of gene products. However, even though the profiles generated group strains in a similar manner to other techniques, the profiles are difficult to compare between laboratories (Maiden *et al.* 1998; Selander *et al.* 1986). Based on MLEE principles and overcoming this limitation, MLST benefited from the advances made in DNA sequencing and has become an important method for typing of epidemiologically important pathogenic strains (Enright & Spratt 1999; Lau *et al.* 2008a; Spratt 1999; Tartof *et al.* 2005).

1.7.2.3.3. Methodology

MLST directly detects the genetic variation of specific genes (housekeeping genes) that are present at different loci by sequencing of the genes. The housekeeping genes are characterized by not being subjected to any unusual selective forces, such as exposure to antimicrobial agents and exhibit neutral genomic variation that accumulates slowly.

Genetic variation of selected housekeeping genes is identified by nucleotide sequencing of 450-500 bp fragments of (usually) seven housekeeping genes. The length of the DNA fragment is selected to enable accurate sequencing of the genome fragment with single pair of primers and to provide a sufficient discriminatory power to allow differentiation of alleles within the population. In addition, the number of house keeping genes used affects the resolution power of the selected scheme (Figure 1.2).

For each gene, different sequences are given different allele numbers and the allele numbers of the seven housekeeping genes provides an allelic profile, which is assigned a sequence type (ST) that defines the genotype of each isolate (Figure 1.3).

The inter-isolate relationship can be defined by the closeness of the assigned ST, where closely related isolates have similar STs or slightly different ST and unrelated isolates have markedly different STs.

The resulting allelic profile is transferable and can be analysed and compared via an international epidemiological database using the internet (Enright & Spratt 1999; Spratt 1999; Urwin & Maiden 2003).

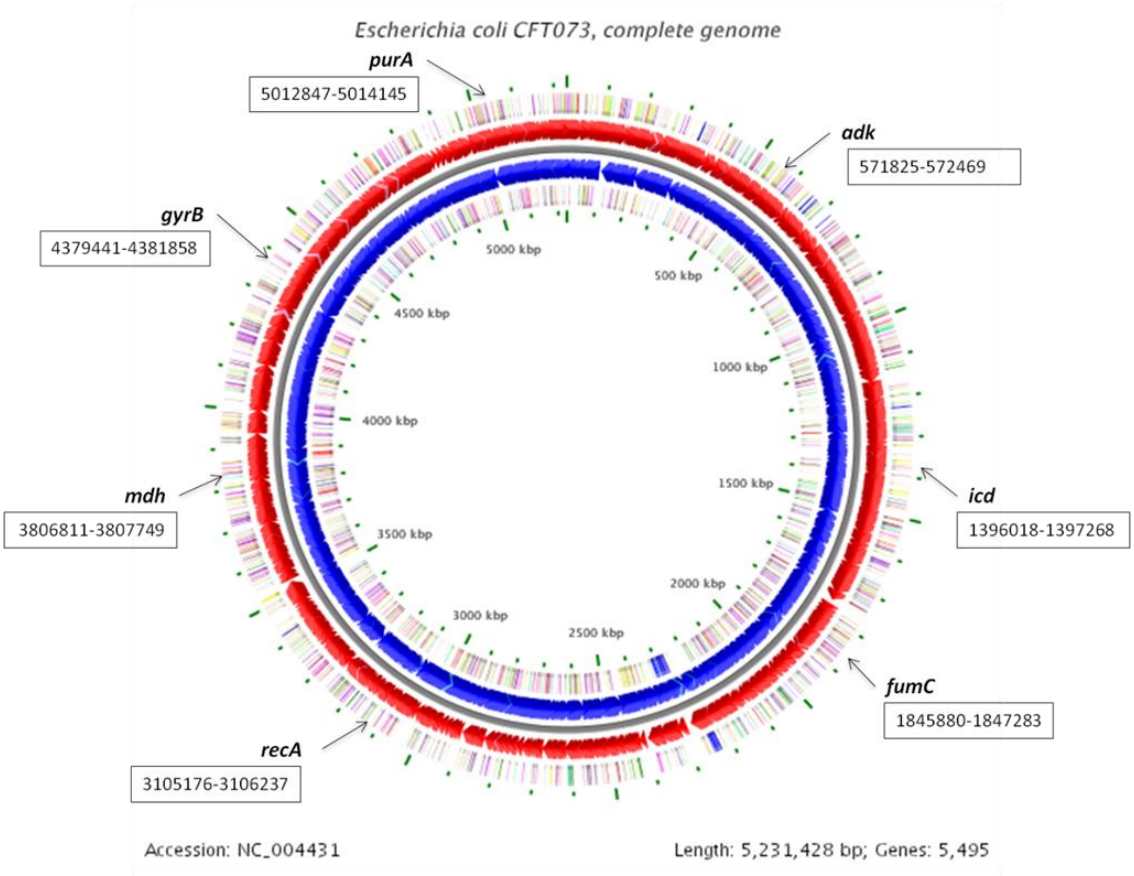


Figure 1. 2 *Escherichia coli* CFT073 complete genome showing the position of the seven house keeping genes used in the Achtman MLST protocol. Adapted from (http://www.ncbi.nlm.nih.gov)

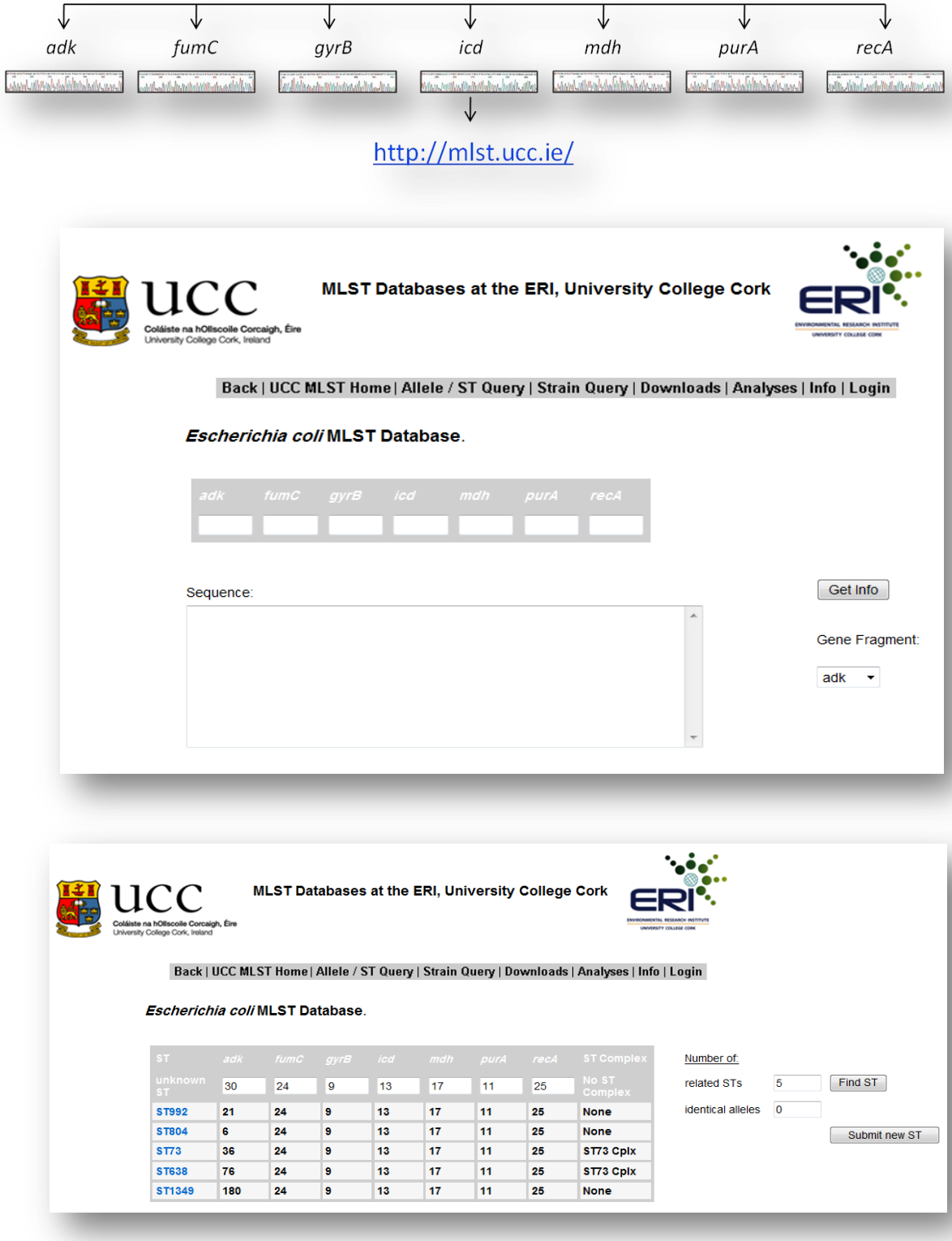


Figure1.3 Snapshots of the Achtman MLST database website illustrating the data submission page and ST assigning page

1.7.2.3.4. The E. coli MLST schemes

The success of an MLST protocol as a typing method greatly depends on the selection of the examined genes. In addition to being characterized by slow genomic variation the examined genes should reflect the epidemiological purpose of the assay and the selected genes should show no genetic relatedness (Adiri, Gophna & Ron 2003; Nemoy *et al.* 2005; Tartof *et al.* 2005). Therefore different MLST schemes were developed to suit different groups of *E. coli* using different housekeeping genes.(Adiri *et al.* 2003; Beutin *et al.* 2005; Noller *et al.* 2003).

In their comparison study Tartof and his colleagues (2005) evaluated a standardized MLST protocol (accessible through the MAX-PLANCK institute) for the typing of UPEC isolates and compare it with PFGE and ERIC2 assays. The study reveals that the MLST used scheme provides high discriminatory power as ERIC2, furthermore it was able to distinguish human clonal group A strains from other (species/nonhuman) sourced clonal group A strains but it failed to supersede PFGE assay in the differentiation of UPEC clonal group A. However, a study by Nemoy *et al.* (2005) revealed that the MLST could offer a better discriminatory power than PFGE if the scheme used is supported by additional genes such as antimicrobial resistance genes (Nemoy *et al.* 2005)

Three distinct MLST schemes associated with discriminative databases are available through three centres. Institut Pasteur's MLST scheme (www.pasteur.fr/mlst); M. Achtman's MLST scheme (www.mlst.ucc.ie) supported by University Collage Cork, Ireland and T. Whittam's MLST scheme (www.shigatox.net) supported by the Shigatoxin *E. coli* Centre based at National food safety and Toxicology Centre, Michigan State University, USA., as shown in Table 1.2. Between these sites, a total of

18 housekeeping genes are targeted for strain discrimination. Based on citations, Achtman's scheme is the most widely used MLST scheme for typing *E. coli* isolates.

Table 1.2 Housekeeping genes supported by internet assessable discriminatory databases

Genes	Gene product	Pasteur's	Achtman's	Whittam's
adk	Adenylate kinase		✓	
fumC	Fumarate hydratase		✓	
gyrB	DNA gyrase		✓	
icd	Isocitrate/isopropylmalate dehydrogenase	*icdA*	✓	*icdA*
mdh	Malate dehydrogenase	✓	✓	✓
purA	Adenylosuccinate dehydrogenase		✓	
recA	ATP/GTP binding motif		✓	
aspC	Aspartate aminotransferase			✓
clpX	ATP-dependent Clp protease			✓
fadD	Cyl-CoA synthetase			✓
lysP	Lysine-specific permease			✓
uidA	Beta-D-glucuronidase	✓	✓	✓
dinB	DNA polymerase IV	✓		
papB	Pap operon regulatory protein	✓		
polB	DNA polymerase II	✓		
putP	proline:sodium symporter	✓		
trpA	tryptophan synthase, alpha subunit	✓		
trpB	tryptophan synthase, beta subunit	✓		

Collected from: http://www.Shigatoxin.net, http://mlst.ucc.ie and http://www.psteur.fr

1.7.2.3.5. Applications of MLST

Due to its high discriminatory power and the portability of the results, in addition to its contribution in the studies of evolution and population biology of organisms, MLST has been widely applied to study a variety of problems including the emergence of antibiotic resistant clones of pathogenic organisms (Nemoy *et al.* 2005; Reinert *et al.* 2005). MLST even shows potential to be used directly on clinical material, which may make MLST as essential a tool for routine clinical diagnosis as it is in epidemiological research (Diggle, Bell & Clarke 2003; Spratt 1999).

1.7.3. Phylogenetic analysis

With the rapid accumulation of genomic sequence data, phylogenetic analyses have evolved from being descriptive and speculative analysis of evolutionary relationships to a more mathematics based science, as a result of advances in mathematical models of sequence evolution and statistical evolutionary tools (Whelan, Lio & Goldman 2001).

For phylogenetic analysis of aligned sequences, all evolution analysis methods describe sequence evolution using phylogenetic trees. Methods used to construct phylogenetic trees fall into two categories: distance methods; and discrete methods.

1.7.3.1. Distance methods

These involve the calculation of matrices of evolutionary distances between taxa on a pairwise basis. The evolutionary distance used for this purpose is usually an estimation of the number of nucleotide substitutions per site (Nei 1996).

There are several distance methods for construction of phylogenetic trees, the most commonly used ones being the unweighted pair group method with arithmetic means (UPGMA) and neighbour joining (NJ) methods.

UPGMA is the simplest method of tree construction and was originally developed to construct taxonomic phenograms, which are trees that reflect the phenotypic similarities between taxa. It uses a sequential clustering algorithm that starts by grouping two taxa with the most similarity and then progressively adding less similar taxa to the groups. It is generally not considered a good algorithm for construction of phylogenetic trees, as it assumes that the evolution rates of different lineages are approximately equal. Although this is not the case in bacterial population biology, UPGMA may prove useful as a quick guide to identifying similar isolates (Morrison 1996).

NJ is based on the minimum-evolution criterion to re-construct a phylogenetic tree from evolutionary distance data i.e. the topology that gives the least total branch length is preferred at each step of the algorithm. Unlike UPGMA constructed trees, NJ does not assume that all lineages evolve at the same rate and it is largely recommended to analyze populations where lineage evolution rate varies. However, as the NJ algorithm seeks to represent the data in the form of an additive tree, NJ may not find the true tree topology with least total branch length. Even though it is sub-optimal in this sense, the reliability of the NJ tree obtained can be extensively tested using bootstrap analysis, which usually finds a tree that is quite close to the optimal tree (Saitou & Nei 1987).

1.7.3.2. Discrete (character-based) methods

Discrete methods consider each nucleotide site directly, rather than on pairwise distances, using optimality criteria to choose among the set of all possible trees. The optimality criteria give each tree a score that is based on the comparison of the tree to other possible trees according to used criteria (Nei 1996).

The commonly used distance methods are the maximum parsimony (MP) and the maximum likelihood (ML) methods (Whelan *et al.* 2001).

ML is one of the standard tools of statistics. In phylogenetic analysis, it evaluates the probability that the chosen evolutionary model will have generated the observed sequences. The preferred trees are those that yield the highest likelihood over the whole tree. It involves two steps: building mathematical models taking in account all possible nucleotide substitution configurations; and then finding the tree that is most likely to have produced the observed data. Because it analyzes each nucleotide position of the multiple alignment, maximum likelihood is CPU consuming and, thus, extremely slow (Felsenstein 1981).

The MP method attempts to determine a phylogenetic tree that requires the smallest number of evolutionary changes to explain the variation in a given set of data. In construction of MP trees, only informative or substitution sites are considered and the sum of the minimum possible substitutions over all sites is called the tree length. The tree with the minimum length is selected as the tree of maximum parsimony (Whelan *et al.* 2001).

1.8. Aims of the project

The spread of highly virulent strains of UPEC such as *E. coli* CGA and the emergence of unique antimicrobial resistance profiles among the UPEC population, provide evidence of epidemic clonal dissemination. To determine the epidemiological significance of these pathogens, a full understanding of the population biology is required.

The overall aims of this thesis were to study the population biology of UPEC in the North West of England with a view to understanding the most significant lineages causing UTI and to inform development of rapid assays to allow identification of members of these lineages. To achieve these overall aims, the following objectives were set:

- Generate a collection (n≥300) of UPEC isolates that meet a selection criterion of identifying *E. coli* causing UTI used by the UK Health Protection Agency (Anonymous 2004).
- Identify successful clones and clonal complexes within the strain collection using multilocus sequence typing (MLST).
- Assess the phylogeny and the microevolution of clones and clonal complexes that constitute the UPEC population and examine their phylogenetic relationships with other *E. coli* pathotypes using different molecular phylogenetic analysis methods and online MLST database resources.
- To further examine isolates of identified clones and clonal complexes using phenotypic and genotypic methods and measure the possible association of key UPEC clones with different pheno- and genotypic profiles.

Materials and Methods

2. Materials and Methods

2.1. Uropathogenic *E. coli* strain selection

A total of 300 isolates of UPEC were collected and, in order to minimize the possibility of collecting non-UPEC, the selection was limited to urine isolates that showed pure culture with a significant total count, accompanied with significant pyuria according to the national standard method (BSOP 41) Health Protection Agency (Anonymous 2004). Urine samples were plated onto CLED agar plates (Oxoid, Ltd, UK), which were incubated at 37^{o}C for 18 hrs in a normal atmosphere.

Of the 300 non-replicated isolates collected, 150 were recovered in June 2007 in the Manchester area (100 isolates) and Preston area (50 isolates), another 150 isolates being collected within the same period of 2009 and with the same geographic representation.

2.2. Isolate collection

The collection was carried in two settings, one in the microbiology laboratory at the Central Manchester and Manchester Children's University Hospital (MRI), which served both the MRI and the surrounding general practitioners in Manchester, and the other in the microbiology laboratory at the Preston Royal Hospital.

To maximize the chances of collecting *E. coli*, all lactose fermenting strains were further examined using chromogenic media CPS3 (Biomérieux), as per manufacturer's directions, on which *E. coli* are identified based on the ability to produce β – glucuronidase.

2.3. Identification and antibiotic susceptibility

The Vitek 2 compact automated system (Biomérieux) was used for the identification and the antibiotic susceptibility testing of the collected isolates. By using the Vitek 2 ID-GNB card (Biomérieux), identification of Gram negative bacilli occurs through testing the organism's metabolic activity in 41 fluorescent biochemical tests including 18 enzymatic tests, 18 fermentation tests, two decarboxylase tests and three other miscellaneous tests.

Antibiotic susceptibility testing is based on kinetic analysis of the bacterial growth in the presence of selected antibiotics (20 antibiotics representing all antibiotic families) and the antibiotic susceptibility profile is then analysed in order to predict the underlying resistance mechanisms present in each isolate.

The antibiotic panel (Vitek 2 AST-054) was selected, as it covers the commonly used antibiotics for UTI and it is one of the standard antibiotic panels used in the UK. It consists of the following antibiotics: Ampicillin, Amoxicillin/Clavulanic acid, Piperacillin, Piperacillin/Tazobactam, Cefalotin, Cefuroxime, Cefuroxime axeil, Cefoxitin, Cefotaxime, Ceftazidime, Cefepime, Aztreonam, Meropenem, Ertapenem, Trimethoprim, Tobramycin, Amikacin, Gentamicin, Nalidixic acid, Ciprofloxacin and Nitrofurantoin.

Isolates that were previously identified as *E. coli* using CPS3 chromogenic media were cultured on Nutrient agar (Oxoid) and incubated for 18-24 hrs at 37ºC. The identification and antimicrobial susceptibility testing was performed following the manufacturer's instructions, and the resulting biochemical activity and antibiotic susceptibility data were recorded in spreadsheets for subsequent analysis.

The fully identified UPEC strains were stored using the microbank system (ProLab Diagnostics, UK) in duplicate (-20ºC and -70ºC).

2.4. DNA extraction

Prior to DNA extraction, strains were cultured on Columbia agar and MacConkey agar plates (Oxoid) at 37ºC for 18 hrs and pure single colonies from Columbia agar plates were used for extraction. Genomic DNA extraction was carried out using PrepMan Ultra sample preparation reagent (Applied Biosystems, USA) following the manufacturer's instructions. Extracted DNA was stored in a freezer at -20ºC, ready to be used for PCR.

2.5. Multilocus sequence typing of UPEC

The previously collected DNA extracts from the 300 UPEC isolates were genotyped using the MLST scheme developed by Achtman and colleagues (Wirth *et al.* 2006).

2.5.1. Selection of the targeted loci

The success of MLST protocol as a typing method greatly depends on the selection of the examined genes. From the few schemes available that are supported by an international database, Achtman's scheme (http://mlst.ucc.ie/mlst/dbs/Ecoli/) provides a high discriminatory power, capable of clustering populations in to different clonal groups and is one of the more widely used schemes.

Seven genes were selected according to the Achtman scheme, *adk:* adenylate kinase; *fumC:* Fumarate hydratase; *gyrB:* DNA gyrase; *icd:* Isocitrate/isopropylmalate

dehydrogenase; *mdh:* Malate dehyrogenase; *purA:* adenylosuccinate; *recA:* ATP/GTP binding motif.

2.5.2. PCR amplification of the targeted genes

Internal fragments of the seven house-keeping genes, which ranged between 583bp to 932bp, were amplified using pairs of primers listed in Table 2.1.

The amplification reactions were carried out in 50 µl volumes containing 5 µl of each primer (10pmol/µl), 5µl of PCR buffer (x10) (Qiagen, Sussex, UK), 10 µl of 1mM deoxynucleoside triphosphates (dNTPs) (Roche, GmbH, Germany) and 0.25 units of Taq DNA polymerase (Qiagen) and 1 µl of chromosomal DNA.

Amplification conditions were: 2 min at 94 ºC, followed by 35 cycles of 1 min at 94 ºC, 1 min at annealing temperature (as listed in Table 2.1), 1 min at 72 ºC followed by 5 min at 72 ºC and finally the reactions were held at 4 ºC. Amplification was carried out using an Eppendorf thermal cycler (Mastercycler; Helena Biosciences, Gateshead, UK).

Table 2.1 PCR primers for the seven selected house keeping genes used in MLST

Gene	Primer Sequence (5' – 3')	Annealing temp. (℃)	Size of product (bp)
adk	*F*'-ATTCTGCTTGGCGCTCCGGG *R* –CCGTCAACTTTCGCGTATTT	60	583
fumC	*F* -TCACAGGTCGCCAGCGCTTC *R* –GTACGCAGCGAAAAAGATTC	63	806
gyrB	*F* –TCGGCGACACGGATGACGGC *R*-ATCAGGCCTTCACGCGCATC	60	911
icd	*F* -ATGGAAAGTAAAGTAGTTGTTCCGGCACA *R* –GGACGCAGCAGGATCTGTT	53	878
mdh	*F* –ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG *R* -TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT	53	932
purA	*F* -CGCGCTGATGAAAGAGATGA *R* –CATACGGTAAGCCACGCAGA	67	816
recA	*F* -CGCATTCGCTTTACCCTGACC *R* –TCGTCGAAATCTACGGACCGGA	60	780

2.5.3. Detection of the PCR product

The presence and the purity of the PCR product was confirmed electrophoretically by mixing 5µl of each PCR product with 5µl of loading buffer [70% (w/v) sterile injectable water (Phoenix Pharmaceuticals, Gloucester, UK), 29.5% (w/v) Glycerol (Sigma) and 0.5% (w/v) Bromophenol Blue (Bio-Rad, Hertfordshire, UK)] and passing the PCR product mixture through 2% (w/v) of agarose (Promega Corporation, Madison, USA) in Tris-borate-EDTA (TBE) (Sigma Dorset, UK) incorporating 15µl of SYBR safe gel stain (Invitrogen, Paisley, UK) at 132 V for 20 min followed by visualization of the bands under UV transillumination (Gel Doc 1000, Bio-Rad Laboratories).

2.5.4. Post PCR purification

The Whatman vacuum filtration system (Whatman, UK) was used to clean the PCR product of unwanted components in the reaction mixture. As recommended by the manufacturer, each sample was treated with 80 µl of binding buffer (4M Guanidine-HCL, 0.75 M Potassium Acetate, pH 406) before being passed through the DNA binding plate under vacuum (-15 and -20 inches Hg), followed by washing the captured DNA with 400 µl of washing buffer (mixture of 40% (v/v) 100mM Tris, 20 mM EDTA, 0.4M NaCl. pH 7.5 and 60% (v/v) ethanol). DNA was then recovered by adding 50µl of elution buffer (10mM Tris, 0.1 mM EDTA, pH 7.5). The purified DNA product was stored at -20 ºC.

2.5.4.1. Sequencing of amplified DNA fragments

SEQ Dye terminator cycle sequencing (DTCS) Quick start kits (Beckman Coulter Inc.) were used for sequencing of the target DNA fragments. As recommended by the manufacturer, the sequencing reactions were carried out in 10 µl volumes, each reaction

containing 2µl of DTCS Quick start master mix (Genome Lab^{TM} DTCS- Quick Start Kit, Beckman Coulter), 1µl of genetix buffer (half CEQ^{TM} Genetix, UK), 0.5µl of primer (10 pmol/µl), as shown in Table 2.2, 0.5µl of purified DNA template and DNA free water to adjust the volume to 10µl.

The sequencing reaction conditions were 40 consecutive cycles of 96 ºC for 20 seconds, 50 ºC for 20 seconds and 60 ºC for 4 minutes, finally held at 4 ºC and were carried out using an Eppendorf thermal cycler (Mastercycler; Helena Biosciences).

Table 2.2 Sequencing primers for analysis of the target gene fragments in MLST

Gene	Primer Sequence
adk	F 5'- GCAATGCGTATCATTCTGCT-3' R 5'- CAGATCAGCGCGAACTTCAG-3'
fumC	F 5'- CCACCTCACTGATTCATGCG-3' R 5'- CGGTGCACAGGTAATGACTG-3'
gyrB	F 5'- CGGGTCACTGTAAAGAAATTATCG-3' R 5'- GTCCATGTAGGCGTTCAGGG-3'
icd	F 5'- TACATTGAAGGTGATGGAATCG-3' R 5'- GTCTTTAAACGCTCCTTCGG-3'
mdh	F 5'- TCTGAGCCATATCCCTACTG-3' R 5'- CGATAGATTTACGCTCTTCCA-3'
purA	F 5'- CTGCTGTCTGAAGCATGTCC-3' R 5'- CAGTTTAGTCAGGCAGAAGC-3'
recA	F 5'- AGCGTGAAGGTAAAACCTGTG-3' R 5'- ACCTTTGTAGCTGTACCACG-3'

2.5.5. Post sequencing purification

To remove unincorporated primers and master mix residues, the ethanol plate precipitation method was used, as recommended by the manufacturer of the sequencing kit (Beckman Coulter). Freshly prepared stop solution/glycogen mixture was prepared as follows. For each sequencing reaction: 1µl of 3M sodium Acetate (pH 5.2) (Sigma), 1µl of 100mM Na2-EDTA (pH 8.0) (Sigma) and 0.5 µl of 20 mg/ml of glycogen (Genetix). 2.5 µl of stop solution and 30 µl of ice cold 95% (v/v) Ethanol/H_2O (kept at -

20ºC) were added to each sequencing reaction and plates were sealed with an aluminium foil lid (Beckman Coulter). The mixture was mixed by inverting the plate 15-20 times vigorously and immediately centrifuged at 14,000 rpm at 4 ºC for 30 minutes using a Beckman Coulter Allegro^{TM} 21R centrifuge (Beckman Coulter Inc.).

After centrifugation, the alcohol was removed by spinning the plate, after inverting it on folded paper towels, at 300 rpm for 2 seconds. The pellets, which formed at the bottom of the wells, were then rinsed twice with 100 µl of ice cold 75% (v/v) Ethanol/H_2O. After each rinse, the plate was immediately centrifuged at 14,000 rpm for 5 minutes at 4 ºC. After centrifugation the supernatant was carefully removed by gently inverting the plate and as mentioned above the excess alcohol was removed by spinning the inverted plate at 300 rpm for 2 seconds.

Finally, the plate was vacuum dried for 20 minutes, and the pellets were suspended in 50 µl of the Sample Loading Solution (supplied with the kit) and each sample was overlaid with one drop of mineral oil (supplied with the kit). The plate was then kept at -20 ºC ready for sequence analysis.

2.5.6. Sequence determination and analysis

The sequences of the reaction products were then determined using a capillary electrophoresis sequencer (CEQ 8000 BECKMAN COULTER). Assembly and analysis of each target gene fragment was performed using Sequencher^{TM} v.4.0 software (Gene Codes Corporation). The sequence fragments were trimmed to a uniform length that corresponded with the region used to identify the target, by reference to allele sequences recovered from the Achtman MLST protocol published on the http://mlst.ucc.ie/mlst/dbs/Ecoli website.

2.5.7. Allele and sequence type assignment

The sequences of target genes fragments were then assigned distinct allele numbers by submitting them to the MLST *E. coli* data base website (http://mlst.ucc.ie/mlst/dbs/Ecoli).

Finally, a Sequence Type (ST) and its associated Sequence Type Complex were assigned for each isolate by comparing its allelic profile with that available in the MLST database. In the case of new alleles being identified, a unique allele number and, subsequently, a new ST was issued by the curator of the database, following examination of the DNA sequence trace files for the respective isolate.

2.6. Phylogenetic analysis

2.6.1. Characteristics of housekeeping genes

The statistical tools implemented in the START v2 program (Jolley *et al.* 2001) were used to describe the molecular characteristics of housekeeping genes. The total number of polymorphic sites and the ratio (*dS/dN*) of synonymous substitutions to non-synonymous substitutions per nucleotide for all target gene fragments were estimated. The *dN/dS* ratio is used to assess the degree of selective pressure operating on target genes. An excess of non-synonymous (*dN*) over synonymous (*dS*) substitutions is indication of positive selection indicating that non-synonymous changes are being fixed faster than they occur by mutation. A *dN/dS* <1 indicates negative selection in which non-synonymous changes are usually deleterious and therefore selected out of the population.

2.6.1.1. Index of association

The extent of linkage equilibrium or clonality among the defined population was estimated by calculating the standardized index of association (I^S_A) values (Haubold & Hudson 2000) as implemented in the START v2 program. Within a monophyletic population an I_A significantly different than Zero in 1000 randomized trials would suggest linkage disequilibrium indicating that the population has a clonal structure, while I_A values not significantly different than Zero indicate linkage equilibrium and the population is considered to be freely recombining.

2.6.1.2. Simpson's index of diversity

The Simpson's Index of Diversity (Simpson 1949) was used to determine the level of diversity within each house keeping gene in the studied population compared to ExPEC and other *E. coli* populations reported in the database. The following equation was used to measure the level of diversity

$$SID = 1 - \frac{\sum n(n-1)}{\mathrm{N}(\mathrm{N}-1)}$$

Where N represents the total number of different STs in the population and n is the total number of STs that have the same allele type.

2.6.2. eBURST

The Based Upon Related Sequence Types (BURST) clustering algorithm was originally designed by Dr. Ed Feil and his colleagues in Imperial College London and upgraded in 2004 when it was entitled eBURST. eBURST is available freely online (http://eburst.mlst.net/). The algorithm attempts to cluster STs into clonal complexes (CC) based on the number of shared gene loci and further predict a founder for each CC based on the abundance of STs that are linked to a particular ST by a single allelic difference. The algorithm draws conclusions about the relatedness of individual STs, describing the ST as single locus variant (SLV), double locus variant (DLV) or triple locus variant (TLV) in relation to the founder ST and to each other. Additionally, eBURST was able to estimate the relative age of a CC by how many SLVs it has and the presence of related subgroups (Feil *et al.* 2004).

2.6.3. UPGMA (unweighted pair group method with arithmetic mean) cluster analysis

As it uses the assumption of an equal evolutionary rate of change in all lineages, this method is not considered a good algorithm for construction of Phylogenetic trees. However, for MLST, in which the target genes were carefully selected to exhibit slow and steady mutation rates, UPGMA can provide a quick guide to identify similarity between STs. Using the START package, UPGMA was performed on allelic profiles of the dataset rather than the nucleotide sequences, which provides additional credibility to the phylogenetic inference obtained.

2.6.4. Neighbour-Joining (NJ) cluster method

The NJ construction tool available in MEGA 4.11.1 software (www.megasoftware.net) was used to draw phylogenetic trees for concatenated sequences of the seven loci and for sequences of each locus alone using the Jukes-Cantor based algorithm and bootstrapping using 1000 data sets for each analysis.

2.6.5. Clonalframe

Clonalframe is a computer package used for the inference of bacterial microevolution using multilocus sequence data and was developed in 2007 by Didelot and Falush and is available freely from http://bacteria.stats.ox.ac.uk/ (Didelot & Falush 2007). It is based on a coalescent approach assuming that all isolates in a dataset had equal chances of reproduction in a constant sized population. As it takes into account both point mutation and homologous recombination, the method provides information on potential recombination sites and, therefore, calculates genealogies after ignoring these recombination events. A consensus tree based on the posterior probabilities of genealogies is generated, showing clonal relationships among the strains and attempting to suggest ancestral STs not found amongst the study population.

Ten Clonalframe runs were computed on unique STs from the dataset, each with 100,000 iterations after 10,000 burn-in iterations. 50% consensus trees and network representations of the analysis were generated using the graphical user interface of the program.

2.6.6. Neighbour-net

Neighbour-net is a distance-based method that constructs phylogenetic networks using nucleotide sequence data. The neighbour-net algorithm is an extension of the neighbor-joining algorithm in which an agglomerative process of pairing nodes is followed by collection of weighted splits that are converted into a split network rather than phylogenetic tree. This has the advantage of resolving conflicting signals generated by evolutionary processes, such as lateral transfer of genes and recombination within a population (Bryant & Moulton 2004).

Concatenated sequences of the seven loci for all STs in the dataset were analyzed to generate a phylogenetic network using the Neighbour-net algorithm implemented in SplitsTree 4.11 (www.splitstree.org) (Huson & Bryant 2006).

2.6.7. Tests for recombination

In addition to index of association, which measures the clonality of a population by determining the contribution of recombination in shaping the population structure, nucleotide sequences of STs in dataset were analysed for evidence of recombination by several methods, each with its own specifications.

2.6.7.1. Maximum Chi-squared test

This is a method in which the sequences are examined in pairs for clusters of substitutions indicating a putative recombination event and to test whether the putative recombination sites are significant. Random trials are run where equal length sequences with the same number of polymorphic sites were created. The significance level was determined as the proportion of trial pairs with greater max X^2 values than were observed for strain data (Smith 1992).

All seven loci of the 100 unique STs in the dataset were tested for recombination using the max X^2 test included in the START v2 package. 1000 random pair comparison trials between allele sequences of each locus were run and putative recombination sites with p values <0.05 were considered significant.

2.6.7.2. Split decomposition

An initial search for recombination was conducted using the split decomposition algorithm, which is a parsimony method that represents conflicting phylogenetic signals in the data as parallel networks. Using the SplitsTree program v. 4.11, evidence for recombination in the concatenated dataset for each ST and for each house keeping gene was examined by split decomposition analysis with default settings (uncorrected P method) and 1000 bootstrap replicates.

2.6.7.3. Pairwise homoplasy index test

Based on the compatibilities between informative sites in a nucleotide sequence alignment, the pairwise homoplasy index test for recombination is able to distinguish recombination from frequent mutations. Using the pairwise homoplasy index test implemented by the SplitsTree program v.4.11, alleles of each house keeping gene were tested for recombination in addition to possible recombination within clonal complexes.

2.7. Triplex PCR for Phylogenetic grouping

This was performed based on methods described by Clermont and colleagues. A triplex PCR reaction was conducted to determine the phylogenetic group of each isolate in the data set (Clermont *et al.* 2000). Fragments of a combination of three DNA markers (*chuA, yjaA* and DNA fragment *tsp*E4C2) were amplified using pairs of primers as

listed in Table 2.3. The amplification reactions were carried out in 50 µl volumes containing 25 µl of BioMix™ Red (Bioline USA Inc.), 1 µl of each primer (10pmol/µl) and 2 µl of 100 ng/µl chromosomal DNA.

Table 2.3 Primers used in Triplex PCR to detect phylogenetic groups

Gene	Primer Sequence (5' – 3')	Size of product (bp)
chuA	F-GACGAACCAACGGTCAGGAT R-TGCCGCCAGTACCAAAGACA	279
yjaA	F-TGAAGTGTCAGGAGACGCTG R-ATGGAGAATGCGTTCCTCAAC	211
tspE4C2	F-GAGTAATGTCGGGGCATTCA R-CGCGCCAACAAAGTATTACG	152

Amplification was carried out using the following conditions: 4min at 94 ºC; followed by 30 cycles of 5 sec at 94 ºC, 10 sec at 59 ºC, 10 sec at 72 ºC; followed by 5 min at 72 ºC and finally the reactions were held at 4 ºC.

The PCR products were analyzed electrophoretically by passing the PCR product through 1.5 % (w/v) agarose (Promega) in Tris-borate-EDTA (TBE) (Sigma) incorporating 15µl of SYBRsafe™ gel stain (Invitrogen) at 132 V for 20 min and visualisation under UV transillumination (Ge Doc 1000, Bio-Rad Laboratories)

Interpretation:

Band patterns on agarose gels were then analyzed to determine the phylogenetic group according to the dichotomous decision tree proposed by Clermont (Clermont *et al.* 2000), (Figures 2.1 & 2.2).

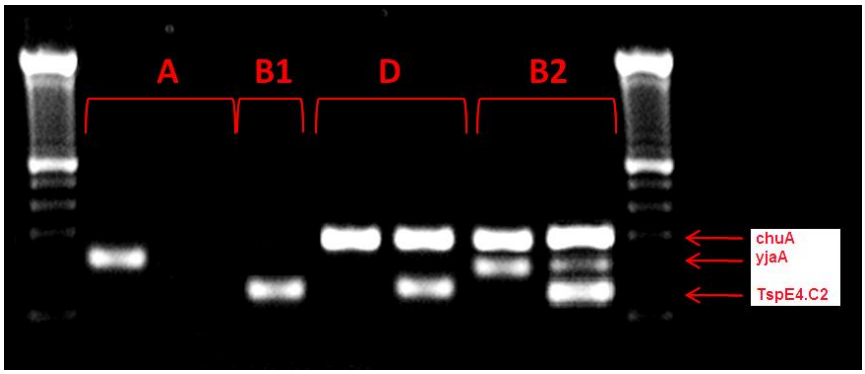


Figure 2.1 Triplex PCR profiles for each phylogenetic group

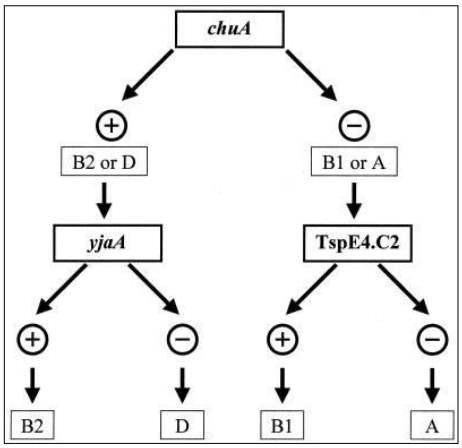


Figure 2.2 Dichotomous decision tree proposed by Clermont *et al* (2000)

2.8. Pulse-field gel electrophoresis (PFGE)

The genetic relatedness of 32 representative isolates of the common UPEC sequence types ST69 (n=10 isolates), ST73 (9) and ST131 (13), taken from different phylogenetic groups, was determined by *XbaI* PFGE analysis performed according to the PulseNet standardized PFGE protocol (Ribot *et al.* 2006). Generated profiles were compared digitally using BioNumerics v.3.5 (Applied Maths). A dendrogram was generated from

the Cluster analysis of Dice similarity indices based on UPGMA. Isolates were considered to belong to the same PFGE group if their Dice similarity index was ≥85%.

2.9. Virulence factor screening

The entire culture collection was screened for the presence of twenty-nine ExPEC associated VF genes encompassing five categories (adhesins, toxins, siderophores, capsule and "miscellaneous"). The adhesin genes investigated included *fimH* (mannose specific adhesin of type I fimbriae), P fimbriae elements (*papAH, papC, papEF* and *papG* alleles I, II, III, *sfaS* (S fimbrial adhesin), *focG* (the putative F1C fimbrial adhesin), *sfa/focDE* (central region of *sfaS* and *focG* operons), *afa/draBC* (Dr antigen-specific adhesin operons), *bmaE* (Blood group M-specific adhesion), *nfaE* (non-fimbrial adhesin), *gaf D* (glucosamine specific G fimbriae). Toxin genes screened were *cnf1* (cytotoxic necrotizing factor), *cdtB* (cytolethal distending toxin) and *hlyA* (α-haemolysin). The siderophore genes studied were *fyuA* (yersiniabactin) and *iutA* (aerobactin). Capsule synthesis associated genes screened were *kpsMT* groups (*II* and *III*) in addition to specifically targeting *K1* and *K5* genes of group II capsules. The miscellaneous VF genes were *cvaC* (colicin V; multifunctional serum resistance-associated plasmids), *traT* (serum resistance associated), *ibeA* (invasion of brain endothelium), and *PAI*, a coding region of unknown significant of a sequenced PAI from archetypal ExPEC strain CFT073 was used as a generic marker for uropathogenic PAIs..

VF profiles of the 300 isolates were determined using primers listed in Table 2.4 and an established multiplex PCR–based assay (Johnson & Stell 2000).

Table 2.4 Primers used in Triplex PCR to detect different VFs

Gene	Primer sequence* (5'-3')	Size of product (pb)
papAH	F-ATGGCAGTGGTGTCTTTTGGTG R-CGTCCCACCATACGTGCTCTTC	720
papC	F-GTGGCAGTATGAGTAATGACCGTTA R-ATATCCTTTCTGCAGGGATGCAATA	200
papEF	F-GCAACAGCAACGCTGGTTGCATCAT R-AGAGAGAGCCACTCTTATACGGACA	336
papG I, II,III ***papG II, III*** ***papG I***	F-CTGTAATTACGGAAGTGATTTCTG R-ACTATCCGGCTCCGGATAAACCAT R-TCCAGAAATAGCTCATGTAACCCG	 1070 1190
allele I	F-TCGTGCTCAGGTCCGGAATTT R-TGGCATCCCCCAACATTATCG	461
allele II	F-GGGATGAGCGGGCCTTTGAT R-CGGGCCCCCAAGTAACTCG	190
Allele III	F-GGCCTGCAATGGATTTACCTGG R-CCACCAAATGACCATGCCAGAC	258
Sfa/focDE	F-CTCCGGAGAACTGGGTGCATCTTAC R-CGGAGGAGTAATTACAAACCTGGCA	410
sfaS	F-GTGGATACGACGATTACTGTG R-CCGCCAGCATTCCCTGTATTC	240
focG	F-CAGCACAGGCAGTGGATACGA R-GAATGTCGCCTGCCCATTGCT	360
Afa/draBC	F-GGCAGAGGGCCGGCAACAGGC R- CCCGTAACGCGCCAGCATCTC	559
bmaE	F-ATGGCGCTAACTTGCCATGCTG R-AGGGGGACATATAGCCCCCTTC	507
gafD	F-TGTTGGACCGTCTCAGGGCTC R-CTCCCGGAACTCGCTGTTACT	952
nfaE	F-GCTTACTGATTCTGGGATGGA R-CGGTGGCCGAGTCATATGCCA	559
fimH	F-TGCAGAACGGATAAGCCGTGG R-GCAGTCACCTGCCCTCCGGTA	508
hlyA	F-AACAAGGATAAGCACTGTTCTGGCT R-ACCATATAAGCGGTCATTCCCGTCA	1177
cnfI	F-AAGATGGAGTTTCCTATGCAGGAG R-CATTCAGAGTCCTGCCCTCATTATT	498
cdtB	F-AAATCACCAAGAATCATCCAGTTA R-AAATCTCCTGCAATCATCCAGTTTA	430
fyuA	F-TGATTAACCCCGCGACGGGAA R-CGCAGTAGGCACGATGTTGTA	880
iutA	F-GGCTGGACATCATGGGAACTGG R-CGTCGGGAACGGGTAGAATCG	300
kpsMT-II	F-GCGCATTTGCTGATACTGTTG R-CATCCAGACGATAAGCATGAGCA	272
kpsMT-III	F-TCCTCTTGCTACTATTCCCCCT R-AGGCGTATCCATCCCTCCTAAC	392
kpsMT K1	F-TAGCAAACGTTCTATTGGTGC R-CATCCAGACGATAAGCATGAGCA	153
kpsMT K2	F-CAGTATCAGCAATCGTTCTGTA R-CATCCAGACGATAAGCATGAGCA	159
rfc	F-ATCCATCAGGAGGGGACTGGA R-AACCATACCAACCAATGCGAG	788
ibeA	F-AGGCAGGTGTGCGCCGCGTAC R-TGGTGCTCCGGCAAACCATGC	170
cvaC	F-CACACACAAACGGGAGCTGTT R-CTTCCCGCAGCATAGTTCCAT	680
traT	F-GGTGTGGTGCGATGAGCACAG R-CACGGTTCAGCCATCCCTGAG	290
PAI	F-GGACATCCTGTTACAGCGCGCA R-TCGCCACCAATCACAGCCGAAC	930

***(Johnson & Stell 2000)

2.10. Identification of isolates from the CTX-M-15 producing O25b ST131 clone

2.10.1. Screening for extended spectrum beta-lactamase enzyme type

ST131 isolates with ESBL phenotypes were screened for *bla*$_{CTX-M}$ alleles by PCR with universal primers MA1 and MA2 (Table 2.5). The amplification reactions were carried out in 50 µl volumes containing 5 µl of each primer (20pmol/µl), 5µl of PCR buffer (x10) (Qiagen), 10 µl of 1mM deoxynucleoside triphosphates (dNTPs) (Roche), 0.25 units of Taq DNA polymerase (Qiagen) and 2 µl of 100 ng/µl chromosomal DNA.

Amplification conditions were: Initial denaturation for 2 min at 94 ºC; followed by 35 cycles of 30 sec. at 94 ºC, 30 sec. at 55 ºC, 45 sec. at 72 ºC; followed by final elongation for 5 min at 72 ºC; and finally the reactions were held at 4 ºC and the resulting amplicons (size 554 bp) were defined by comparison to a standard 100bp DNA ladder (Invitrogen).

Determination of different *bla*$_{CTX-M}$ group was carried out using specific primers for each of the 5 groups (Table 2.5) following the amplification reaction conditions described above.

2.10.2. Detection of CTX-M-15 O25b ST131 clone

ST131 isolates identified as being positive for group 1 CTX-M were investigated for inclusion of the CTX-M-15 producing O25b clone. This was carried out by PCR using specific primers listed in Table 2.5. The amplification reactions were carried out in 50 µl volumes containing 25 µl of BioMix™ Red (Bioline USA Inc.), 1 µl of each primer (20pmol/µl) and 2 µl of 100 ng/µl chromosomal DNA.

Table 2.5 Primers used to screen for different CTX-M alleles and for the CTX-M-15 producing O25b clone

Target Description	Primer	Primer Sequence (5' – 3')	Size of product (bp)	*Reference
Universal CTX-M	MA	F-SCSATGTGCAGYACCAGTAA R-CCGCRATATGRTTGGTGGTG	554	1
CTX-M-G1	*bla*$_{CTX-M-1}$	F-AAAAATCACTGCGCCAGTTC R-AGCTTATTCATCGCCACGTT	415	2
CTX-M-G2	*bla*$_{CTX-M-2}$	F-CGACGCTACCCCTGCTATT R-CCAGCGTCAGATTTTTCAGG	552	2
CTX-M-G8	*bla*$_{CTX-M-8}$	F-GCACGATGACATTCGGG R-AACCCACGATGTGGGTAGC	666	2
CTX-M-G9	*bla*$_{CTX-M-9}$	F-CAAAGAGAGTGCAACGGATG R-ATTGGAAAGCGTTCATCACC	205	2
CTX-M-G25	*bla*$_{CTX-M-25}$	F-GCACGATGACATTCGGG R-AACCCACGATGTGGGTAGC	327	2
CTX-M-15	*bla*$_{CTX-M-15}$	F-ATAAAACCGGCAGCGGTG R-GAATTTTGACGATCGGGG	500	3
O25b	*rfb.1bis* *rfbO25b.r*	F-ATACCGACGACGCCGATCTG R-TGCTATTCATTATGCGCAGC	300	3

*1 (Saladin *et al.* 2002) ; 2 (Woodford, Fagan & Ellington 2006); 3 (Blanco *et al.* 2009); *Abbreviations: S: G or C, Y: C or T, R: A or G*

Amplification conditions were: Initial denaturation for 5 min at 94 ºC; followed by 30 cycles of 30 sec. at 94 ºC, 30 sec. at 55 ºC, 90 sec. at 72 ºC; followed by final elongation for 7 min at 72 ºC and finally the reactions were held at 4 ºC (Blanco *et al.* 2009). The produced amplicons (300 and 483bp) were identified using a 100bp DNA Ladder (Invitrogen) as a marker.

2.10.3. Detection of mechanisms of quinolone resistance in ST131 isolates

All quinolone resistant ST131 isolates were screened for the presence of different resistance mechanisms using published methods**.**

Mutations in QRDRs of *gyrA* and *parC* were detected as described by Yue and colleagues (Yue *et al.* 2008). The amplification reactions were carried out in 50 µl volumes containing 5 µl of each primer (10pmol/µl) (Table 2.6), 5µl of PCR buffer (x10) (Qiagen), 10 µl of 1mM deoxynucleoside triphosphates (dNTPs) (Roche) and 0.25 units of Taq DNA polymerase (Qiagen) and 1 µl of chromosomal DNA. The resulting amplicons were purified using a Whatman vacuum filtration system (Whatman) and were sequenced using a Dye terminator cycle sequencing (DTCS) Quick start kit (Beckman Coulter), as described in Section 2.5.4. The final assembly and analysis of each target gene fragment was carried out using Sequencher™ v.4.0 software (Gene Codes Corporation) and comparison with relevant native sequences for the *gyrA* (X06373) and *parC* genes (M58408) in the genebank™ database.

PCR-based Restriction Fragment Length Polymorphism (PCR-RFLP) assay was used to identify the *aac (6')-lb-cr* variants. PCR reactions were performed to amplify all *aac (6')-lb* sequences using primers and PCR condition described previously (Jones *et al.* 2008). The amplified DNA was purified as mentioned above (Section 2.5.4) and the cleaned amplicons were digested with *NdeI* and *FokI* restriction enzymes (Sigma) as described previously (Jones *et al.* 2008).

Screening for the three known *qnr* genes was carried out using multiplex PCR (Cattoir *et al.* 2007) and three strains were used as positive controls, two of which (*qnrA* and *qnrB*) were kindly provided by Dr. Neil Woodford (ARMRL, HPA, UK) and the *qnrS* positive strain was obtained from the National Collection of Type Cultures (NCTC) (HPA, UK). The expected product sizes for *qnrA, qnrB* and *qnrS* were 580, 264 and 428

bp, respectively (Figure 2.3). Amplification products were sized using a 100bp DNA Ladder (Invitrogen).

Table 2.6 Primers used in PCR reactions to detect different quinolone resistance mechanisms

Gene	Primer sequence (5'- 3')	Size of product (pb)	*Reference
gyrA	F- GGATAGCGGTTAGATGAGC R- CGTTCACCAGCAGGTTAGG	521	1
parC	F- AATGAGCGATATGGCAGAGC R- TTGGCAGACGGGCAGGTAG	376	1
qnrA	F- AGAGGATTTCTCACGCCAGG R- TGCCAGGCACAGATCTTGAC	580	2
qnrB	F- GGMATHGAAATTCGCCACTG R- TTTGCYGYYCGCCAGTCGAA	264	2
qnrS	F- GCAAGTTCATTGAACAGGGT R- TCTAAACCGTCGAGTTCGGCG	428	2
aac(6')-lb	F- ATGACTGAGCATGACCTTGC R- TTAGGCATCACTGCGTGTTC	519	3

*1 (Yue *et al.* 2008); 2 (Cattoir *et al.* 2007); 3 (Jones *et al.* 2008). *Abbreviations: M: A or C, Y: C or T, H: A or C or T.*

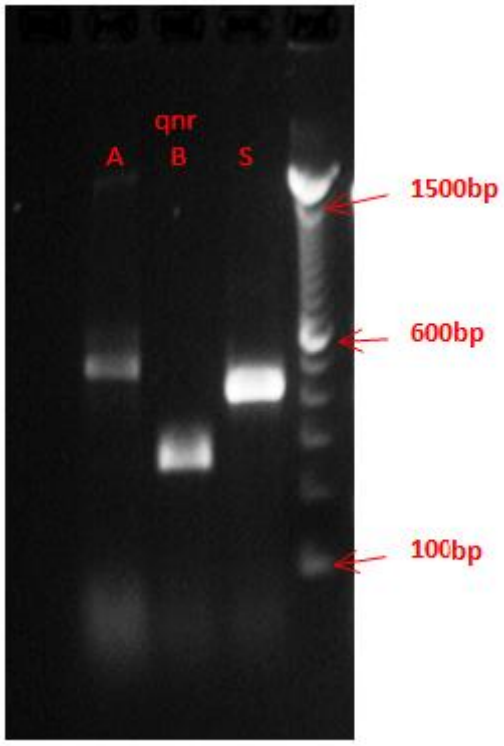


Figure 2.3 Three Positive *qnrA*, *qnrB* and *qnrS* controls subjected to electrophoresis in 1.5% agarose gels with a 100bp DNA marker (Invitrogen)

2.11. Statistical analysis

Statistical analysis was performed using SPSS software version 16.0 (SPSS Inc.) and the add-in Excel statistical tool software XLSTAT–Pro (Addinsoft Digital River GmbH. Germany).

Associations between different parameters were analysed by using Fisher's Exact or Mann-Whitney U test and the threshold for statistical significance was a *P* value of $\leq$ 0.05 and, where appropriate, more stringent criteria for statistical significance were used.

Aggregate scores for metabolic activities, resistance to different antimicrobial drugs and carriage of virulence factors were calculated for each isolate as the sum of all parameters for which the isolates tested positive divided by the total number of the tested parameters. Correlation analysis between biochemical scores, antimicrobial resistance scores and virulence scores were analysed using Pearson's correlation coefficients. Correlation analysis was also used to describe the association of different STs to the aggregate scores calculated for metabolic activities, resistance to different antimicrobial drugs and carriage of virulence factors and the significance criterion was a *P* value of ≤ 0.05.

Correspondence analysis was used to represent the interrelationships of the UPEC isolates according to their STs and on the basis of their phylogenetic groups and to describe the distribution of urovirulence traits among different phylogenetic types.

To assess the best combination of variables to identify ST131 and the CTX-M-15 producing O25b clone, a stepwise logistic multivariate regression analysis was applied with 95% confidence intervals (CIs).

Results

3. Results

Full data relating to the genotypic and phenotypic analyses of the isolates studied here are given in Appendices I to VI.

3.1. UPEC strains

A total of 300 *E. coli* isolates (200 from Manchester and 100 from Preston in the Northwest of England) were typed using MLST.

3.2. Allelic variation

Allelic characterisation of isolates in the study population was carried out using START v2. For the seven housekeeping gene fragments amplified and sequenced in the present study, a total of 195 different alleles were detected. These consisted of 187 (96%), that had been described previously and eight (4%) novel alleles. The number of alleles for each targeted gene ranged from 22 (*icd* and *mdh*) to 33 (*fumC* and *gyrB*) with an average of 28 per locus (Table 3.1).

Table 3.1 Allelic characterisation of UPEC housekeeping genes

Gene	Size of sequenced fragment (bp)	No. of alleles	No. of polymorphic sites (%)	d_N/d_S
adk	536	26	44 (8.2)	0.0056
fumC	469	33	76 (16.2)	0.0143
gyrB	460	33	33 (7.1)	0.0168
icd	518	22	37 (7.1)	0.0087
mdh	452	22	26 (5.7)	0.0126
purA	478	31	34 (7.1)	0.0154
recA	510	28	26 (5.0)	0.0000
mean	**489**	**28**	**39 (8.0)**	**0.0105**

The degree of sequence diversity within the housekeeping genes was relatively high with the proportion of polymorphic sites ranging from 5.0% (*recA*) to 16.2% (*fumC*) with an average of 8.0%. The ratio of non-synonymous to synonymous d_N/d_S changes for each gene locus were calculated using START v2, and all observed ratios were less than 1, indicative of negative or stabilising selection, with *gyrB* noted to have the highest d_N/d_S ratio (0.0168) whilst *recA* had the lowest ratio (0.000; Table 3.1).

To confirm that the population was not under focal positive selection leading to a flourish of allelic diversity, the allelic structure seen in the collected population was compared with that of STs reported in the international MLST database. Despite the relatively low Simpson's Index of Diversity (SID) detected for most enteric *E. coli* (EC) related target gene alleles, when compared to ExPEC all of the housekeeping gene fragments in the current dataset demonstrated diversity consistent with that of the entire ExPEC collection (Table 3.2).

Table 3.2 Diversity of NW-UPEC STs compared to that of ExPEC and EC in the international MLST database.

Gene	NW-UPEC (100)		International database			
			ExPEC (293)		EC (525)	
	No. of Alleles	SID	No. of Alleles	SID	No. of Alleles	SID
adk	26	0.89	53	0.90	102	0.84
fumC	33	0.94	68	0.96	117	0.89
gyrB	33	0.93	64	0.94	101	0.93
icd	22	0.90	49	0.93	103	0.93
mdh	22	0.89	40	0.92	78	0.92
purA	31	0.91	49	0.91	87	0.84
recA	28	0.94	43	0.93	70	0.86

SID: Simpson's index of diversity

3.3. Allele specificity

The majority of the alleles identified in the present study were found in both ExPEC and EC databases (Figure 3.1). The incidence of each allele of the seven housekeeping gene fragments differed little between our collection (195 alleles) and the ExPEC database (366 alleles) with only three alleles (fumC39, mdh8, mdh17) showing significant differences in their frequency of incidence. On the other hand when compared with the 658 alleles in the EC database, 49 alleles were significantly associated with either the NW-UPEC collection or the EC database (Table 3.3).

3.4. Sequence types and clonal complexes

From 300 UPEC isolates tested, a total of 100 sequence types (STs) and 17 clonal complexes (CCs) were identified, CCs accounted for 54.7% of the tested population (Figure 3.2). All of the 17 CCs reported here were previously described and available in the international MLST database (http://mlst.ucc.ie). A total of 74 STs in the dataset were not able to be assigned to a CC. Forty-four STs were novel, of which 8 contained new alleles, the remainder were novel combinations of previously characterised alleles (Table 3.4).

Figure 3.3 shows the frequency of repeatedly detected STs among collections from Manchester and Preston, where 9 STs were consistently detected in both collections. Although the number of isolates was too small to allow examination of statistical significance, some notable changes occurred between 2007 and 2009 with increases in the proportion of ST73 in Preston and ST131 in Manchester, but a decrease in ST69 in Manchester.

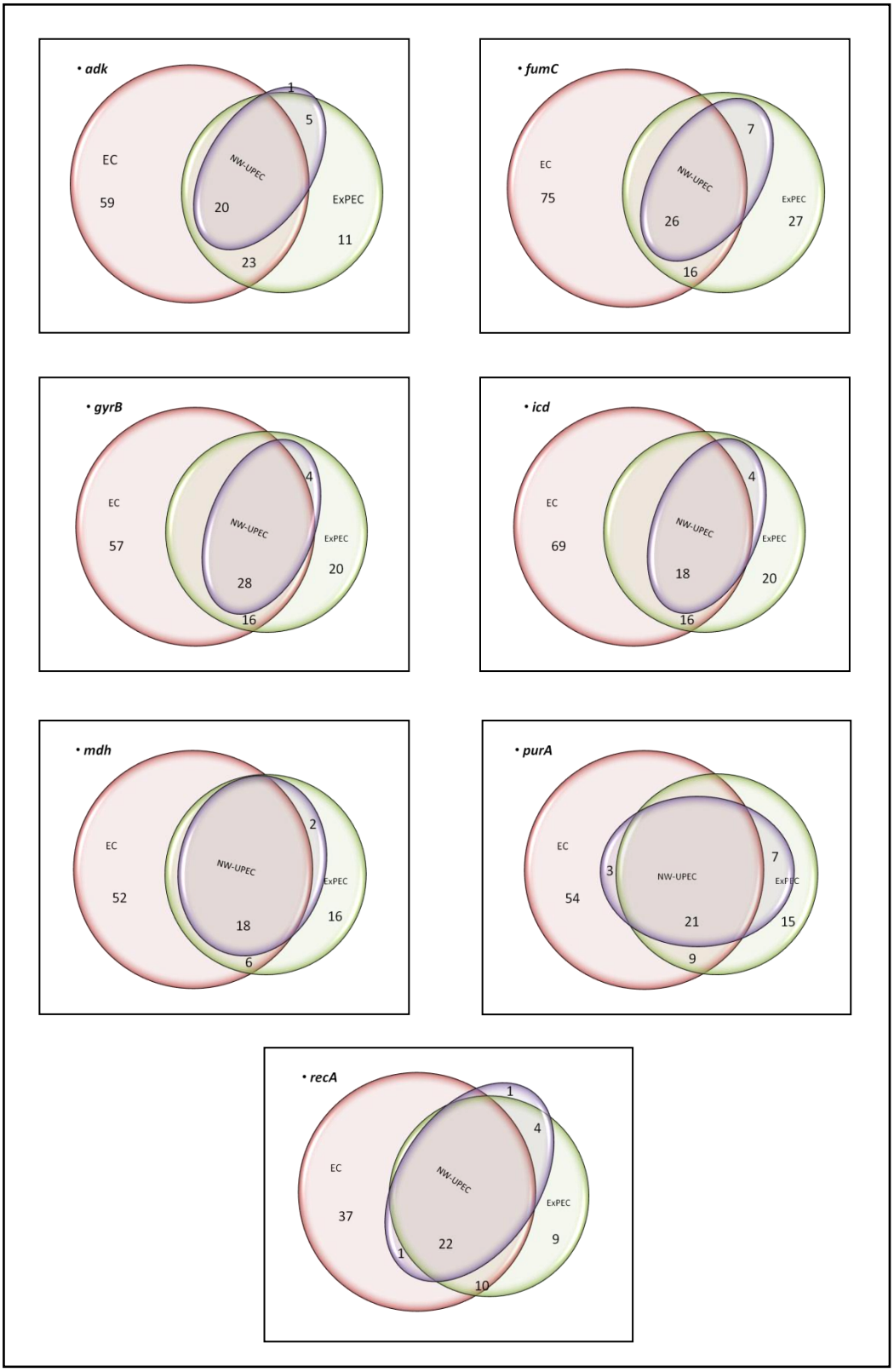


Figure 3.1 Proportional Venn-diagrams showing the number of alleles shared between the three screened populations (NW-UPEC, ExPEC and EC databases)

Table 3.3 NW-UPEC allelic specificity compared to that reported for database ExPEC and EC isolates.

	No. (%) of Sequence type																											
	adk				***fumC***				***gyrB***				***icd***				***mdh***				***purA***				***recA***			
Allele	**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**		**NW UPEC**		**ExPEC**	
8	0	(0.0)	1	(0.3)	1	(1.0)	2	(0.7)	0	(0.0)	2	(0.7)	4	(4.0)	21	(7.2)	3	(3.0)	27	(9.2)	15	(15.2)	61	(20.8)	0	(0.0)	0	(0.0)
17	0	(0.0)	1	(0.3)	0	(0.0)	0	(0.0)	3	(3.0)	6	(2.0)	0	(0.0)	0	(0.0)	28	(28.3)	53	(18.1	0	(0.0)	0	(0.0)	2	(2.0)	10	(3.4)
39	0	(0.0)	1	(0.3)	4	(4.0)	2	(0.7)	1	(1.0)	1	(0.3)	0	(0.0)	1	(0.3)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)

Allele	**NW UPEC**		**EC**		**NW UPEC**		**EC**		**NW UPEC**		**EC**		**NW UPEC**		**EC**		**NW UPEC**		**EC**		**NW UPEC**		**EC**		**NW UPEC**		**EC**	
2	0	(0.0)	0	(0.0)	0	(0.0)	4	(0.8)	1	(1.0)	12	(2.)3	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(1.0)	22	(4.2)	9	(9.1)	114	(21.7)
4	2	(2.0)	13	(2.5)	13	(0.0)	114	(21.7)	10	(10.1)	89	(17.0)	0	(0.0)	2	(0.4)	5	(5.1)	6	(1.1)	0	(0.0)	3	(0.6)	6	(6.1)	30	(5.7)
5	0	(0.0)	1	(0.2)	0	(0.0)	1	(0.2)	3	(3.0)	44	(8.4)	0	(0.0)	0	(0.0)	11	(11.1)	33	(6.3)	13	(13.1)	25	(4.8)	0	(0.0)	0	(0.0)
6	24	(24.2)	165	(31.4)	4	(4.0)	26	(5.0)	0	(0.0)	0	(0.0)	6	(6.1)	9	(1.7)	0	(0.0)	2	(0.4)	0	(0.0)	0	(0.0)	8	(8.1)	88	(16.8)
7	0	(0.0)	3	(0.6)	1	(1.0)	28	(5.3)	0	(0.0)	17	(3.2)	0	(0.0)	2	(0.4)	3	(3.0)	40	(7.6)	6	(6.1)	51	(9.7)	11	(11.1)	109	(20.8)
8	0	(0.0)	8	(1.5)	1	(1.0)	4	(0.8)	0	(0.0)	3	(0.6)	4	(4.0)	63	(12.0)	3	(3.0)	74	(14.1)	15	(15.2)	178	(33.9)	0	(0.0)	1	(0.2)
9	0	(0.0)	18	(3.4)	0	(0.0)	1	(0.2)	16	(16.2)	8	(1.5)	0	(0.0)	5	(1.0)	6	(6.1)	67	(12.8)	0	(0.0)	4	(0.8)	5	(5.1)	3	(0.6)
10	3	(3.0)	46	(8.8)	0	(0.0)	0	(0.0)	8	(8.1)	12	(2.3)	4	(4.0)	22	(4.2)	0	(0.0)	5	(1.0)	3	(3.0)	5	(1.0)	10	(10.1)	5	(1.0)
11	0	(0.0)	0	(0.0)	4	(4.0)	70	(13.3)	0	(0.0)	1	(0.2)	0	(0.0)	2	(0.4)	6	(0.0)	51	(0.0)	20	(20.2)	19	(3.6)	0	(0.0)	5	(1.0)
13	17	(17.2)	25	(4.8)	5	(5.1)	5	(1.0)	0	(0.0)	16	(3.0)	24	(24.2)	27	(5.1)	0	(0.0)	0	(0.0)	2	(2.0)	14	(2.7)	0	(0.0)	2	(0.4)
14	5	(5.1)	0	(0.0)	7	(7.1)	2	(0.4)	1	(1.0)	13	(2.5)	9	(9.1)	5	(1.0)	0	(0.0)	0	(0.0)	1	(1.0)	11	(2.1)	1	(1.0)	24	(4.6)
15	0	(0.0)	7	(1.3)	0	(0.0)	3	(0.6)	0	(0.0)	23	(4.4)	0	(0.0)	5	(1.0)	0	(0.0)	4	(0.8)	0	(0.0)	2	(0.4)	0	(0.0)	4	(0.8)
16	0	(0.0)	11	(2.1)	0	(0.0)	1	(0.2)	0	(0.0)	2	(0.4)	4	(4.0)	56	(10.7)	6	(6.1)	4	(0.8)	1	(1.0)	17	(3.2)	1	(1.0)	6	(1.1)
17	0	(0.0)	1	(0.2)	0	(0.0)	1	(0.2)	3	(3.0)	5	(1.0)	0	(0.0)	5	(1.0)	28	(28.3)	32	(6.1)	0	(0.0)	1	(0.2)	2	(2.0)	7	(1.3)
18	1	(1.0)	7	(1.3)	0	(0.0)	2	(0.4)	0	(0.0)	4	(0.8)	2	(2.0)	43	(8.2)	2	(2.0)	6	(1.1)	2	(2.0)	20	(3.8)	0	(0.0)	1	(0.2)
19	0	(0.0)	7	(1.3)	3	(3.0)	7	(1.3)	16	(0.0)	31	(0.0)	0	(0.0)	2	(0.4)	0	(0.0)	0	(0.0)	1	(1.0)	0	(0.0)	2	(2.0)	12	(2.3)
21	4	(4.0)	5	(1.0)	0	(0.0)	6	(1.1)	0	(0.0)	1	(0.0)	0	(0.0)	4	(0.8)	0	(0.0)	4	(0.8)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
24	0	(0.0)	1	(0.2)	14	(14.1)	16	(3.0)	1	(1.0)	0	(0.0)	0	(0.0)	6	(1.1)	1	(1.0)	35	(6.7)	0	(0.0)	0	(0.0)	1	(1.0)	1	(0.2)
25	0	(0.0)	0	(0.0)	0	(0.0)	5	(1.0)	1	(1.0)	1	(0.2)	7	(7.1)	8	(1.5)	0	(0.0)	1	(0.2)	3	(3.0)	4	(0.8)	12	(12.1)	15	(2.9)
27	1	(1.0)	0	(0.0)	0	(0.0)	10	(1.9)	3	(0.0)	0	(0.0)	0	(0.0)	5	(1.0)	0	(0.0)	0	(0.0)	1	(1.0)	1	(0.2)	0	(0.0)	0	(0.0)
28	1	(1.0)	0	(0.0)	0	(0.0)	2	(0.4)	1	(0.0)	1	(0.0)	0	(0.0)	4	(0.8)	1	(1.0)	2	(0.4)	4	(4.0)	1	(0.2)	2	(2.0)	0	(0.0)
30	0	(0.0)	0	(0.0)	0	(0.0)	2	(0.4)	2	(0.0)	0	(0.0)	0	(0.0)	1	(0.2)	2	(0.0)	4	(0.8)	0	(0.0)	1	(0.2)	1	(0.0)	1	(0.2)
34	2	(2.0)	5	(1.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.2)	0	(0.0)	4	(0.8)	4	(0.0)	0	(0.0)
36	7	(7.1)	3	(0.6)	2	(2.0)	3	(0.6)	0	(0.0)	0	(0.0)	6	(6.1)	8	(1.5)	3	(0.0)	3	(0.6)	0	(0.0)	1	(0.2)	0	(0.0)	5	(1.0)
37	8	(8.1)	8	(1.5)	4	(4.0)	1	(0.2)	0	(0.0)	0	(0.0)	8	(8.1)	12	(2.3)	0	(0.0)	3	(0.6)	1	(1.0)	4	(0.8)	0	(0.0)	2	(0.4)
38	2	(2.0)	0	(0.0)	9	(0.0)	12	(0.0)	0	(0.0)	1	(0.2)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	4	(0.8)
39	0	(0.0)	0	(0.0)	4	(0.0)	4	(0.0)	1	(0.0)	6	(1.1)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)
40	5	(5.1)	2	(0.4)	2	(0.0)	3	(0.0)	0	(0.0)	1	(0.2)	0	(0.0)	0	(0.0)	0	(0.0)	2	(0.0)	1	(0.0)	9	(0.0)	0	(0.0)	0	(0.0)
56	0	(0.0)	29	(5.5)	0	(0.0)	1	(0.0)	0	(0.0)	2	(0.4)	0	(0.0)	2	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	4	(0.0)	0	(0.0)	9	(1.7)
68	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	3	(0.6)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	5	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)
69	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	0	(0.0)	5	(1.0)	0	(0.0)	0	(0.0)	2	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.2)
73	0	(0.0)	1	(0.2)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	2	(0.4)	4	(4.0)	2	(0.4)
74	0	(0.0)	1	(0.2)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.0)	2	(2.0)	0	(0.0)
99	0	(0.0)	1	(0.2)	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.2)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	2	(2.0)	0	(0.0)

Significant differences were identified using Fisher's exact test. Highlights indicate significant (*P≤0.05)* differences in the incidence of alleles in the different populations.

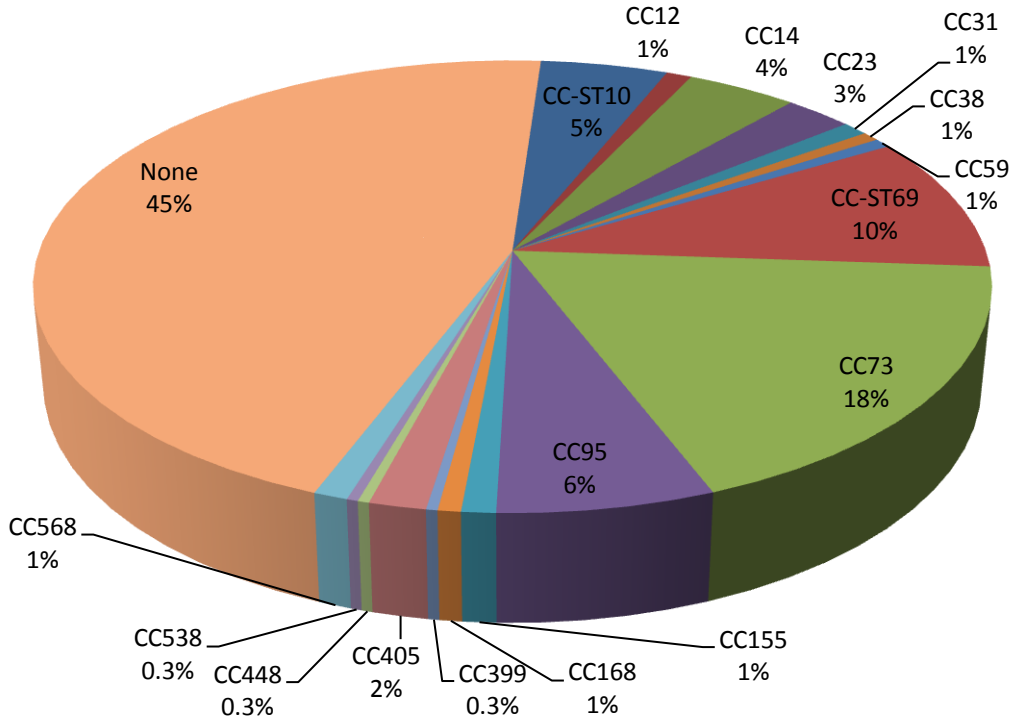


Figure 3.2 The frequency of clonal complexes among the 300 UPEC isolates

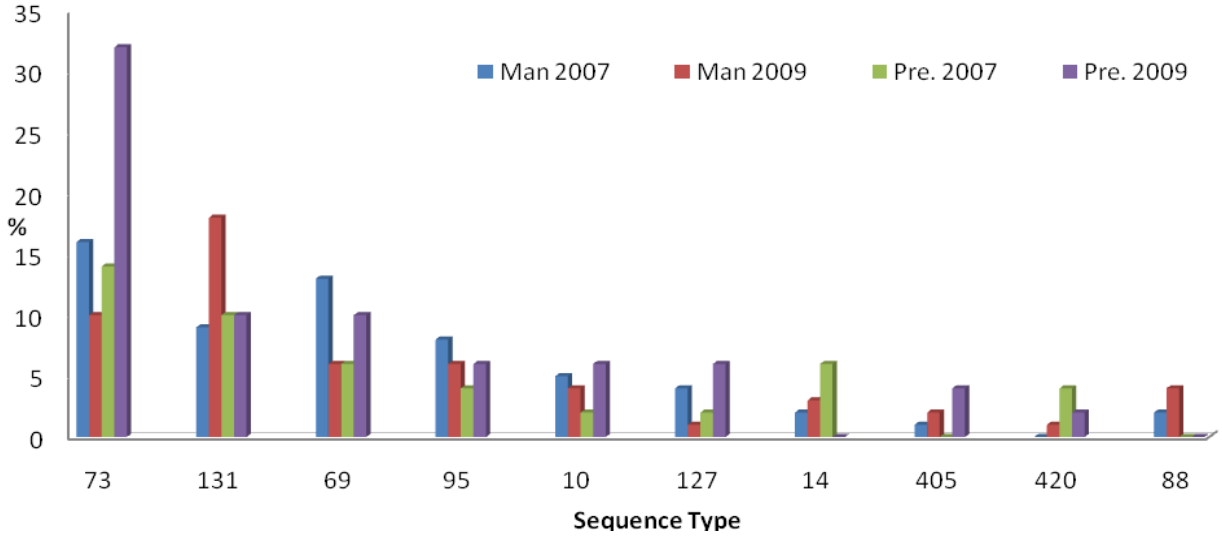


Figure 3.3 Frequency of different STs detected in both Manchester and Preston

Table 3.4 Frequency of detected sequence types and their allelic profiles

ST	CC	Profile							Frequency	% of Database
		adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*		
73	ST-73	36	24	9	13	17	11	25	50	16.67
131	-	53	40	47	13	36	28	29	37	12.33
69	ST-69	21	35	27	6	5	5	4	27	9
95	ST-95	37	38	19	37	17	11	26	19	6.33
10	ST-10	10	11	4	8	8	8	2	13	4.33
127	-	13	14	19	36	23	11	10	11	3.67
14	ST-14	14	14	10	14	17	7	10	8	2.67
88	ST-23	6	4	12	1	20	12	7	6	2
372	-	88	103	19	36	23	44	26	5	1.67
405	ST-405	35	37	29	25	4	5	73	5	1.67
141	-	13	52	10	14	17	25	17	4	1.33
420	-	13	38	84	13	17	68	34	4	1.33
58	ST-155	6	4	4	16	24	8	14	3	1
62	-	28	33	25	29	7	11	24	3	1
393	ST-31	18	106	17	6	5	5	4	3	1
404	ST-14	14	14	10	14	17	7	74	3	1
843	-	40	24	19	14	23	1	10	3	1
12	ST-12	13	13	9	13	16	10	9	2	0.67
38	ST-38	4	26	2	25	5	5	19	2	0.67
59	ST-59	27	32	24	29	26	19	22	2	0.67
80	ST-568	13	24	19	14	23	1	10	2	0.67
93	ST-168	6	11	4	10	7	8	6	2	0.67
104	ST-73	13	24	9	13	17	11	25	2	0.67
106	ST-69	21	35	27	6	5	8	4	2	0.67
117	-	20	45	41	43	5	32	2	2	0.67
362	-	62	100	17	31	5	5	4	2	0.67
410	ST-23	6	4	12	1	20	18	7	2	0.67
550	ST-14	14	14	10	14	17	92	10	2	0.67
638	ST-73	76	24	9	13	17	11	25	2	0.67
808	-	36	38	19	37	17	11	26	2	0.67
48	ST-10	6	11	4	8	8	8	2	1	0.33
92	-	40	14	19	36	23	11	10	1	0.33
136	-	38	39	30	13	17	11	28	1	0.33
167	ST-10	10	11	4	8	8	13	2	1	0.33
210	-	6	8	4	1	9	48	7	1	0.33
224	-	6	4	33	16	11	8	6	1	0.33
295	-	6	4	12	1	9	2	7	1	0.33
297	-	6	65	32	26	9	8	2	1	0.33
345	-	6	4	14	1	20	62	7	1	0.33
399	ST-399	6	4	1	95	69	8	20	1	0.33
448	ST-448	6	6	5	16	11	8	7	1	0.33
491	-	13	39	19	36	30	14	82	1	0.33
493	ST-12	40	13	9	13	16	10	9	1	0.33
538	ST-538	13	40	19	13	36	28	30	1	0.33
568	-	13	24	19	14	17	1	10	1	0.33
636	-	13	108	10	97	18	68	93	1	0.33
648	-	92	4	87	96	70	58	2	1	0.33
681	-	38	39	30	13	17	25	28	1	0.33
706	-	88	24	49	36	17	11	91	1	0.33
779	-	37	38	19	37	17	11	2	1	0.33
780	-	37	38	19	37	17	11	25	1	0.33

Table 3.4. Continued…

ST	CC	Profile							Frequency	% of Database
		adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*		
781	-	76	39	50	16	16	37	25	1	0.33
782	-	21	35	27	6	5	5	99	1	0.33
783	-	6	4	33	10	20	12	7	1	0.33
784	-	13	167	19	13	36	28	10	1	0.33
786	-	36	24	9	13	17	27	25	1	0.33
787	-	141	37	29	25	4	5	73	1	0.33
804	-	6	24	9	13	17	11	25	1	0.33
805	-	6	19	52	18	9	13	6	1	0.33
806	-	14	4	12	1	20	12	7	1	0.33
807	-	37	14	10	14	17	7	74	1	0.33
809	-	36	96	121	13	17	11	10	1	0.33
842	-	76	24	9	13	16	11	25	1	0.33
844	-	20	45	41	43	5	8	2	1	0.33
845	-	13	38	9	37	18	76	16	1	0.33
999	-	13	166	47	13	23	28	122	1	0.33
1000	-	37	24	9	172	17	11	25	1	0.33
1001	-	6	23	4	18	7	7	6	1	0.33
1002	-	35	37	4	25	4	5	73	1	0.33
1003	-	34	36	39	87	67	131	4	1	0.33
1004	-	6	19	3	171	11	8	6	1	0.33
1005	-	36	24	9	13	17	113	25	1	0.33
1058	-	37	38	19	37	17	11	34	1	0.33
1059	-	6	6	5	136	11	130	6	1	0.33
1303	-	10	7	4	8	12	35	2	1	0.33
1524	-	6	19	64	26	11	8	6	1	0.33
1525	-	6	4	1	13	69	8	20	1	0.33
1526	-	13	52	41	14	17	25	17	1	0.33
1527	-	40	13	9	13	16	126	9	1	0.33
1528	-	6	6	4	1	9	8	6	1	0.33
1529	-	40	1	84	13	17	68	34	1	0.33
1530	-	13	24	19	36	23	11	34	1	0.33
1531	-	37	38	19	37	17	5	26	1	0.33
1532	-	6	6	5	10	11	8	7	1	0.33
1533	-	13	13	9	13	16	76	9	1	0.33
1534	-	36	24	9	13	30	11	25	1	0.33
1535	-	52	116	55	10	35	40	38	1	0.33
1536	-	126	160	131	6	9	5	99	1	0.33
1537	-	1	4	12	1	20	12	7	1	0.33
1538	-	6	220	3	26	28	7	10	1	0.33
1539	-	34	36	28	25	5	16	4	1	0.33
1540	-	14	14	10	14	20	7	10	1	0.33
1541	-	37	38	19	37	17	68	26	1	0.33
1542	-	6	4	60	1	31	18	7	1	0.33
1543	-	4	26	17	25	5	5	19	1	0.33
1544	-	100	37	29	25	4	5	73	1	0.33
1545	-	21	35	67	6	5	5	44	1	0.33
1546	-	13	108	10	97	17	68	93	1	0.33
1547	-	36	24	9	13	4	11	25	1	0.33
1548	-	13	13	9	13	20	10	9	1	0.33

3.5. Phylogenetic analysis

3.5.1. eBURST analysis of STs

Using the eBURST v3 (http://eburst.mlst.net) clustering algorithm, 17 CCs encompassing 59 STs represented 259 UPEC isolates. The remaining 41 isolates appeared as singletons, which do not fall into any eBURST CC (Table 3.5). For clarity it should be mentioned that eBURST CC have a nomenclature that is not related to the conventional MLST naming schemes. Clonal Complexes were identified according to the number of single locus variants (SLVs) and double locus variants (DLVs) shared between isolates, where only STs that share six or more loci are designated to a defined CC. The largest CC (CC1) accounted for 20% (60) of the tested population and was dominated by a single ST (ST 73). ST73 was identified by eBURST as the founder of the CC1 group with the greatest number of SLVs and DLVs and with a bootstrap value of 96%. Additionally, there were eight other major groups comprising of more than 2 STs for which eBURST was able to identify founders with a reasonable level of confidence indicated by bootstrap values ranging between 98% and 29%.

Figure 3.4 illustrates eBURST clonal lineages showing in addition to SLVs, numerous inter-clonal and intra-clonal DLVs suggesting that frequent recombinational events occur within the defined population. CC1 shows the highest DLVs, some of which, for example ST1000, exist outside of the CC. Similarly CC5 reveals numerous DLVs with many outsider STs (ST295, ST345, ST783 and ST1542).

According to eBURST default definitions, each ST that has at least two descendent SLVs beside the SLV that is its progenitor, will be assigned as a subgroup founder (represented by a yellow dot in Figures 3.3, 3.4 and 3.5). In the dataset presented there

are three CCs with a subgroup founder (ST638 for CC1, ST404 for CC4 and ST493 for CC3).

As many of the successful STs, such as ST131 and ST 127, failed to be assigned to a well defined clonal lineage, even when less stringent conditions were applied (i.e. sharing identical alleles at four of seven loci), the clonality of NW-UPEC dataset was examined in an eBURST comparison analysis with the international MLST database to explore the possibility of a common ancestor not present in current dataset. Firstly, with ExPEC associated STs, both ST131 and ST127 were assigned as founders of distinct CCs with six and five STs, respectively associated with numerous SLVs and DLVs with bootstrap values of 97% and 94% but none of which stretched to involve other CCs (Figure 3.5). Secondly when expanding the comparison to involve the whole *E. coli* database, in addition to the re-enforcement of the existing clonal lineage the layout of clonal lineages dramatically changed as CC7, with its founder ST10 became the predominant CC in the whole *E. coli* snapshot with the most SLVs and DLVs and subgroup founders. CC131 and CC127 remained with inclusive SLVs and DLVs (Figure 3.6).

3.5.2. Phylogenetic Trees

The phylogenetic relationship between STs was presented by generating a UPGMA dendrogram using graphical tools integrated within the START v2.0 software. Figure 3.7 shows an unrooted UPGMA dendrogram resolving the 100 allelic profiles of the seven genes into 2 large clusters (GI & GII) at 0.9 linkage distance. This distance represented the number of loci that were identical between two allelic profiles, which ranged from zero (100% similarity) to 1 (dissimilar). These results correlated well with those obtained using eBURST.

Table 3.5 The 17 clonal complexes obtained from eBURST analysis of the UPEC population, showing the frequency of each sequence type and their corresponding SLVs and DLVs, the predicted founder of each group is highlighted and the level of confidence in assignation of the founding genotype is indicated by the bootstrap value.

CC 1: No. Isolates = 60 | No. STs = 9 | Predicted Founder = 73						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
73	50	7	1	0	96%	97%
638	2	4	4	0	25%	14%
104	2	3	5	0	1%	0%
804	1	3	5	0	3%	0%
1534	1	2	6	0	0%	0%
1547	1	2	6	0	0%	0%
1005	1	2	5	1	0%	0%
786	1	2	5	1	0%	0%
842	1	1	5	2	0%	0%

CC 2: No. Isolates = 26 | No. STs = 7 | Predicted Founder = 95						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
95	19	6	0	0	98%	95%
779	1	3	3	0	4%	1%
1058	1	3	3	0	5%	1%
780	1	3	3	0	10%	1%
1531	1	2	4	0	0%	0%
1541	1	2	4	0	0%	0%
808	2	1	5	0	0%	0%

CC 3: No. Isolates = 6 | No. STs = 5 | Predicted Founder = 12						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
12	2	3	1	0	63%	17%
493	1	2	2	0	17%	0%
1533	1	1	3	0	0%	0%
1527	1	1	2	1	0%	0%
1548	1	1	2	1	0%	0%

CC 4: No. Isolates = 15 | No. STs = 5 | Predicted Founder = 14						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
14	8	3	1	0	61%	16%
404	3	2	2	0	17%	0%
550	2	1	2	1	0%	0%
1540	1	1	2	1	0%	0%
807	1	1	1	2	0%	0%

CC 5: No. Isolates = 10 | No. STs = 4 | Predicted Founder = 88						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
88	6	3	0	0	61%	16%
1537	1	2	1	0	5%	0%
806	1	2	1	0	8%	0%
410	2	1	2	0	0%	0%

CC 6: No. Isolates = 8 | No. STs = 4 | Predicted Founder = 405						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
405	5	3	0	0	63%	13%
787	1	2	1	0	4%	0%
1544	1	2	1	0	10%	0%
1002	1	1	2	0	0%	0%

CC 7: No. Isolates = 15 | No. STs = 3 | Predicted Founder = 10						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
10	13	2	0	0	29%	0%
48	1	1	1	0	0%	0%
167	1	1	1	0	0%	0%

CC 8: No. Isolates = 6 | No. STs = 3 | Predicted Founder = 80						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
80	2	2	0	0	30%	0%
843	3	1	1	0	0%	0%
568	1	1	1	0	0%	0%

CC 9: No. Isolates = 30 | No. STs = 3 | Predicted Founder = 69						
ST	FREQ	SLV	DLV	TLV	ST Bootstrap	
					Group	Subgrp
69	27	2	0	0	31%	0%
106	2	1	1	0	0%	0%
782	1	1	1	0	0%	0%

CC 10: No. Isolates = 12 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
127	11	1	0	0	-
92	1	1	0	0	-

CC 11: No. Isolates = 2 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
1532	1	1	0	0	-
448	1	1	0	0	-

CC 12: No. Isolates = 5 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
141	4	1	0	0	-
1526	1	1	0	0	-

CC 13: No. Isolates = 3 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
117	2	1	0	0	-
844	1	1	0	0	-

CC 14: No. Isolates = 2 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
1525	1	1	0	0	-
399	1	1	0	0	-

CC 15: No. Isolates = 3 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
38	2	1	0	0	-
1543	1	1	0	0	-

CC 16: No. Isolates = 2 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
636	1	1	0	0	-
1546	1	1	0	0	-

CC 17: No. Isolates = 2 | No. STs = 2 | Predicted Founder = None					
ST	FREQ	SLV	DLV	TLV	ST Bootstrap
136	1	1	0	0	-
681	1	1	0	0	-

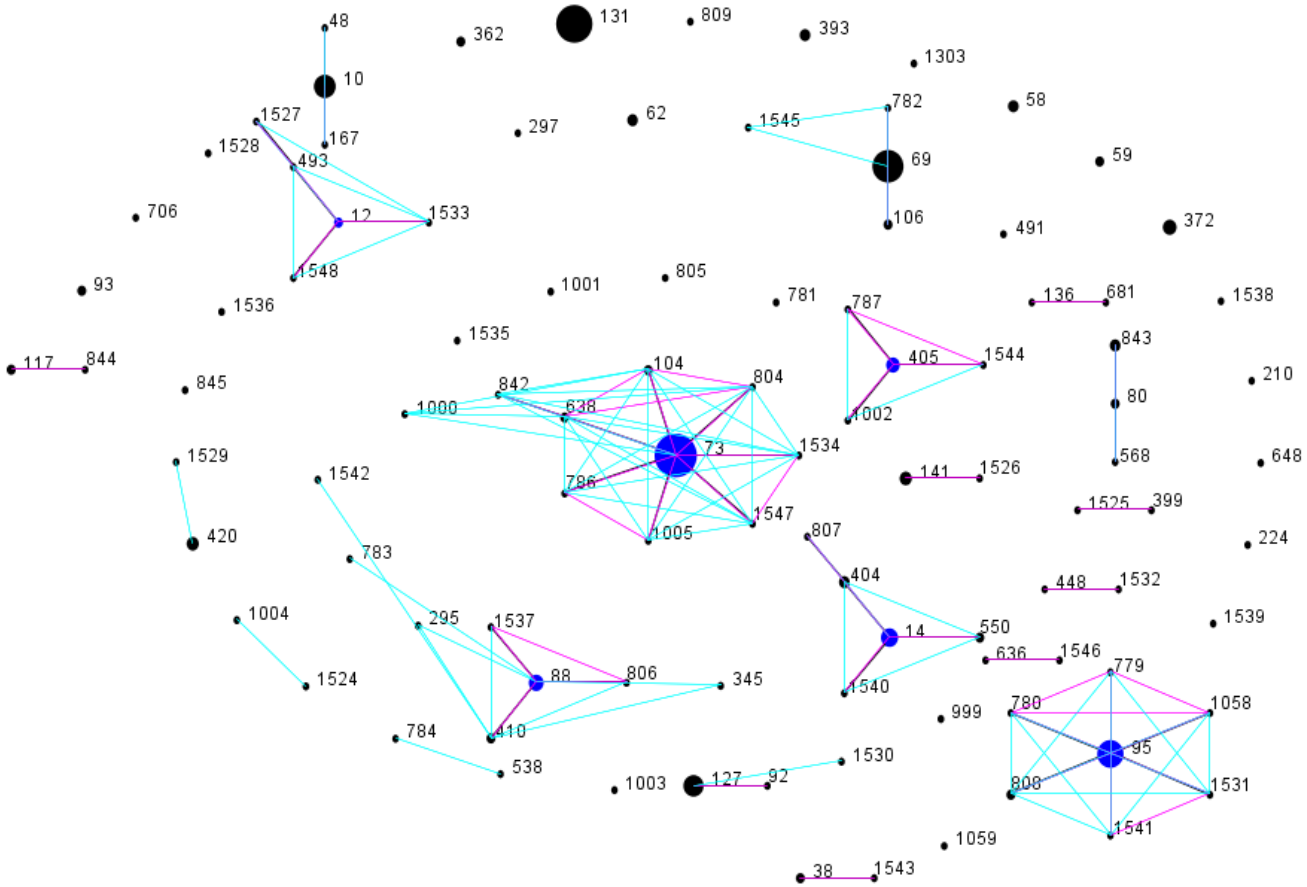


Figure 3.4 eBURST diagram of NW-UPEC showing clusters of related STs (CC) and individual singleton STs of the 300 UPEC. Seventeen CCs are identified, each ST is represented by a circle, the size of the circle indicates the frequency of each ST. Predicted founders are positioned centrally in each cluster and shown in blue and SLVs and DLVs shown in pink and blue, respectively

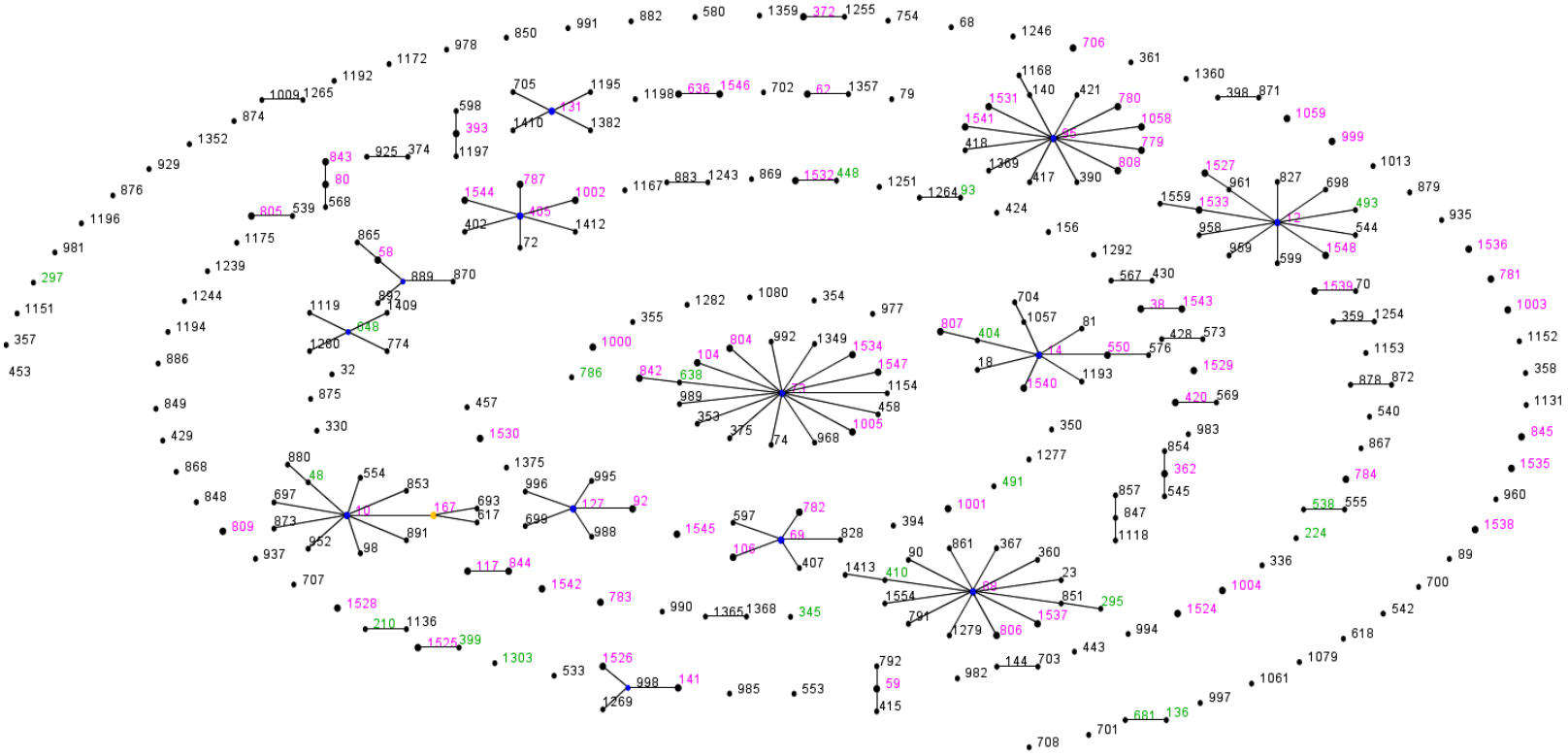


Figure 3.5 eBURST snapshot of NW-UPEC projected over ExPEC related STs, showing clusters of related STs (CC) and individual singleton STs of the NW-UPEC in pink, ExPEC from the database in black and shared STs in green. Predicted founders are positioned centrally in each cluster and shown in blue, subgroup founders are shown in yellow.

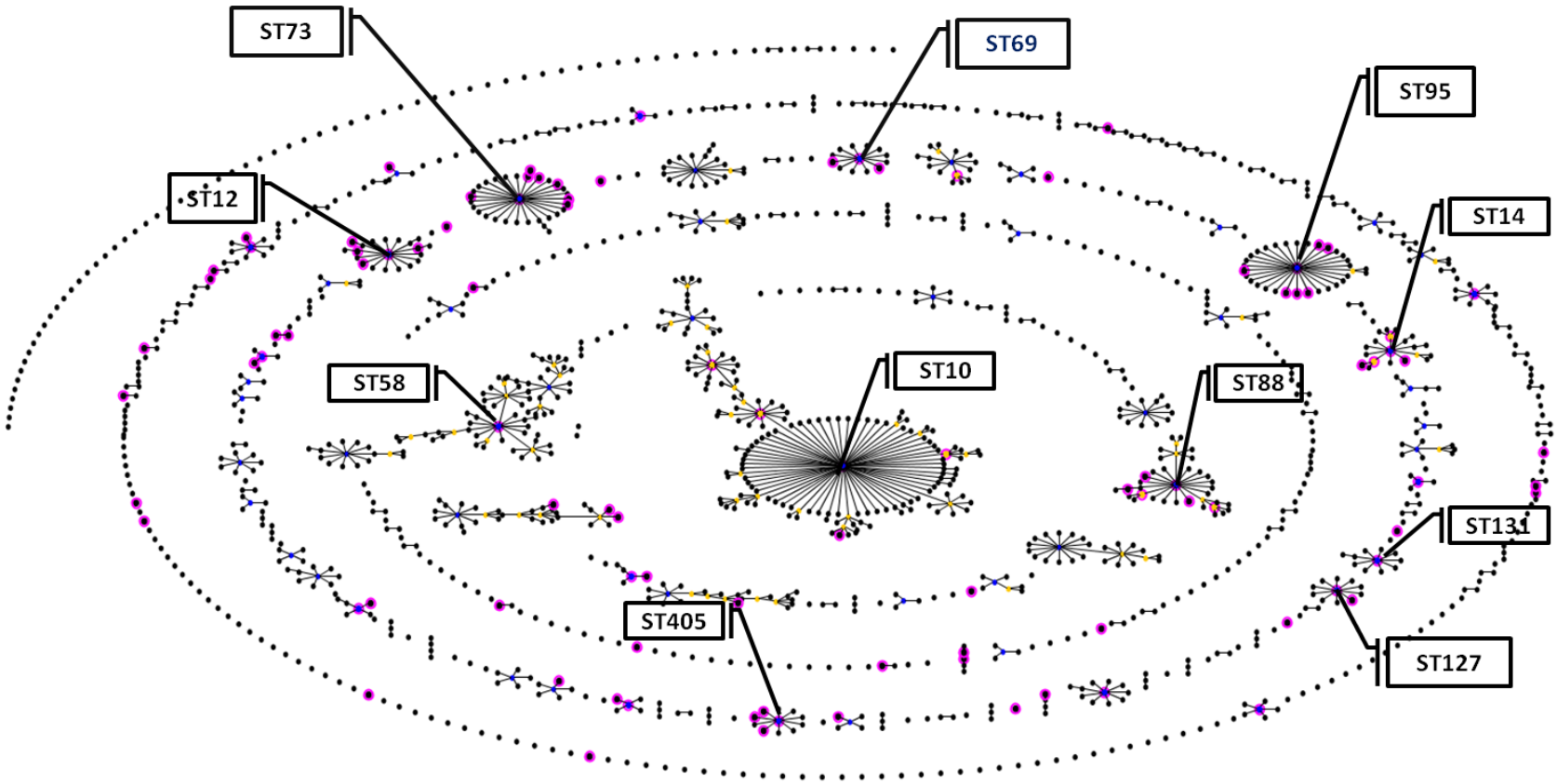


Figure 3.6 eBURST snapshot E. coli population illustrating the structure of common uropathogenic CCs and their related STs and showing clusters of related STs (CCs) and individual singleton STs of the NW-UPEC in pink and other *E. coli* from the database in black. Predicted founders are positioned centrally in each cluster and shown in blue, subgroup founders are shown in yellow

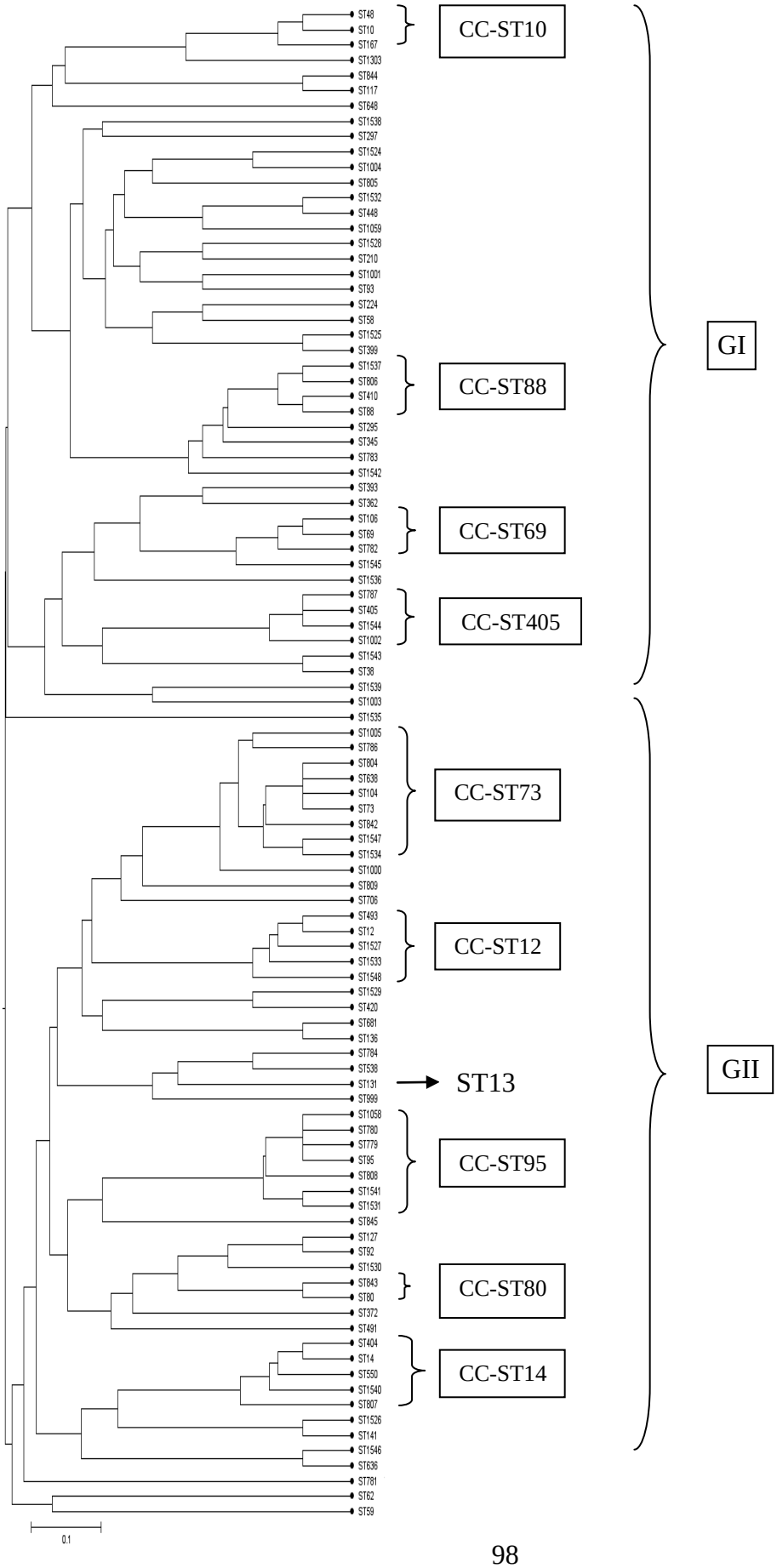

ST48
ST10
ST167
ST1303
ST844
ST117
ST648
ST1538
ST297
ST1524
ST1004
ST805
ST1532
ST448
ST1059
ST1528
ST210
ST1001
ST93
ST224
ST58
ST1525
ST399
ST1537
ST806
ST410
ST88
ST295
ST345
ST783
ST1542
ST393
ST362
ST106
ST69
ST782
ST1545
ST1536
ST787
ST405
ST1544
ST1002
ST1543
ST38
ST1539
ST1003
ST1535
ST1005
ST786
ST804
ST638
ST1104
ST73
ST842
ST1547
ST1534
ST1000
ST809
ST706
ST493
ST12
ST1527
ST1533
ST1548
ST1529
ST420
ST681
ST136
ST784
ST538
ST131
ST999
ST1058
ST780
ST779
ST95
ST808
ST1541
ST1531
ST845
ST127
ST92
ST1530
ST843
ST80
ST372
ST491
ST404
ST14
ST550
ST1540
ST807
ST1526
ST141
ST1546
ST1636
ST781
ST62
ST59
0.1
CC-ST10
CC-ST88
CC-ST69
CC-ST405
CC-ST73
CC-ST12
ST13
CC-ST95
CC-ST80
CC-ST14
GI
GII

Figure 3.7 (previous page) Unrooted UPGMA dendrogram, based on allelic profiles of 100 STs among the collected UPEC.

The nucleotide sequences of the seven gene fragments of unique STs were concatenated in the order *adk, fumC, gyrB, icd, mdh, purA* and *recA* using START v2.0 software and the resulting 3405bp sequence alignments were phylogenetically analysed using the Neighbour-joining alignment method integrated in MEGA 4.0 software with the Jukes-Cantor based algorithm and 1000 bootstrapping replications. Figure 3.8 shows the unrooted phylogenetic dendrogram of the 100 identified STs. In general the dataset was divided in to 3 large clusters with 60% of STs previously assigned to GI in the UPGMA tree split to form a distinct cluster (GIb) associated to (GII). Despite this shift in the relation of some CCs, the overall layout of the STs within CCs is consistent with that provided by eBURST analysis.

An alignment of unique concatenated sequences was tested by ClonalFrame to infer their phylogenetic relatedness. The majority-role consensus tree constructed, (Figure 3.9) again shows the two groups observed with UPGMA and NJ and almost agrees with the eBURST lineage display. In the eBURST analysis, CC95 is a consistent clone with seven STs (ST95, ST88, ST1531, ST1541, ST779, ST780 and ST1058). In ClonalFrame, this clone was reduced to only 4 ST (ST95, ST88, ST1531, ST1541) and the remaining STs joined other singletons arising from unknown ancestors. In contrast, CC73 extended its membership to include ST1000, which in eBURST analysis showed various DLVs with many CC73 members.

The network representation of the ClonalFrame analysis output illustrated in Figure 3.10 suggests that the common ancestor has not yet been identified and that numerous genotypes descended from this common ancestor have also not been identified among

the studied population. However, almost all STs appear to descend from two unknown ancestors, which in turn directly arise from the common ancestor.

Based on eBURST analysis, ST14 and ST405 were the founding genotypes for their respective clonal complexes. ClonalFrame analysis, however, does not put particular weight on ST14 and ST405 as clone founders and only regards ST14 as a descendant strain of an as yet unknown ancestor and ST 405 as descended from ST1544. Interestingly, ST127 which, based on eBURST analysis, fails to form its own clonal lineage in the dataset presented here, appears as a clonal founder in the ClonalFrame analysis.

Using the Neighbour-net method integrated in splitstree 4.11, a phylogenetic network was constructed using the concatenated nucleotide sequences representing the dataset. In general, Neighbour-net identified groups in a similar way to that presented in the ClonalFrame output. Neighbor-net was able to resolve the conflicting signals introduced to nucleotide sequences by recombination, as these events are displayed by parallel paths in the phylogenetic network, as shown in Figure 3.11. ST806 was separated from CC88, which might indicate that multiple alleles were acquired by homologous recombination. CC73 was mostly arranged with a radial symmetry indicating that most diversity was generated by mutation. Parallel paths play an important role in the formation of most other CCs indicating the possibility of substantial recombination.

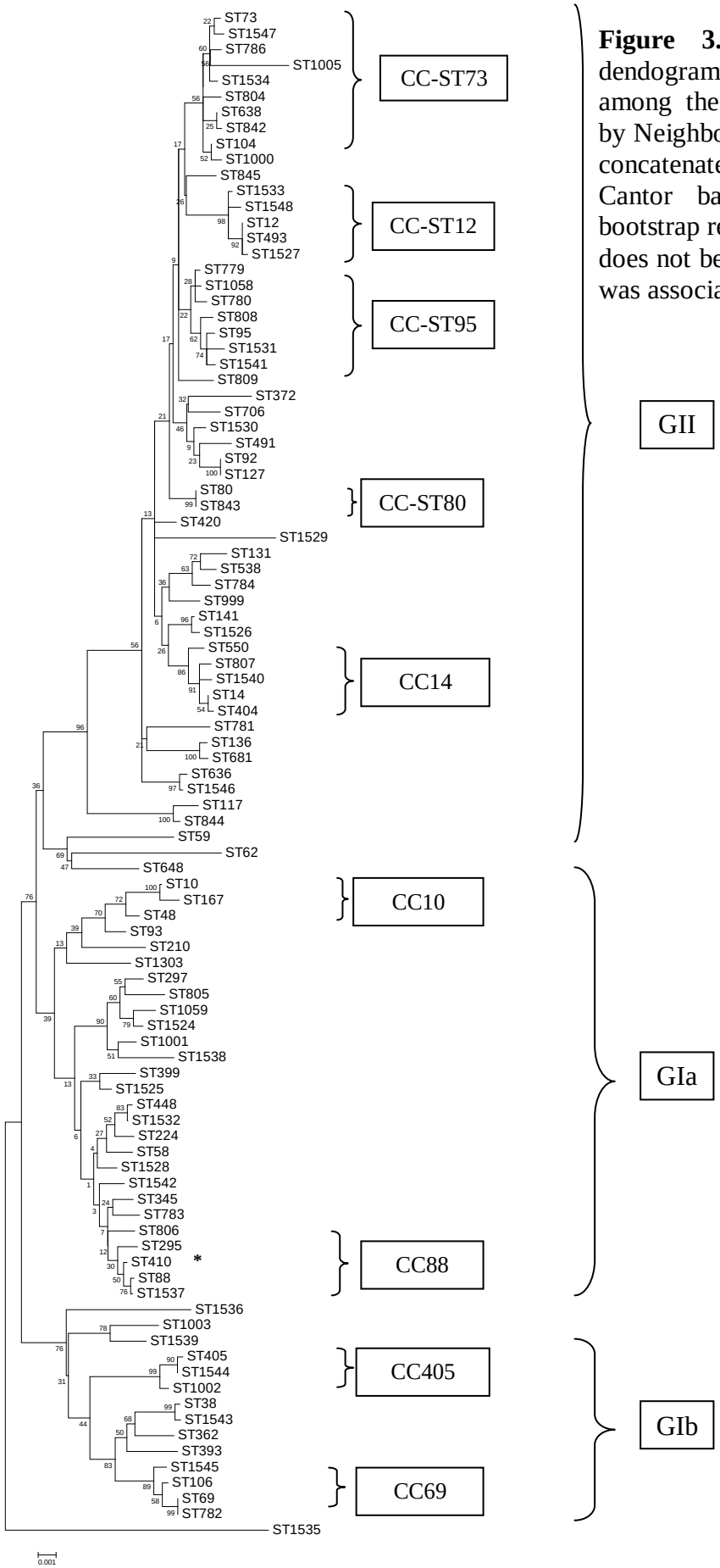


Figure 3.8 Unrooted phylogenetic dendogram of the identified 100 STs among the collected UPEC, analysed by Neighbour-joining method based on concatenated sequences (with Jukes-Cantor based algorithm and 1000 bootstrap replications). * denotes an ST does not belong to the assigned CC but was associated with it by DLVs.

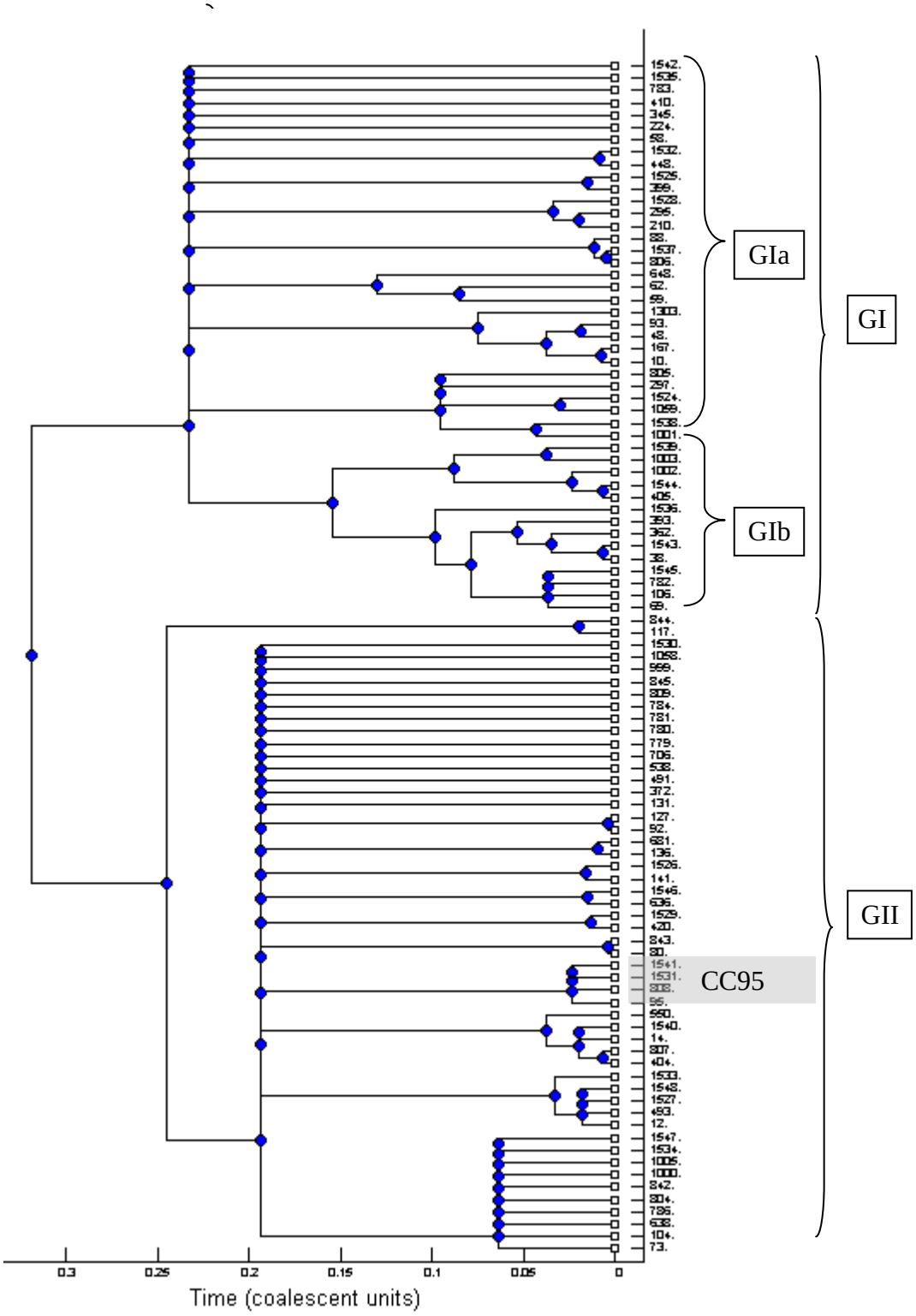


Figure 3.9 A 50% consensus of 10 evolutionary trees generated for NW-UPEC dataset using ClonalFrame1.1

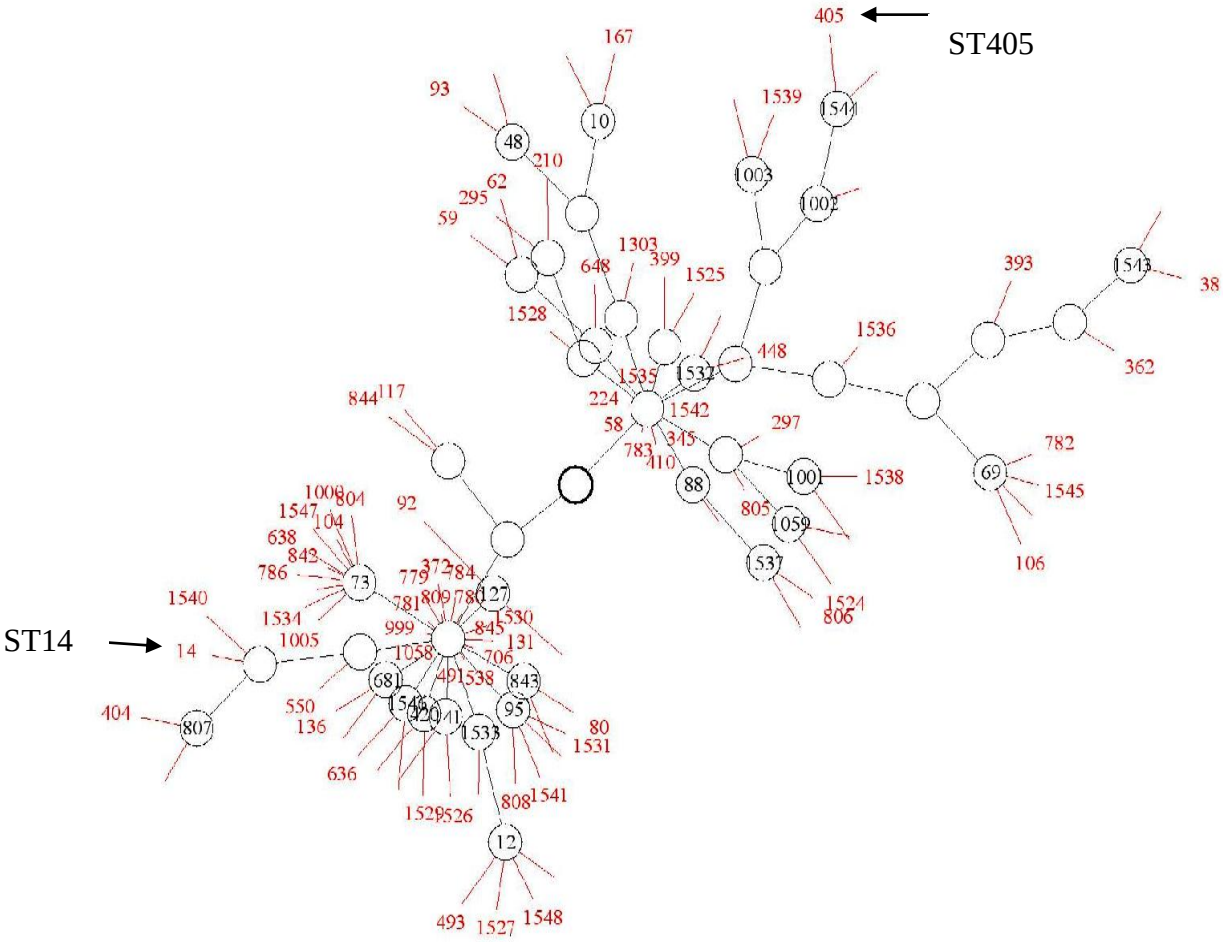


Figure 3.10 A network representation of the ClonalFrame output. The network shows inferred ancestral nodes in black and the STs in red with each line indicating a single ST. Nodes whose ancestral ST is not found amongst the study population are shown as an empty circle. The ancestral node of the network is indicated by a darker circle.

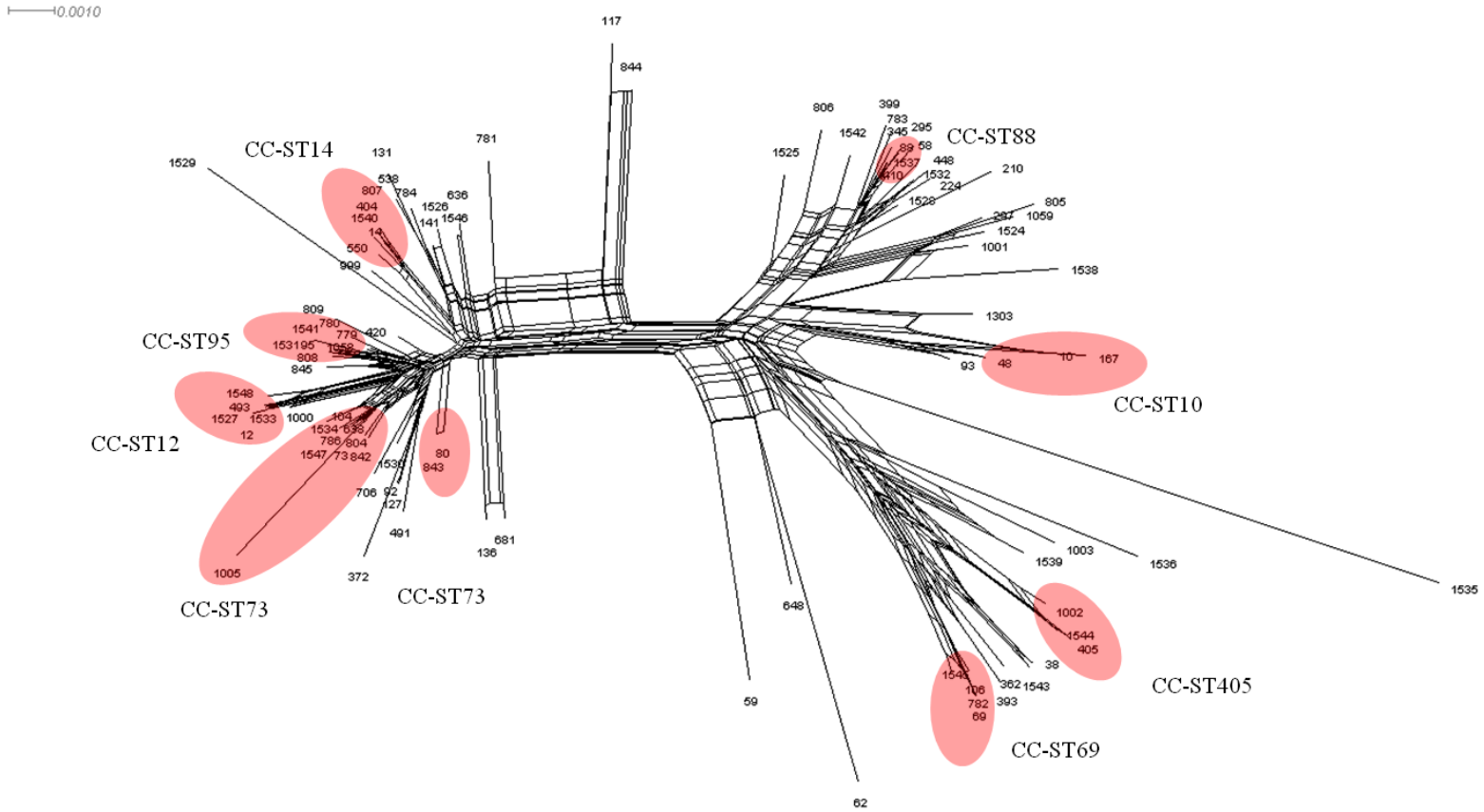


Figure 3.11 Neighbour-net based on concatenated sequence alignment of STs representing the NW-UPEC population. Shadowed areas indicate CCs

3.5.3. Analysis of recombination

From eBURST and ClonalFrame analysis it is clear that the dataset population evolved primarily by accumulation of single mutations with some recombinational events. So it is important to estimate the contribution of recombination in the evolution of the dataset population and the clonal groups.

3.5.3.1. Maximum X^2 test

Nucleotide sequences of target genes were analysed to identify potential recombination events between alleles of each locus. A large number of putative recombination sites were identified ranging from 231 sites in *icd* to 528 sites in *fumC* and *gyrB* from which few were considered significant after 1000 random comparison trials. Table 3.6 shows significant putative recombination sites in each locus. Most target genes show multiple possible recombination sites except *gyrB* where, from 528 identified sites, none were considered significant and only one site was assigned as a significant recombination site in *purA*.

Table 3.6 Frequency of significant putative recombination sites in each target gene

	Max X^2	
Alleles	Putative significant recombination sites	*P* value
adk	24	
fumC	12	
gyrB	0	
icd	10	<0.05
mdh	12	
purA	1	
recA	113	

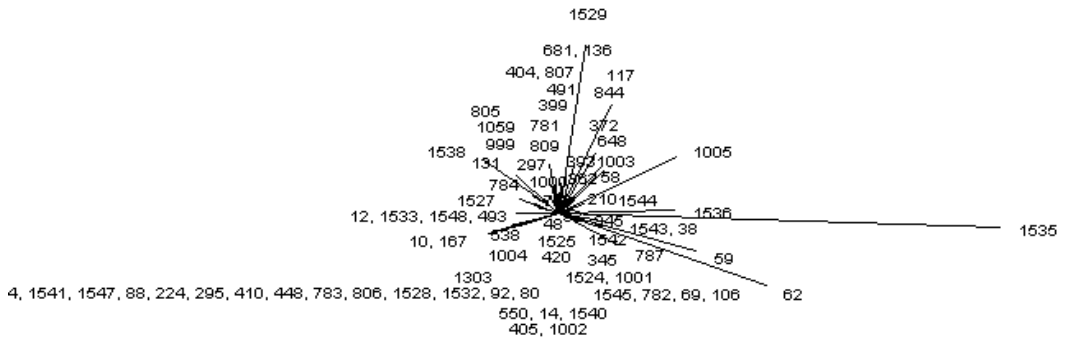
3.5.3.2. Split decomposition analysis

Concatenated allele sequences of the target genes from the 100 STs were analysed for evidence of recombination using a split decomposition algorithm in which the extent of recombination within the population is reflected by the degree of networks computed by the program. The star phylogeny displayed in Figure 3.11a is consistent with a clonal population structure.

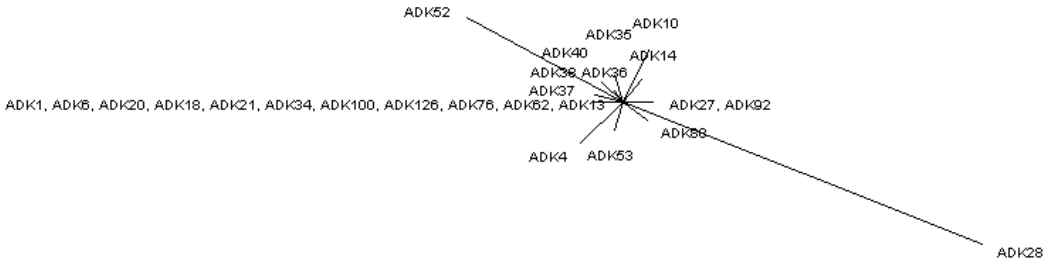
To further analyse the population structure, the target genes were analysed separately. The results are shown in Figures 3.12b-h. The structure of the split graphs varied substantially between the different loci. The split tree graph obtained with *mdh* presented an interconnected network indicating evidence of recombination whereas the remaining loci gave a tree-like structure consisting of a single central origin from which single branches radiate showing no evidence of recombination.

Split tree decomposition

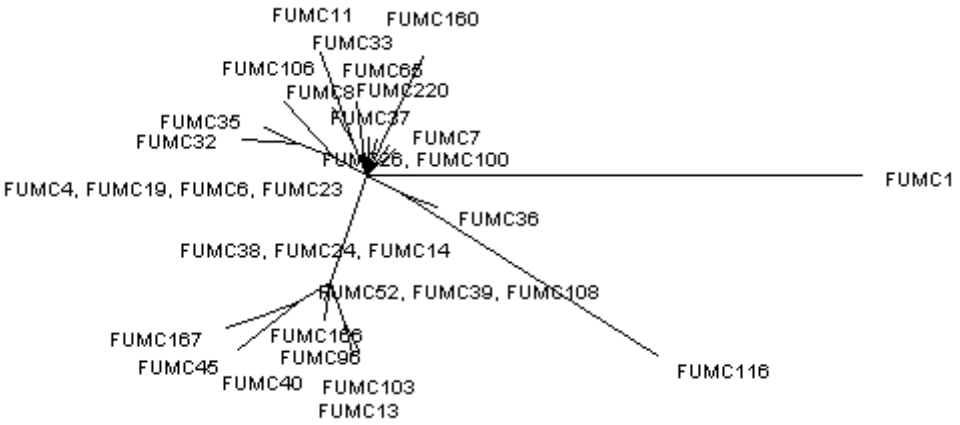
a) Concatenated sequences



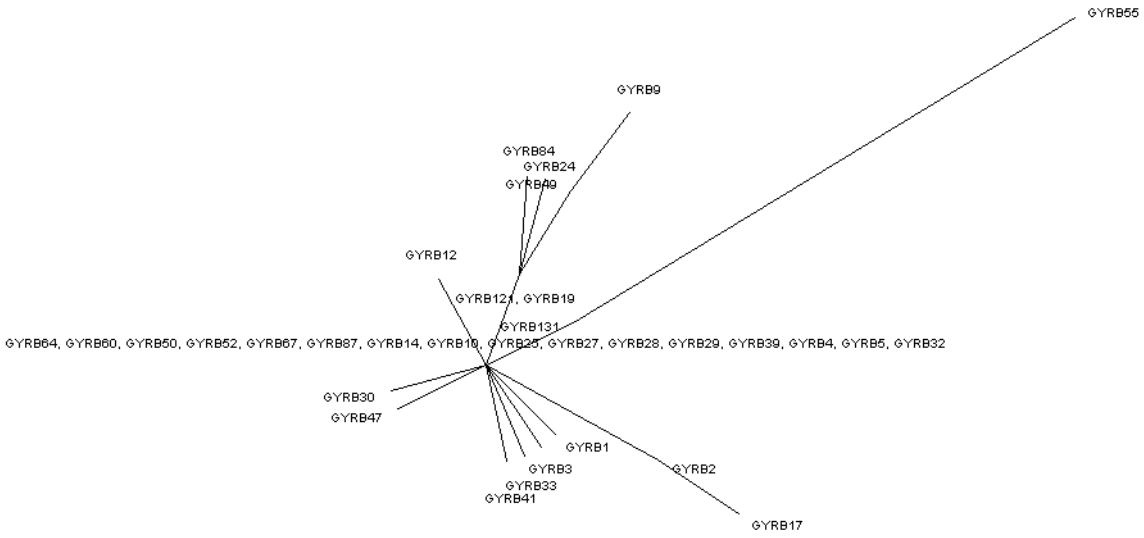
b) adk



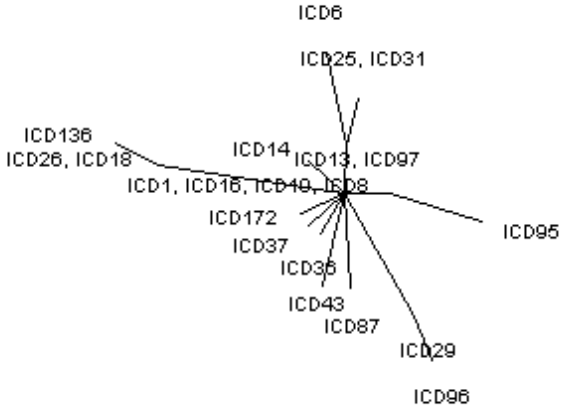
c) fumC



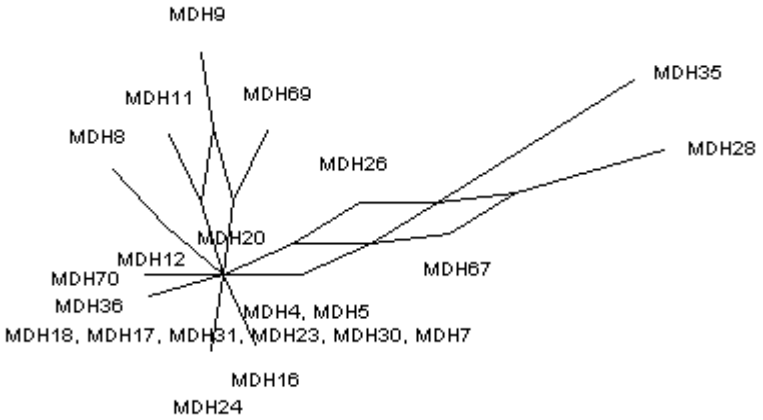
d) gyrB


GYRB55
GYRB9
GYRB84
GYRB24
GYRB49
GYRB12
GYRB121, GYRB19
GYRB131
GYRB64, GYRB60, GYRB50, GYRB52, GYRB67, GYRB87, GYRB14, GYRB10, GYRB25, GYRB27, GYRB28, GYRB29, GYRB39, GYRB4, GYRB5, GYRB32
GYRB30
GYRB47
GYRB1
GYRB3
GYRB33
GYRB41
GYRB2
GYRB17

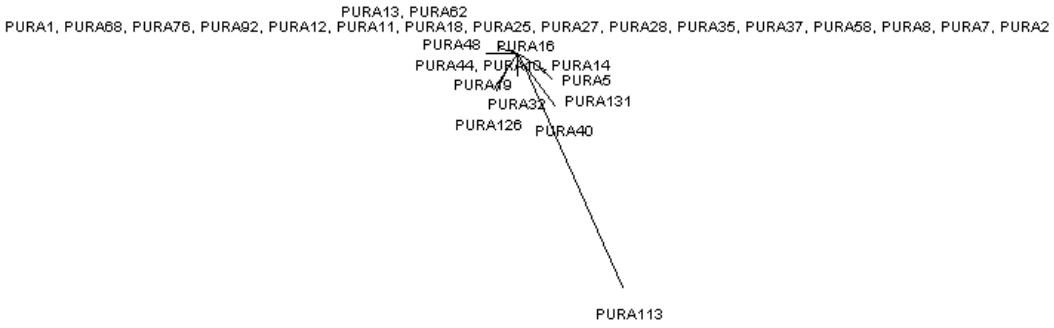
e) icd


ICD6
ICD25, ICD31
ICD136
ICD26, ICD18
ICD14
ICD13, ICD97
ICD1, ICD15, ICD49, ICD8
ICD172
ICD95
ICD37
ICD36
ICD43
ICD87
ICD29
ICD96

f) mdh


MDH9
MDH11
MDH69
MDH35
MDH8
MDH26
MDH28
MDH20
MDH12
MDH67
MDH70
MDH36
MDH4, MDH5
MDH18, MDH17, MDH31, MDH23, MDH30, MDH7
MDH16
MDH24

g) purA


PURA13, PURA62
PURA1, PURA68, PURA76, PURA92, PURA12, PURA11, PURA18, PURA25, PURA27, PURA28, PURA35, PURA37, PURA58, PURA8, PURA7, PURA2
PURA48
PURA16
PURA44, PURA19, PURA14
PURA5
PURA131
PURA32
PURA126
PURA40
PURA113

h) recA

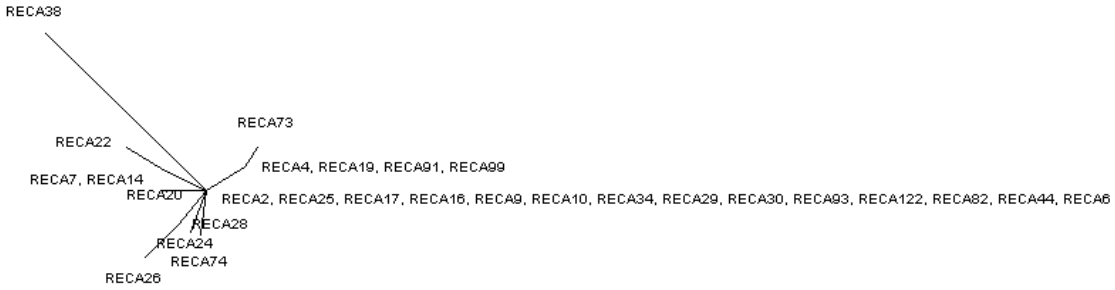


Figure 3.12 Split decomposition analysis(*a*) for concatenated allele sequences of the target genes from the 100 STs and (*b* - *h*) for individual allele sequences of each target gene

3.5.3.3. Pairwise homoplasy index (Φ_w) test

Since conflicting signals of recombination were observed for target genes either calculated in the max X^2 test or displayed in split tree diagrams, another statistical test was applied. The Φ_w test is considered a powerful tool to identify recombination and is able to distinguish recurrent mutation from recombination. However, with considerable parsimony informative sites, the Φ_w test was able to detect evidence of significant recombination in all loci except *adk* and *recA* (Table 3.7).

Table 3.7 Pairwise homoplasy index, (Φw) test of significant recombination events in each target gene

Allele	No. of alleles	Φ_w test		
		Informative sites	*P* value	Interpretation
adk	26	19	0.25	N.R
fumC	33	41	0.0	**S.R**
gyrB	33	17	0.0	**S.R**
icd	22	14	0.0	**S.R**
mdh	22	26	0.0	**S.R**
purA	31	17	0.012	**S.R**
recA	28	17	0.249	N.R

S.R:significant recombination; N.R: non significant recombination.

3.5.3.4. Index of association

The level of linkage between alleles at different loci was estimated by calculating the index of association (I^S_A) values using the standardised method introduced by Haubold and Hudson in 2000 and integrated within the START2 program. The I^S_A value found for all isolates was 0.535 (*P <0.0001*) but reduced to 0.224 (*P<0.0001*) when only one representative of each ST was considered. This indicates that the population is in linkage disequilibrium and thus the population structure of the present dataset is clonal (Table 3.8). The contribution of recombination was also estimated within previously identified clonal complexes using both the Φw test and linkage disequilibrium analysis (Table 3.9).

Table 3.8 Multilocus linkage disequilibrium analysis of the 300 UPEC isolates

	No.	V_e	V_o	I^S_A	P^a
Total isolates	300	0.8341	3.512	0.535	<0.0001
Total STs	100	0.5318	1.247	0.224	<0.0001

Ve, expected variance; *Vo*, observed variance; I^S_A,standardized index of association.P^a Probability of observing an Vo/Ve ratio as or more extreme than that found in the original data based on1000 trials.

Table 3.9 Contribution of recombination in different phylogenetic structures within the NW-UPEC population

CC	No of STs	Informative site	Φw	I^S_A value	*P* value	Linkage disequilibrium
dataset	100	184	0.0	0.2268	0.000	LD
GI	27	133	0.0	0.2226	0.000	LD
CC73	8	2	1.0	-0.0873	0.996	-
CC95	7	4	1.0	-0.1003	1.0	-
CC14	5	0	0.95	-0.0586	0.946	-
CC12	5	0	1.0	-0.0566	1.0	-
GII	14	85	0.0	0.1582	0.000	LD
CC69	3	0	1.0	-0.0188	0.878	-
CC405	4	0	1.0	-0.0524	1.0	-
CC10	3	0	1.0	-0.0417	1.0	-
CC88	4	0	1.0	-0.0524	1.0	-

3.6. Phylogenetic typing

E. coli may be grouped into four major phylogenetic groups A, B1, B2 and D. Of the 300 UPEC isolates tested 175 (58%) belonged to group D followed by B2 with 77 (26%) isolates with 36 (12%) and 11 (4%) belonging to groups A and B1 respectively.

To assess the phylogenetic distribution among defined STs, correspondence analysis was performed on frequently occurring STs and STs belonging to the nine eBURST CCs, whereas the remaining STs were represented collectively as singletons. Significant associations were detected between phylogenetic groups and different STs (*P <0.0001*) and the results expressed by the projection of each ST on a plane defined by two axes (F1 and F2), which account for most of the total variance (81%) allowing sorting of the 300 UPEC in the data set among the phylogenetic groups (Figure 3.13). Ten STs accounted for 139 (46%) isolates clustered between B2 and D indicating shared phylogenetic groups, whereas group D was represented by 30 STs accounting for 69 (23%) isolates, group A by seven STs with 27 (9%) isolates and six STs accounted for seven (2%) isolates associated with B1. Singletons were collectively distributed between groups D (the majority of singletons), A and B.

Correspondence analysis was also carried out to describe the distribution of the phylogenetic groups among the eBURST CCs. Despite the presence of mixed phylogenetic groups within some of the CCs, a significant linkage between the phylogenetic groups and CCs (*P <0.0001*) was observed. CC10 and CC88 were classified as group A, CC80 as B2 and CC69, CC95, CC12 and CC405 as group D. In CC73, while most of the STs were classified as group D, ST73, which is the founder of the CC, included a mixture of D and B2 isolates. Furthermore ST786 and ST1547 of CC73 were classified as group B2. This mosaic, which clearly occurs in mature CCs could indicate genetic exchange between isolates of different CCs.

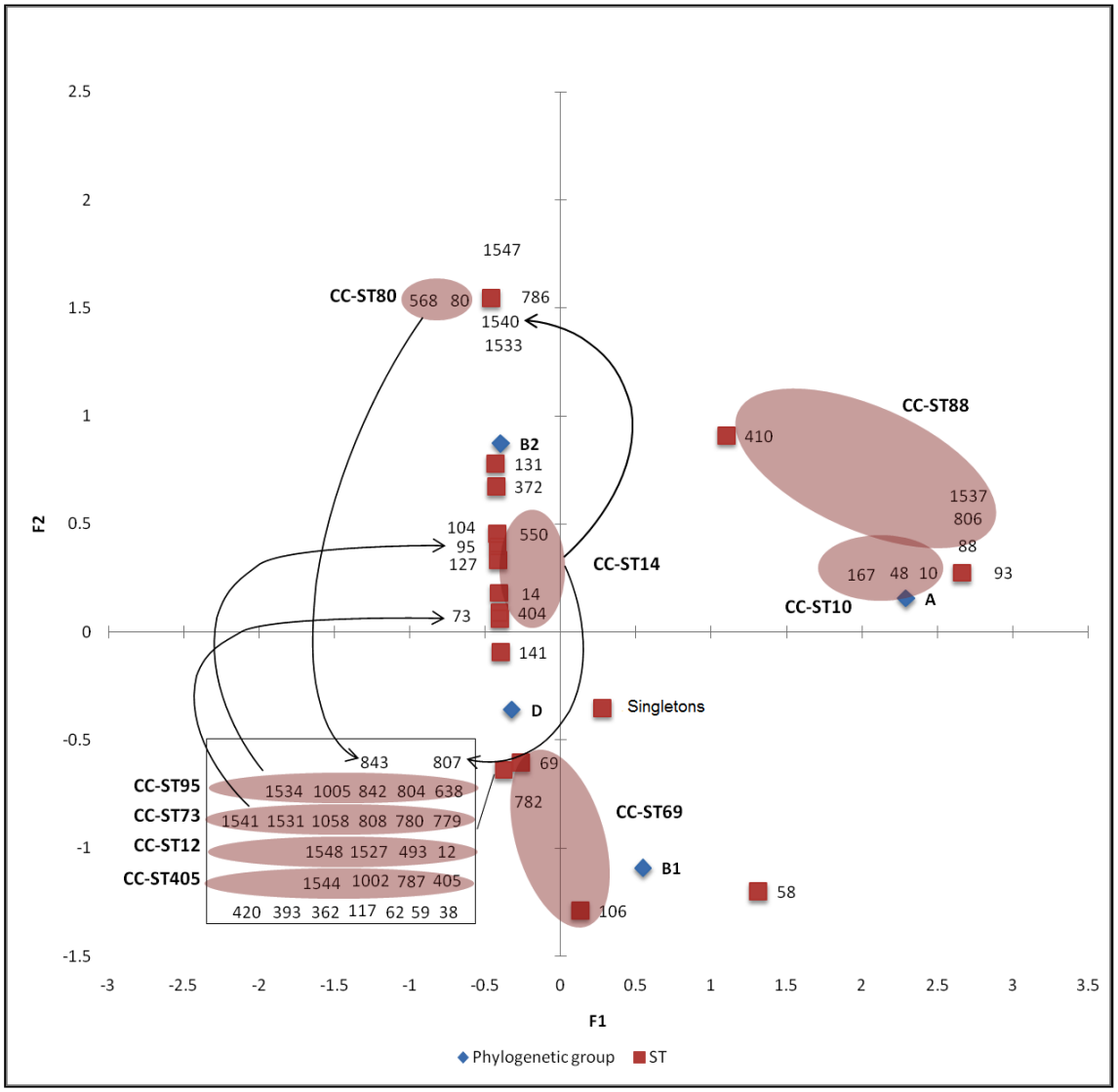


Figure 3.13 Graphical representation of the correspondence analysis of UPEC based on the MLST data. The F1/F2 plane accounted for 81% of total variance, with factors F1 and F2 accounting for 57% and 24%, respectively. Arrows show the position of individual STs when they arise away from their respective CCs.

3.7. Pulsed field gel electrophoresis

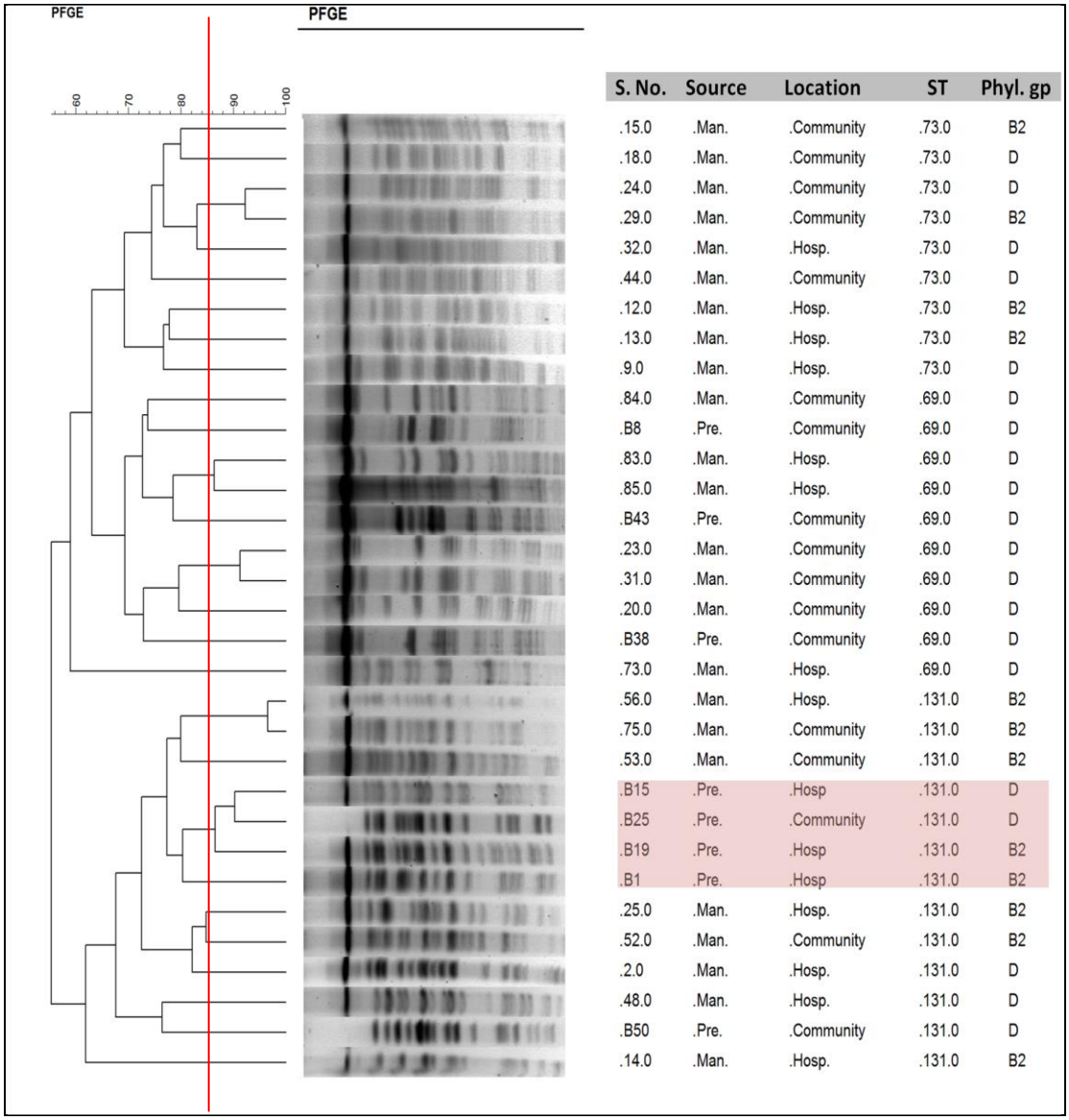
XbaI PFGE analysis was performed with modifications according to the CDC standardised PFGE protocol and the generated profiles were compared digitally using BioNumerics software. Cluster analysis of Dice similarity indices based on the UPGMA was used to generate a dendrogram describing the relationships among PFGE profiles of 32 representative isolates of the common UPEC STs with different phylogenetic groups, 10 ST69 (n=10 isolates), ST73 (n=9) and ST131 (n=13; Figure 3.14).

As expected, finer resolution was obtained by PFGE analysis and each ST constituted a separate cluster. The ST131 cluster (defined at the 65% similarity level) was tied to the two other cluster groups (ST69 and ST73) at less than 60% similarity.

Isolates were considered to belong to the same PFGE group if their Dice similarity index was ≥85% and using this criterion ST73 and ST69 clusters comprised of eight PFGE profiles each.

The ST 131 cluster in turn, comprised of nine different profiles with most of the isolates collected from Preston forming a sub-cluster at >80% similarity. There was no evidence for a geographical effect in the sub-clusters of both ST73 and ST69.

Interestingly isolates 24 and 29 of the ST73 cluster showed different phylogenetic groups (B2 and D) despite the fact that they both share the same PFGE profiles. Similarly, within the Preston ST131 sub-cluster isolates 15, 19 and 25 shared the same PFGE profile when they differed in their phylogenetic groups.



S. No.	Source	Location	ST	Phyl. gp
.15.0	.Man.	.Community	.73.0	B2
.18.0	.Man.	.Community	.73.0	D
.24.0	.Man.	.Community	.73.0	D
.29.0	.Man.	.Community	.73.0	B2
.32.0	.Man.	.Hosp.	.73.0	D
.44.0	.Man.	.Community	.73.0	D
.12.0	.Man.	.Hosp.	.73.0	B2
.13.0	.Man.	.Hosp.	.73.0	B2
.9.0	.Man.	.Hosp.	.73.0	D
.84.0	.Man.	.Community	.69.0	D
.B8	.Pre.	.Community	.69.0	D
.83.0	.Man.	.Hosp.	.69.0	D
.85.0	.Man.	.Hosp.	.69.0	D
.B43	.Pre.	.Community	.69.0	D
.23.0	.Man.	.Community	.69.0	D
.31.0	.Man.	.Community	.69.0	D
.20.0	.Man.	.Community	.69.0	D
.B38	.Pre.	.Community	.69.0	D
.73.0	.Man.	.Hosp.	.69.0	D
.56.0	.Man.	.Hosp.	.131.0	B2
.75.0	.Man.	.Community	.131.0	B2
.53.0	.Man.	.Community	.131.0	B2
.B15	.Pre.	.Hosp	.131.0	D
.B25	.Pre.	.Community	.131.0	D
.B19	.Pre.	.Hosp	.131.0	B2
.B1	.Pre.	.Hosp	.131.0	B2
.25.0	.Man.	.Hosp.	.131.0	B2
.52.0	.Man.	.Community	.131.0	B2
.2.0	.Man.	.Hosp.	.131.0	D
.48.0	.Man.	.Hosp.	.131.0	D
.B50	.Pre.	.Community	.131.0	D
.14.0	.Man.	.Hosp.	.131.0	B2

Figure 3.14 *XbaI* PFGE dendrogram for 32 UPEC isolates as produced by the UPGMA algorithm based on Dice similarity coefficients. Highlighted area indicates ST131 isolates collected in Preston. The red line shows 85% cutoff**.**

3.8. Demographic distribution of STs

Demographic parameters such as age, gender, specimen type and geographic origin of specimen were retrieved from laboratory records. Of the 300 MLST typed isolates, 82% were from females with patient's age ranging from <1 month to 95 years. Table 3.10 shows the distribution of patients by age and sex.

Table 3.10 Distribution of patients according to age and gender

Gender	Age groups(yr)										Total
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	≥ 90	
Female	6	16	43	26	28	30	34	30	25	8	246
Male	3	2	2	3	6	5	17	9	3	4	54
Total	9	18	45	29	34	35	51	39	28	12	300

One hundred of the specimens were collected from patients attending inpatient wards while the remaining 200 were from different general practitioner's clinics. In Table 3.11, the distribution of patients by gender, specimen collection type and location of collection showed no significant different in male/female ratio or between isolates collected from Manchester and those from Preston. However, distribution of specimens according to collection site showed a significant difference (p<0.0001) with specimens collected from Preston appearing to be more commonly associated with patients attending general practitioner's clinics than was seen in the Manchester collection.

Various specimens were collected including catheter urine, mid-stream urine and directly collected urine. Mid-stream urines accounted for 66% of the specimens and catheter urine specimens accounted for approximately 6% of the collected specimens.

Although no significant association between different STs and patient's age groups was observed, the distribution of STs among other demographic parameters revealed some statistically significance associations. ST131 showed a stronger link to hospital acquired than community acquired infections with a significant statistical difference (18% vs. 10%, $P < 0.05$; Figure 3.17). Additionally, distribution of STs according to gender showed a significant association of ST69 with females (11% vs. 0%, $P < 0.05$; Figure 3.18).

Table 3.11 Distribution of patients according to gender, time and collection site

	2007		2009	
	M	F	M	F
	No.(%)	No.(%)	No.(%)	No.(%)
Manchester	11(11)	89(89)	22(22)	78(78)
Preston	10(20)	40(80)	11(22)	39(78)

	2007		2009	
	Hospital aq.	Community aq.	Hospital aq.	Community aq.
	No.(%)	No.(%)	No.(%)	No.(%)
Manchester	40(40)	60(60)	43(43)	57(57)
Preston	11(22)	39(78)	6(12)	44(88)

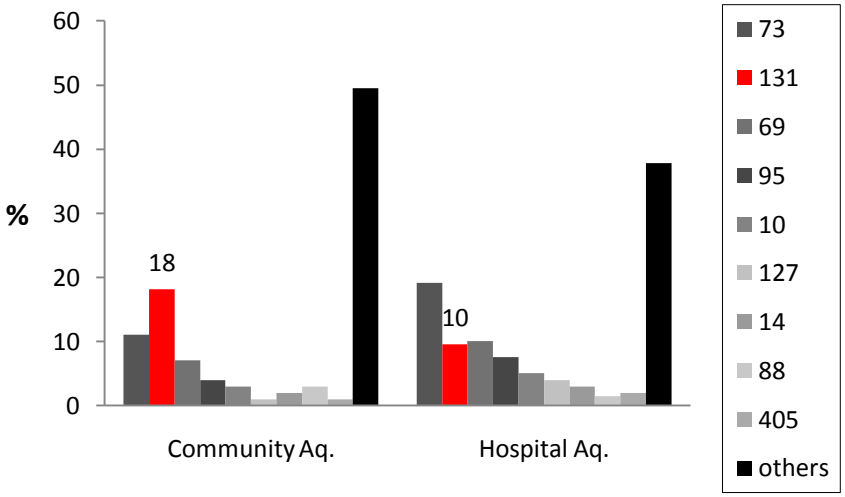


Figure 3.15 Distribution of common STs according to specimen collection site

.

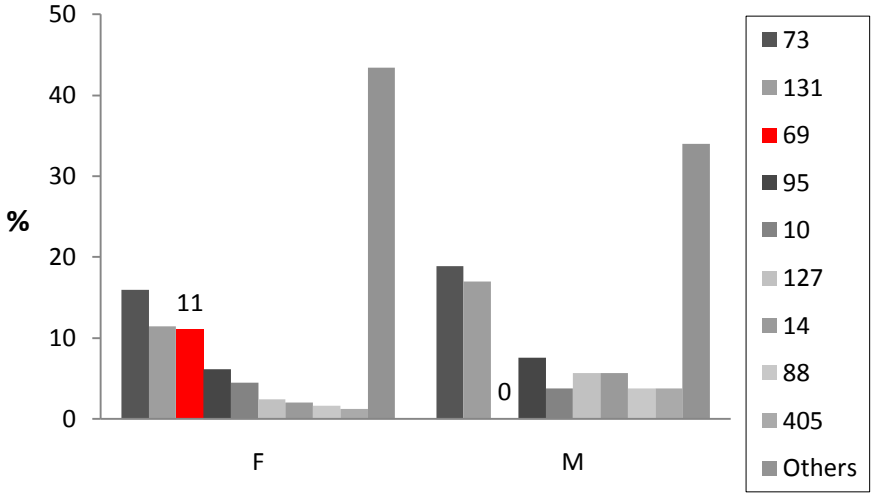


Figure 3.16 Distribution of common STs according to gender

3.9. Biochemical activity

Metabolic profiles were determined using the Vitek 2 ID-GNB card (Biomérieux), to test the organism's metabolic activity in 41 fluorescence based biochemical tests including 18 enzymatic tests, 18 fermentation tests, 2 decarboxylase tests and 3 other miscellaneous tests.

On average, the biochemical activities of the tested UPEC isolates were similar to those of *E. coli* reported by others (Ewing 1986; Farmer *et al.* 1985). A notable difference was observed concerning α-Galactosidase activity, where 69% of the tested isolates appeared to produce an α-Galactosidase enzyme compared to almost 100% reported by others (*P<0.0001*; Table 3.12).

3.9.1. Association between metabolic activities

Statistical analysis of association between results of different biochemical tests reveals complex associations with 52 positive associations and 26 negative associations involving wide range of tests. When the margin of confidence was increased and a *P* value of <0.001 considered as statistically significant, the associations were limited to 14 biochemical tests with 22 positive associations and 12 negative association, of which 70% occurred as overlapping associations between tests related to three types of metabolic reactions namely production of peptidases (tyrosine arylamidase and proline arylamidase), decarboxylases (ornithine decarboxylase and lysine decarboxylase) and alkalinisation of L- lactate and succinate. Additionally resistance to O129 was significantly associated with assimilation of sucrose and production of ornithine decarboxylase (Table 3.13).

Table 3.12 Biochemical activity profiles of UPEC studied here, compared to those previously reported by other groups (bold face type indicates major difference)

Test	% Activity by of UPEC	% Activity for *E. coli* as reported by	
		(Ewing, 1986)	(Farmer *et al*. 1985)
Ala-Phe-Pro-Arylamidase	0	ND	ND
Adonitol	0.7	5.6	5
L-Pyrrolydonyl-Arylamidase	0	ND	ND
L-Arabitol	0	ND	5
D-Cellobiose	0	2.4	2
Beta-Galactosidase	98.7	ND	95
H2O production	0	ND	1
Beta-N-Acetyl-Glycosaminidase	0	ND	ND
Glutamyl Arylamidse pNA	0	ND	ND
D-Glucose	100	100	100
Gamma-Glutamyl-Transferase	0.7	ND	ND
Fermentation/Glucose	96.7	ND	100
Beta-Glucosidase	0	ND	ND
D-Maltose	97.4	89.9	95
D-Mannitol	100	96.8	98
D-Mannose	100	ND	98
Beta-Xylosidase	0	ND	ND
Beta-Alanine Arylamidase pNA	0	ND	ND
L-Proline Arylamidase	9.3	ND	ND
Lipase	0	0	0
Palatinose	0	ND	ND
Tyrosine Arylamidase	36.4	ND	ND
Urease	0.7	0	1
D-sorbitol	92.1	93.4	94
Saccharose/Sucrose	49.7	48.9	50
D-Tagatose	7.3	ND	ND
D-Trehalose	99.3	98.8	98
Citrate	0	ND	1
Malonate	0.7	0	0
5-Keto-D-Gluconate	49	ND	ND
L-Lactate alkalinisation	23.8	ND	ND
Alpha-Glucosidase	0	0	0
Succinate alkalinisation	33.8	ND	ND
Beta-N-Acetyl-Galactosaminidase	0	ND	ND
Alpha-Galactosidase	**69**	**ND**	**99**
Phosphatase	6	ND	ND
Glycine Arylamidase	0	ND	ND
Ornithine Decarboxylse	78.8	64.2	65
Lysine Decarboxylase	95.4	88.7	90
L-Histidine assimilation	0	ND	ND
Courmarate	100	ND	ND
Beta-Glucoronidase	94	ND	ND
O/129 Resistance	60.9	ND	ND
Glu-Gly-Arg-Arylamidase	0	ND	ND
L-Malate assimilation	2	ND	ND
Ellman	96.7	ND	ND
L-Lactate assimilation	0	ND	ND

Table 3.13 Associations between metabolic activities of 300 isolates representing NW-UPEC population

Biochemical test		ProA	TyrA	URE	dSOR	SAC	dTAG	dTRE	MNT	5KG	ILATk	SUCT	AGAL	PHOS	GlyA	ODC	LDC	CMT	BGUR	O129R	IMLTa	ELLM
	NO.	28	162	10	284	167	19	297	1	123	120	128	208	36	7	227	280	297	275	175	7	292
TyrA	162	28																				
	(P value)	*0.0001*																				
URE	10	2	10																			
	(P value)	–	–																			
dSOR	284	25	152	10																		
	(P value)	–	–	–																		
SAC	167	20	68	6	162																	
	(P value)	–	*0.0001*	–	–																	
dTAG	19	5	13	0	18	13																
	(P value)	–	–	–	–	–																
dTRE	297	28	162	10	283	166	19															
	(P value)	–	–	–	–	–	–															
MNT	1	1	1	0	1	0	0	1														
	(P value)	–	–	–	–	–	–	–														
5KG	123	7	77	6	115	41	6	123	0													
	(P value)	–	–	–	–	*0.0001*	–	–	–													
ILATk	120	28	118	9	112	55	12	120	1	48												
	(P value)	*0.0001*	*0.0001*	–	–	–	–	–	–	–												
SUCT	128	27	118	10	120	55	13	128	1	58	100											
	(P value)	*0.0001*	*0.0001*	*0.0002*	–	*0.0002*	–	–	–	–	*0.0001*											
AGAL	208	28	142	10	201	106	15	208	1	96	106	118										
	(P value)	*0.0001*	*0.0001*	–	–	–	–	–	–	–	*0.0001*	*0.0001*										
PHOS	36	14	32	3	35	24	7	36	0	11	27	28	35									
	(P value)	*0.0001*	*0.0001*	–	–	–	–	–	–	–	*0.0001*	*0.0001*	*0.0001*									
GlyA	7	5	7	0	7	4	1	7	0	1	7	6	5	4								
	(P value)	*0.0001*	–	–	–	–	–	–	–	–	–	–	–	–								
ODC	227	17	108	9	216	123	13	226	0	91	73	86	165	22	3							
	(P value)	–	*0.0001*	–	–	–	–	–	–	–	*0.0001*	–	–	–	–							
LDC	280	20	145	10	266	153	14	278	1	115	107	115	193	28	6	217						
	(P value)	*0.0001*	–	–	–	–	–	–	–	–	–	–	–	*0.0009*	–	–						
CMT	297	28	162	10	282	165	17	295	1	123	120	128	208	36	7	225	278					
	(P value)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–					
BGUR	275	18	139	10	263	151	12	274	1	116	98	111	187	28	3	211	258	273				
	(P value)	*0.0001*	*0.0001*	–	–	–	*0.0003*	–	–	–	*0.0001*	–	–	–	–	–	–	–				
O129R	175	22	91	3	163	117	13	174	1	59	73	78	116	26	4	120	162	173	158			
	(P value)	–	–	–	–	*0.0001*	–	–	–	–	–	–	–	–	–	*0.0006*	–	–	–				
IMLTa	7	6	7	0	5	7	1	7	0	1	7	6	6	3	3	4	5	7	2	7		
	(P value)	*0.0001*	–	–	–	–	–	–	–	–	–	–	–	–	*0.0003*	–	–	–	*0.0001*	–		
ELLM	292	25	157	10	280	165	19	291	1	121	116	125	205	33	7	225	274	290	269	169	6	
	(P value)	–	–	–	*0.0003*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
ILATa	3	3	3	0	3	3	0	3	0	0	3	3	3	2	2	3	3	3	0	3	3	3
	(P value)	*0.0007*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*0.0005*	–	*0.0001*	–

Statistical analysis of association between biochemical tests using Fisher's exact test showing only $P \leq 0.001$ which reflects statistical significance; not shown are biochemical tests that yielded P values > 0.001.

3.9.2. Metabolic characteristics of common STs

To assess whether isolates of certain ST share specific metabolic characteristics, two analytical approaches were used. First, a non-parametric comparison analysis using Mann-Whitney U test was used to compare the aggregate bio-scores generated for each ST against all others. A significant association was observed for ST131 with high bio-scores when compared to all other STs ($P \leq 0.0001$), on the other hand ST127 showed an significant association with low bio-scores with a significant difference from the other STs combined ($P \leq 0.003$) (Table 3.14).

Table 3.14 Aggregate Bio-score in relation to STs

ST (No)	Aggregate Bio-score Median (range)	*P value*
ST10 (13)	0.55 (0.48-0.65)	*0.670*
ST14 (8)	0.59 (0.55-0.72)	*0.145*
ST69 (27)	0.55 (0.48-0.72)	*0.967*
ST73 (50)	0.59 (0.45-0.65)	*0.770*
ST88 (6)	0.52 (0.51-0.58)	*0.216*
ST95 (19)	0.55 (0.45-0.62)	*0.973*
ST127 (11)	0.52 (0.45-0.55)	***0.003***
ST131 (37)	0.62 (0.48-0.76)	***0.0001***
ST405 (5)	0.55 (0.41-0.69)	*0.771*

P values (Mann-Whitney U test) shows the significance of the different Bio-score of each ST compared to the other STs.

Secondly, the statistical association of individual metabolic activity with common STs was assessed using Fisher's exact test. Of the 41 metabolic activities tested, common STs were significantly associated with 11 tests. Remarkably, ST131 exhibited significant associations with seven tests, which indicates distinctive biochemical characteristics (Table 3.15) and ST73 had a significant association to four tests. In contrast, isolates of ST14 and ST405 showed no difference in their metabolic activity compared to the rest of the tested isolates. However, production of ornithine decarboxylase and assimilation of sucrose appear to be key characteristics in many

successful STs, as they are significantly associated with the most frequently isolated STs.

Table 3.15 Statistical analysis of association between STs and metabolic activity

		Biochemical test										
		ADO	**ProA**	**SAC**	**5KG**	**ILATk**	**AGAL**	**PHOS**	**ODC**	**BGUR**	**O129R**	**ILATa**
ST	**(No.)**	(6)	(28)	(167)	(123)	(120)	(208)	(36)	(227)	(275)	(175)	(3)
ST10	(13)	***0.0012***	–	–	–	–	–	–	***0.0001***	–	–	–
ST14	(8)	–	–	–	–	–	–	–	–	–	–	–
ST69	(27)	–	–	*0.0001*	*0.0017*	–	–	–	***0.0001***	–	–	–
ST73	(49)	–	–	***0.0001***	*0.0001*	–	–	***0.0014***	*0.0001*	–	–	–
ST88	(6)	–	–	–	–	–	–	–	–	–	–	–
ST95	(19)	–	–	***0.0001***	–	–	–	–	–	–	***0.0001***	–
ST127	(9)	–	–		–	–	–	–	–	–	–	–
ST131	(37)	–	*0.0001*	*0.0001*	***0.0001***	*0.0041*	*0.0041*	–	*0.0034*	*0.0002*	–	***0.0017***
ST405	(5)	–	–	–	–	–	–	–	–	–	–	–

P values (by Fisher's exact test) were calculated for each ST compared to the rest of the population shown only where $P \leq 0.005$. Bold *P* values are for negative associations.

3.10. Antibiotic susceptibility

In the present study, nearly 30% of the isolates were susceptible to all tested antibiotics, while the rest of the isolates demonstrated a wide range of resistance from resistance to just one antibiotic (11%) to resistance to 19 out of 21 tested antibiotics (< 1%). A reasonably high level of resistance was demonstrated to a number of commonly used antibiotics: ampicillin (55%); first generation cephalosporins (38%); trimethoprim (37%); and nalidixic acid (21%), as shown in Figure 3.19. Interestingly, isolates collected from Manchester appeared to be more resistant to most antibiotics tested than those from Preston, with significant different ($P < 0.05$) in resistance to many commonly used antibiotics such as trimethoprim, nalidixic acid and ciprofloxacin.

According to the generated antibiotic susceptibility profiles, different mechanisms of resistance were predicted against the major antibiotic families. Among the detected mechanisms of resistance against β-lactam antibiotics, acquired penicillinase (A.P) appeared to be the predominant mechanism (47%), while ESBL activity was seen in only 4% of the isolates. On the other hand, three patterns of enzymatic modification

resistance were detected against aminoglycosides mainly through two enzymes, N-Acetytransferase (AAC) and O-Adenyltransferase (ANT) and were collectively acquired by 11% of the isolates.

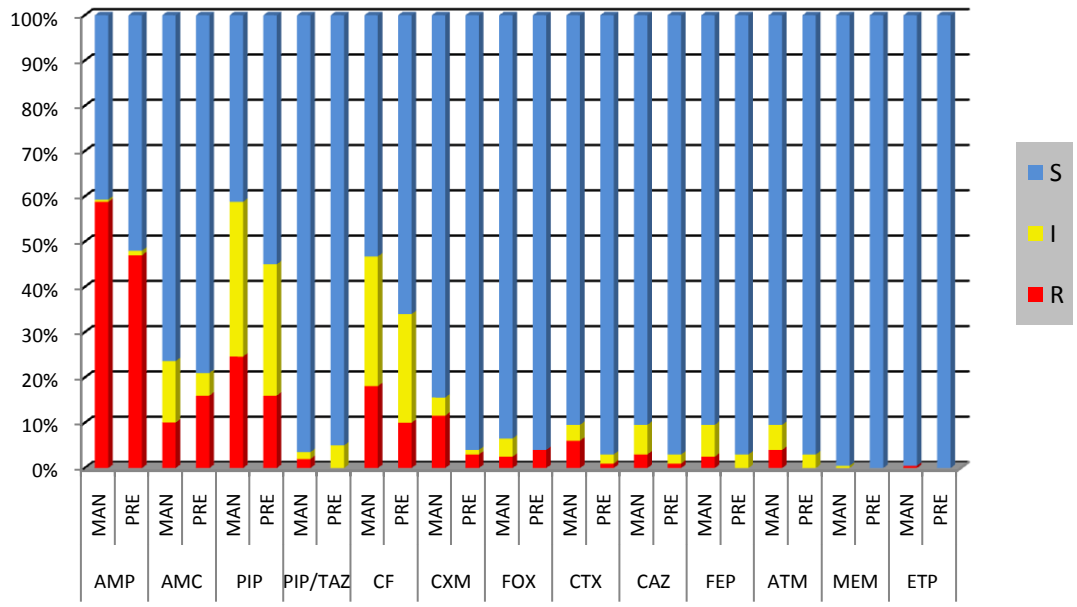


Figure 3.17A Distribution of antibiotic susceptibilities according to geographic site of isolation. AMP:ampicillin; AMC: amoxicillin/clavulanic acid; PIP: piperacillin; PIP/TAZ: piperacillin/tazobactam; CF: cefalotin; CXM: cefuroxime; FOX: cefoxitin; CTX: cefotaxime; CAZ: ceftazidime; FEP: cefepime; ATM: aztreonam; MEM: meropenem; ETP: ertapenem; S: susceptible; R: resistant; I : intermediate susceptibility.

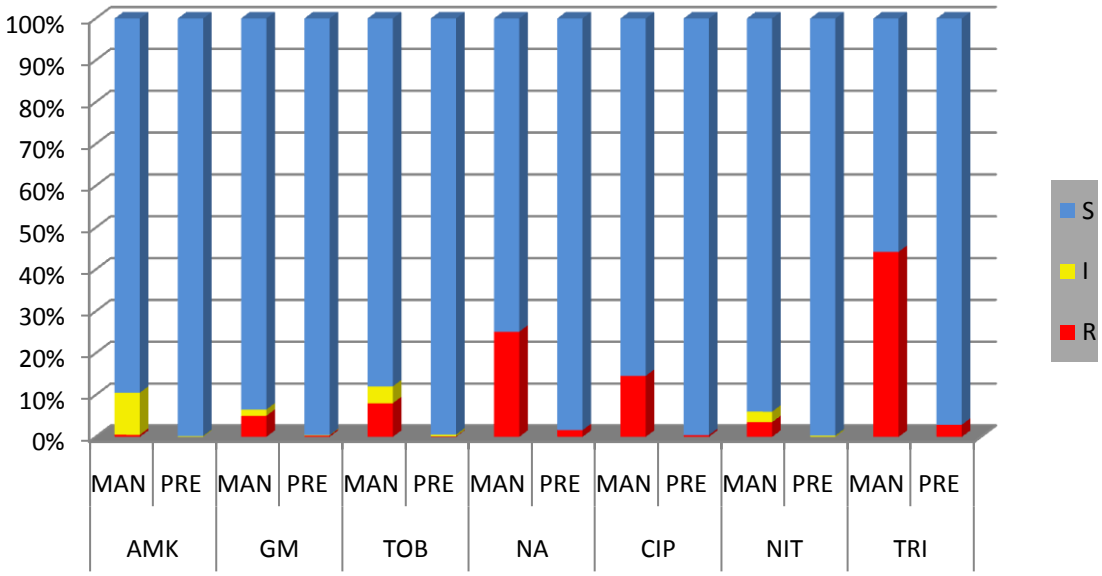


Figure 3.19 B Distribution of antibiotic susceptibilities according to geographic site of isolation. AMK: amikacin; GM: gentamicin; TOB: tobramycin; NA: nalidixic acid; CIP: ciprofloxacin; NIT: nitrofurantoin; TRI: trimethoprim; S: susceptible; R: resistant; I : intermediate susceptibility.

3.10.1. Association of STs with antibiotic resistance

Sequence Types and their susceptibility to commonly used antibiotics were assessed using Fisher's exact test. Overall, STs vary considerably in their antibiotic susceptibilities from the most susceptible, ST95, with the lowest resistance-scores (median 0.0 and range 0.0-0.29) to the most resistant, ST131, with the highest score (median 0.38 and range 0.0-0.9). Table 3.16 shows the distribution of antibiotic resistant isolates according to their ST and mechanism of resistance predicted by VITEK2 AES. ST131 showed a distinctive antibiogram that was significantly associated with resistance to most of the tested antibiotic families. This antibiogram is characterised by production of ESBLs and AAC(6) in addition to resistance to ciprofloxacin, nitrofurantoin and trimethoprim. These associations, when compared to all other STs, were all significant (*P*≤0.008*)*. Isolates of ST69 were characterised by resistance to trimethoprim (*P ≤ 0.0001*).

Table 3.16 Association of STs and resistance to different antibiotic families

		β-Lactams		Aminoglycosides			Quinolones		Furanes	Trimethoprim	Resistance-score
		A.P	ESBL	AAC(3)	AAC(3),ANT(2)	AAC(6)	PR	R	R	R	median(range)
ST	(No.)	(141)	(24)	(2)	(5)	(23)	(31)	(34)	(9)	(112)	
ST10	(13)	4	2	0	0	0	0	1	0	6	0.1(0.0-0.48)
P value		–	–	–	–	–	–	–	–	–	
ST14	(8)	4	0	0	0	0	3	0	0	3	0.07(0.0-0.19)
P value		–	–	–	–	–	–	–	–	–	
ST69	(27)	19	0	0	2	0	3	1	0	20	0.14(0.0-0.33)
P value		–	–	–	–	–	–	–	–	*0.0001*	
ST73	(49)	25	0	0	0	2	3	1	0	11	0.1(0.0-0.38)
P value		–	–	–	–	–	–	–	–	–	
ST88	(6)	4	0	0	0	0	2	0	1	2	0.14(0.05-0.43)
P value		–	–	–	–	–	–	–	–	–	
ST95	(19)	7	0	0	0	0	2	0	0	2	0.0(0.0-0.29)
P value		–	–	–	–	–	–	–	–	–	
ST127	(9)	5	1	1	0	1	0	0	0	1	0.1(0.0-0.67)
P value		–	–	–	–	–	–	–	–	–	
ST131	(37)	16	13	1	0	13	3	19	3	24	0.38(0.0-0.9)
P value		–	*0.0001*	–	–	*0.0001*	–	*0.0001*	*0.008*	*0.0004*	
ST405	(5)	1	2	0	0	1	1	1	0	3	0.1(0.0-0.9)
P value		–	–	–	–	–	–	–	–	–	

P values (by Fisher's exact test) shown only where *P≤0.005*.

3.11. Cluster analysis of STs based on biochemical profiles in relation to antibiotic susceptibility

Unweighted pair grouping method with arithmetic mean cluster analysis based on the biochemical profiles of isolates from the major STs revealed a large cluster comprised almost entirely of ST131 (19/21) and included 51% of the total ST131 isolates. The cluster was defined at the 70% similarity level and characterised by resistance to multiple antibiotics. The remaining ST131 isolates were found scattered through the dendrogram. Although there is a tendency for isolates with the same ST to cluster, only ST73 and ST69 gathered in clusters primarily consisting of their respective ST. Other than ST131, no correlation was found between the defined clusters and antibiotic resistance profiles (Figure 3.20).

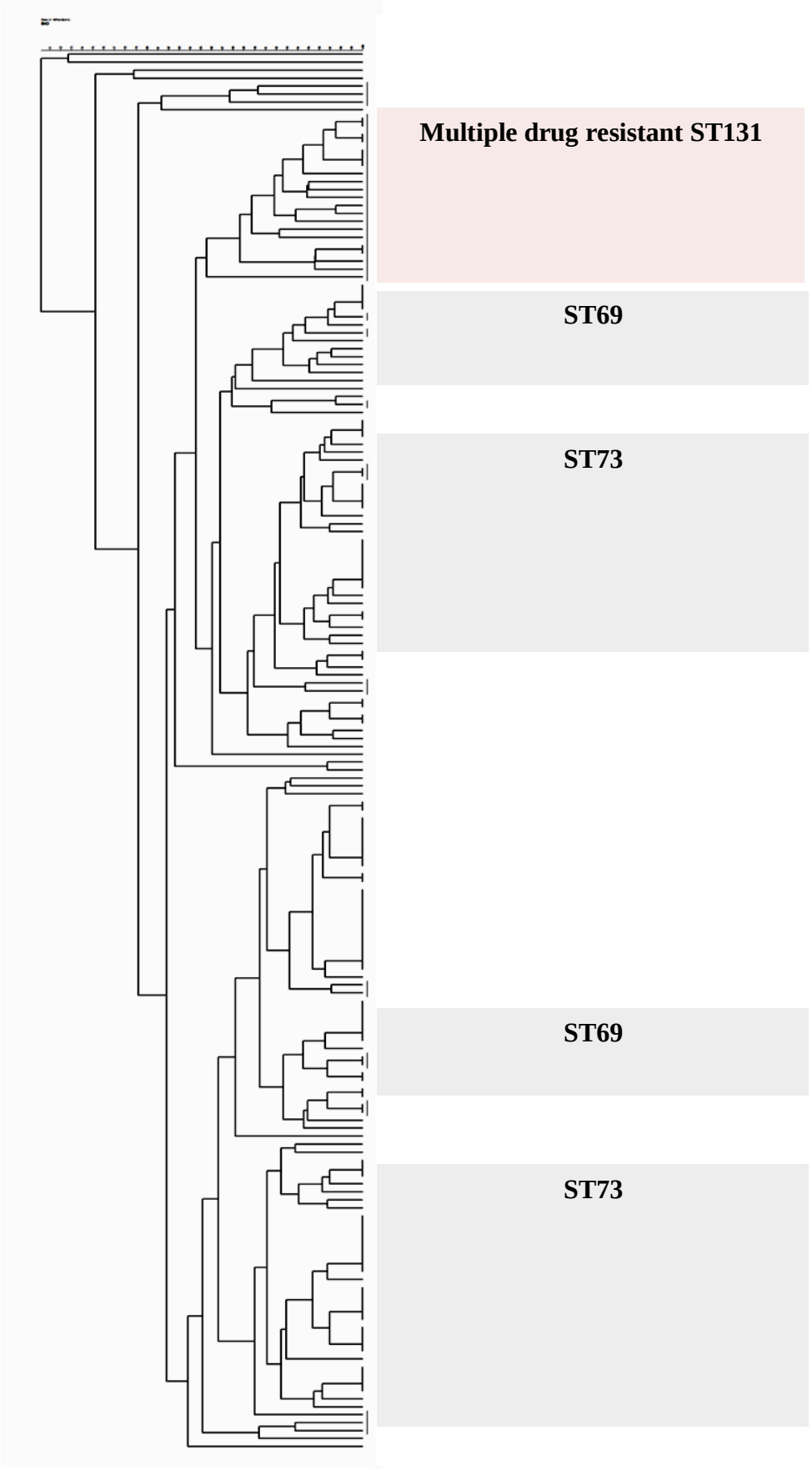


Figure 3.18 UPGMA Cluster analysis based on the biochemical profiles of major sequence types

3.12. Prevalence of virulence factors

Of the 29 virulence genes tested for, 26 were detected with a prevalence ranging from 0.3% (*nfaE*) to 97% (*fimH*). Among the adhesin genes various Pap elements occurred in more than 40% of the isolates. Although the toxin associated gene, *hlyA* was equally prevalent to *cnf1*, both were significantly more prevalent than the colicin V production gene *cvaC* (*P< 0.001*). Of the siderophores, *fyuA* (yersiniabactin) was significantly more frequent than *iut* (aerobactin; *P<0.0001*). Nearly 60% of the isolates were positive for the group II capsule synthesis gene with K5 accounting for 61% of *kpsM II* positive isolates. The serum resistance related gene (*traT*) occurred in 72% of the isolates while the pathogenicity island (*PAI*) marker gene was present in 60% of the isolates (Figure 3.21).

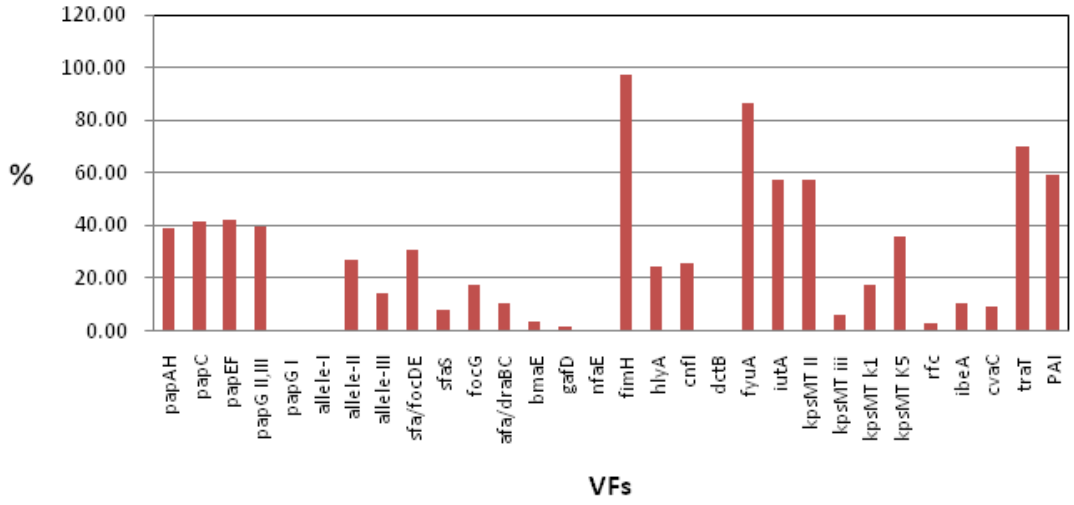


Figure 3.19 Prevalence of VFs genes among 300 NW-UPEC isolates

3.12.1. Distribution of VFs by STs

Genotyping of virulence traits revealed a strong correlation between VF profiles and STs. For frequently detected STs, each ST consisted of closely related VF profiles with similarity ranging from (65% to100%) which suggests vertical transmission of VFs. Overall, the STs varied considerably in VF contents, from ST10 with the lowest VF score (mean 3.7, range from 1 to 8), to ST127 with the highest VF score (mean 13.4 and range 9 to 17). Table 3.17 shows distinctive virulence traits that were significantly associated with the commonly detected STs. Although different Pap elements were important virulence determinants in many STs, lack of Pap elements appeared to significantly characterise some successful STs including ST131 and ST10. In addition to its deficit of Pap elements, ST131 isolates showed a significant lack of the Sialosyl-binding adhesion gene ($p<0.0001$), ST127 was significantly associated with *papG III* ($p<0.0001$) and, ST69 isolates were negatively associated with *PAI* marker ($P<0.0001$).

3.12.2. Phylogenetic distribution of VFs

Projection of the VF traits onto the phylogenetic background of the isolates revealed, as expected, an association of most VFs with B2 and D phylogenetic groups and to a lesser extent to group A, whereas none of the VF traits were found to be associated with group B1 (Figure 3.22). Although acquisition of VFs was more frequently associated with groups D and B2 and to a lesser extent group A, group A is the group with the greatest number of significant associations with VF's, with significant associations with 17 VFs, most of which are negative associations. Group D was significantly associated with Pap elements with frequency of occurrence ranging between 35% - 54% compared to 16% - 27% in other groups ($P \leq 0.0001$).

Table 3.17 Distinctive virulence characteristics of common clonal groups among UPEC

Category	specific VF	Total no. 300	prevalence (no.) of VF by clonal group					
			ST10 (no=13)	ST69 (no=27)	ST73 (no=49)	ST95 (no=19)	ST127 (no=9)	ST131 (no=37)
adhesion	papA	120	8 (1)	52 (14)	55 (27)	**89 (17)**	**100 (9)**	**8 (3)**
	papC	127	**0 (0)**	52 (14)	57 (28)	**100 (19)**	**100 (9)**	**5 (2)**
	papEF	129	**8 (1)**	**70 (19)**	55 (27)	**84 (16)**	**100 (9)**	**8 (3)**
	papGII	84	0 (0)	44 (12)	41 (20)	**84 (16)**	11 (1)	**5 (2)**
	papGIII	43	0 (0)	4 (1)	22 (11)	10 (2)	**100 (9)**	3 (1)
	sfa/foc DE	95	0 (0)	**4 (1)**	**84 (41)**	10 (2)	**77 (7)**	**0 (0)**
	afa/draBC	33	23 (3)	4 (1)	**0 (0)**	0 (0)	0 (0)	**43 (16)**
	sfaS	23	0 (0)	0 (0)	8 (4)	10 (2)	**55 (5)**	0 (0)
	focG	55	0 (0)	7 (2)	**61 (30)**	16 (3)	11 (1)	**0 (0)**
Toxins	hlyA	78	8 (1)	**0 (0)**	**69 (34)**	5 (1)	**67 (6)**	**8 (3)**
	cnf1	82	**0 (0)**	**0 (0)**	**69 (34)**	10 (2)	**77 (7)**	**8 (3)**
	cvaC	28	8 (1)	4 (1)	**0 (0)**	**42 (8)**	0 (0)	**0 (0)**
siderophore	fyuA	262	**38 (5)**	85 (23)	**96 (47)**	95 (18)	100 (9)	**97 (36)**
	iutA	177	54 (7)	70 (19)	65 (32)	47 (9)	**22 (2)**	**86 (32)**
capsule	kpsM II	177	**8 (1)**	55 (15)	61 (30)	**100 (19)**	**100 (9)**	57 (21)
	K1	53	0 (0)	7 (2)	**4 (2)**	**100 (19)**	0 (0)	**3 (1)**
	K5	109	**8 (1)**	51 (14)	43 (21)	**0 (0)**	**89 (8)**	**54 (20)**
Miscellaneous	ibeA	32	0 (0)	0 (0)	**0 (0)**	10 (2)	0 (0)	**24 (9)**
	traT	216	54 (7)	**89 (24)**	**57 (28)**	89 (17)	67 (6)	**86 (32)**
	PAI	181	**23 (3)**	**11 (3)**	**92 (45)**	**84 (16)**	78 (7)	**92 (34)**
VF score	mean,median,(range)		3.7,3(1-8)	7.8,9,(5-7)	11,11,(6-16)	11.8,12,(6-14)	13.4,13,(9-17)	6.9,6,4-14)

Bold values indicate significant associations ($P \leq 0.05$); whereas underline values indicates negative association

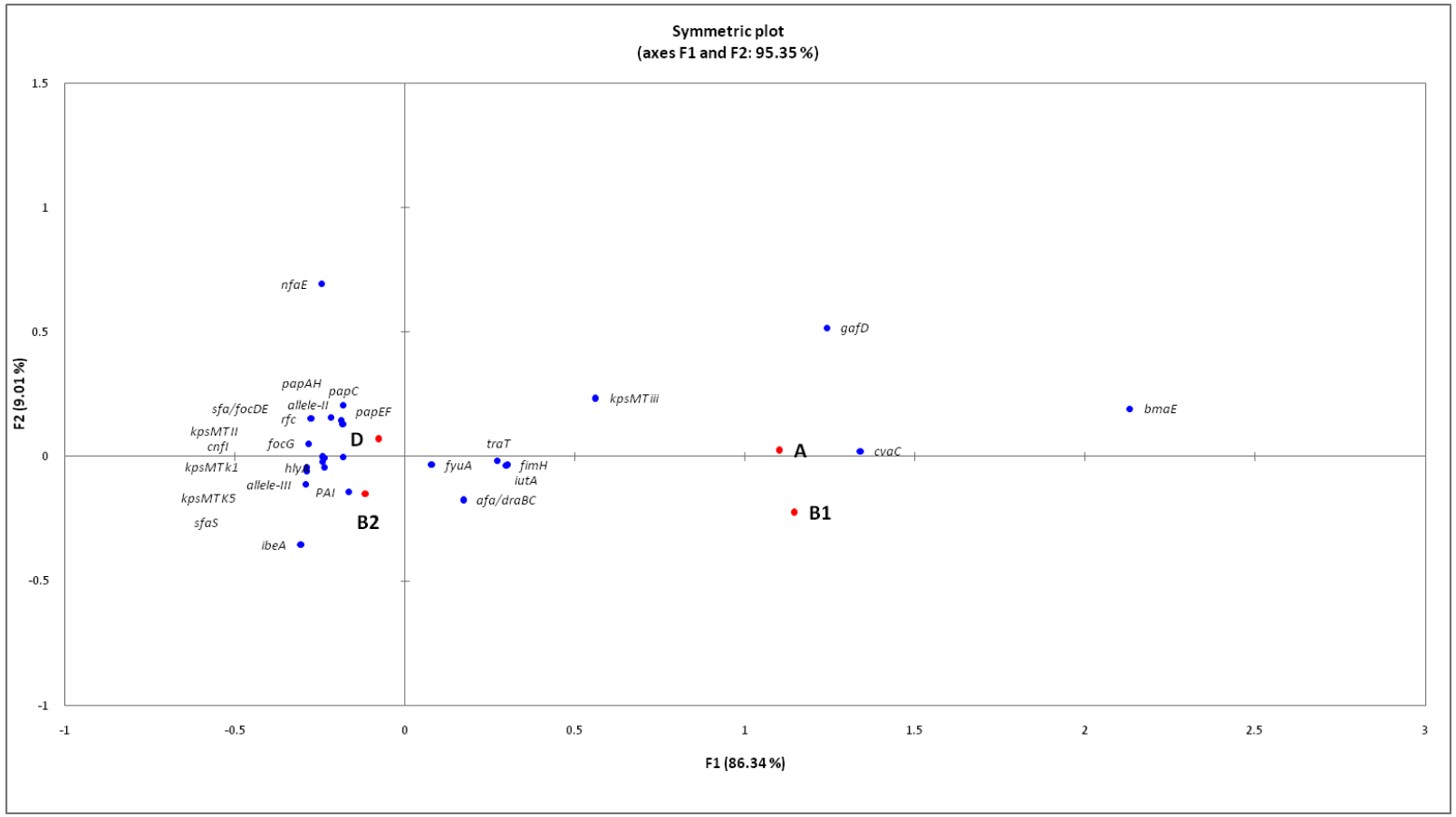


Figure 3.20 Graphical representation of the correspondence analysis of UPEC associated virulence traits based on phylogenetic types. The F1/F2 plane accounted for 95% of total variance, with factors F1 and F2 accounting for 86% and 9%, respectively.

Only three VFs (*fyuA, ibeA* and PAI) appeared to be significantly associated with group B2. Siderophore *fyuA* occurred in 100% of B2 isolates compared to 88% in group D, 12 % in group A and 1.7% in group B1. Finally, although the marker gene for pathogenicity island (PAI) was detected in 60% of group D isolates, it was positively associated with group B2 with a frequency of 84% compared to 50% in other groups ($P \leq 0.0001$) (Table 3.18). However, no significant differences were observed in the incidence of any virulence factors between phylogenetic groups within STs clones (Table 3.19).

Table 3.18 Distribution of VF genes in relation to phylogenetic group among UPEC isolates

Phy. gp	(No.)	VF																
		papAH	*papC*	*papEF*	*allele-II*	*allele-III*	*sfa/focDE*	*focG*	*bmaE*	*hlyA*	*cnfI*	*fyuA*	*kpsMT II*	*kpsMT k1*	*kpsMT K5*	*ibeA*	*cvaC*	*PAI*
		(117)	(124)	(126)	(81)	(43)	(92)	(53)	(11)	(74)	(78)	(259)	(173)	(53)	(108)	(31)	(28)	(177)
A	(36)	3	3	2	2	0	0	0	6	1	1	22	2	0	3	0	10	5
P value		*0.0001*	*0.0001*	*0.0001*	*0.0011*	*0.0042*	*0.0001*	*0.0008*	*0.0006*	*0.0004*	*0.0002*	*0.0001*	*0.0001*	*0.0008*	*0.0001*	*0.0351*	*0.0005*	*0.0001*
B1	(11)	0	0	0	0	0	1	0	1	0	0	3	0	0	0	0	2	1
P value		*0.0082*	*0.0033*	*0.0031*								*0.0001*						*0.0008*
B2	(78)	30	31	31	18	16	29	16	0	25	25	78	55	18	34	15	4	66
P value												*0.0001*				*0.0046*		*0.0001*
D	(175)	85	91	94	62	27	62	38	4	48	52	155	116	35	71	16	12	105
P value		*0.0001*	*0.0001*	*0.0001*	*0.0001*		*0.0042*	*0.0321*										

P values (by Fisher's exact test) shown only where $P \leq 0.05$.

Table 3.19 Distribution of virulence factors associated with common STs according to phylogenetic groups B2 and D

	ST14		ST73		ST95		ST127		ST131	
	B2 (3)	D (5)	B2 (16)	D (34)	B2 (9)	D (10)	B2 (4)	D (5)	B2 (24)	D (13)
papAH	2	4	6	21	8	9	4	5	3	0
papC	1	4	8	20	9	10	4	5	2	0
papEF	2	4	6	21	8	8	4	5	3	0
papG II,III	1	4	9	20	9	9	4	3	2	0
allele-II	1	3	6	14	7	9	1	0	2	0
allele-III	0	0	4	7	1	1	4	5	1	0
sfa/focDE	0	1	10	31	1	1	4	3	0	0
afa/draBC	0	0	0	0	0	0	0	0	10	6
nfaE	0	0	0	0	0	0	0	0	0	1
fimH	3	4	16	33	9	10	4	5	24	12
hlyA	1	3	11	23	1	0	2	4	3	0
cnfI	0	0	9	25	1	1	4	3	3	0
fyuA	3	4	16	31	9	9	4	5	24	12
iutA	1	4	12	20	5	4	1	1	22	10
kpsMT II	3	4	10	20	9	10	4	5	14	7
kpsMT k1	0	0	2	0	9	10	0	0	0	1
kpsMT K5	3	2	6	15	0	0	4	4	13	7
ibeA	1	0	0	0	1	1	0	0	6	3
cvaC	0	0	0	0	4	4	0	0	0	0
traT	2	3	9	19	8	9	3	3	19	13
PAI	3	4	15	30	7	9	2	5	22	12

3.13. Association between VFs and antibiotic susceptibility

The relationship between antibiotic resistance and carriage of VFs was statistically examined and *P* values <0.005 used to indicate significant relationships. None of the adhesion related genes, such as Pap elements or S/F1C fimbriae genes, appeared to have a positive association with any antibiotic resistance traits. The aerobactin receptor gene (*iutA*) was positively associated with resistance to ampicillin and piperacillin as well as to nalidixic acid, ciprofloxacin and trimethoprim. The serum resistance gene (*traT*) on the other hand shows an association with ampicillin and piperacillin in addition to trimethoprim. Overall, correlation analysis reveals that resistance to advanced antibiotics such as third generation cephalosporins or fluoroquinolones is significantly associated with absence of virulence capacities starting with adhesion, as isolates resistant to those antibiotics are negatively associated with the presence of different adhesion related genes though *iutA* and *traT* remain appreciably prevalent (Table 3.20).

3.14. Interrelation of metabolic, virulence and antibiogram profiles among UPEC

To explore the interrelation between the three characteristics, Pearson's correlation coefficients were calculated for the three possible relationships that can be drawn from these characteristics using the aggregation score of each. The correlation coefficients ranged between (0.0459 – 0.1476) which indicate weak correlations despite the significant statistical differences *(<0.0001, 0.001* and *0.005*) observed for Bio-score vs. AST, VF vs. AST and VF vs. Bio-scores, respectively (Figure 3.23).

Table 3.20 Association of resistant to antibiotics and various VFs traits in UPEC isolates

Antibiotic		VFs										
		papAH	*papC*	*papEF*	*papG -III*	*sfa/focDE*	*sfaS*	*afa/draBC*	*hlyA*	*cnfI*	*iutA*	*traT*
	(No.)	(117)	(124)	(126)	(43)	(92)	(23)	(32)	(74)	(78)	(173)	(211)
AMP	(166)	59	59	67	15	43	9	23	35	38	109	129
P value					*0.0046*						*0.0022*	*0.0022*
AMC	(68)	19	19	23	6	21	3	15	18	18	46	51
P value								*0.0013*				
PIP	(162)	56	56	64	13	41	8	23	34	36	108	126
P value					*0.0009*						*0.0007*	*0.0024*
CF	(127)	35	38	38	10	31	7	18	27	28	79	93
P value		*0.0005*	*0.0006*	*0.0004*								
CXM	(35)	6	6	7	3	4	2	7	3	5	23	25
P value			*0.0017*									
CTX	(22)	2	2	3	1	1	1	6	1	2	16	17
P value		*0.0025*	*0.0012*			*0.0035*						
CAZ	(22)	2	2	3	1	1	1	6	1	2	16	17
P value		*0.0025*	*0.0012*			*0.0035*						
FEP	(22)	2	2	3	1	1	1	6	1	2	16	17
P value		*0.0025*	*0.0012*			*0.0035*						
ATM	(22)	2	2	3	1	1	1	6	1	2	16	17
P value		*0.0025*	*0.0012*			*0.0035*						
NA	(64)	15	17	21	1	5	0	14	6	5	52	51
P value		*0.0039*			*0.0004*	*0.0001*		*0.0024*		*0.0001*	*0.0001*	
CIP	(33)	5	5	8	1	2	0	8	3	3	27	25
P value		*0.0024*	*0.0012*			*0.0005*					*0.0027*	
TRI	(111)	32	35	38	5	18	2	22	13	15	83	91
P value					*0.0001*	*0.0001*	*0.0028*	*0.0002*	*0.0001*	*0.0001*	*0.0001*	*0.0007*

P values (by Fisher's exact test) shown only where *P≤0.005*. AMP:ampicillin; AMC: amoxicillin/clavuanic acid; PIP: piperacillin; CF: cefalotin; CXM: cefuroxime; CTX: cefotaxime; CAZ: ceftazidime; FEP: cefepime; ATM: aztreonam; NA: nalidixic acid; CIP: ciprofloxacin; TRI: trimethoprim.

Although no significant correlations were observed between metabolic scores and virulence score or resistance and antibiotic score, correlation between AST vs. Bio scores revealed an appreciable correlation that justified further investigation.

Subsequent examination of the correlations among each ST revealed a weak but significant positive association between Bio-score and AST-score (*P≤0.01*) among ST131 isolates that could explain the correlation observed between Bio-score and AST-score in the general population (Table 3.21).

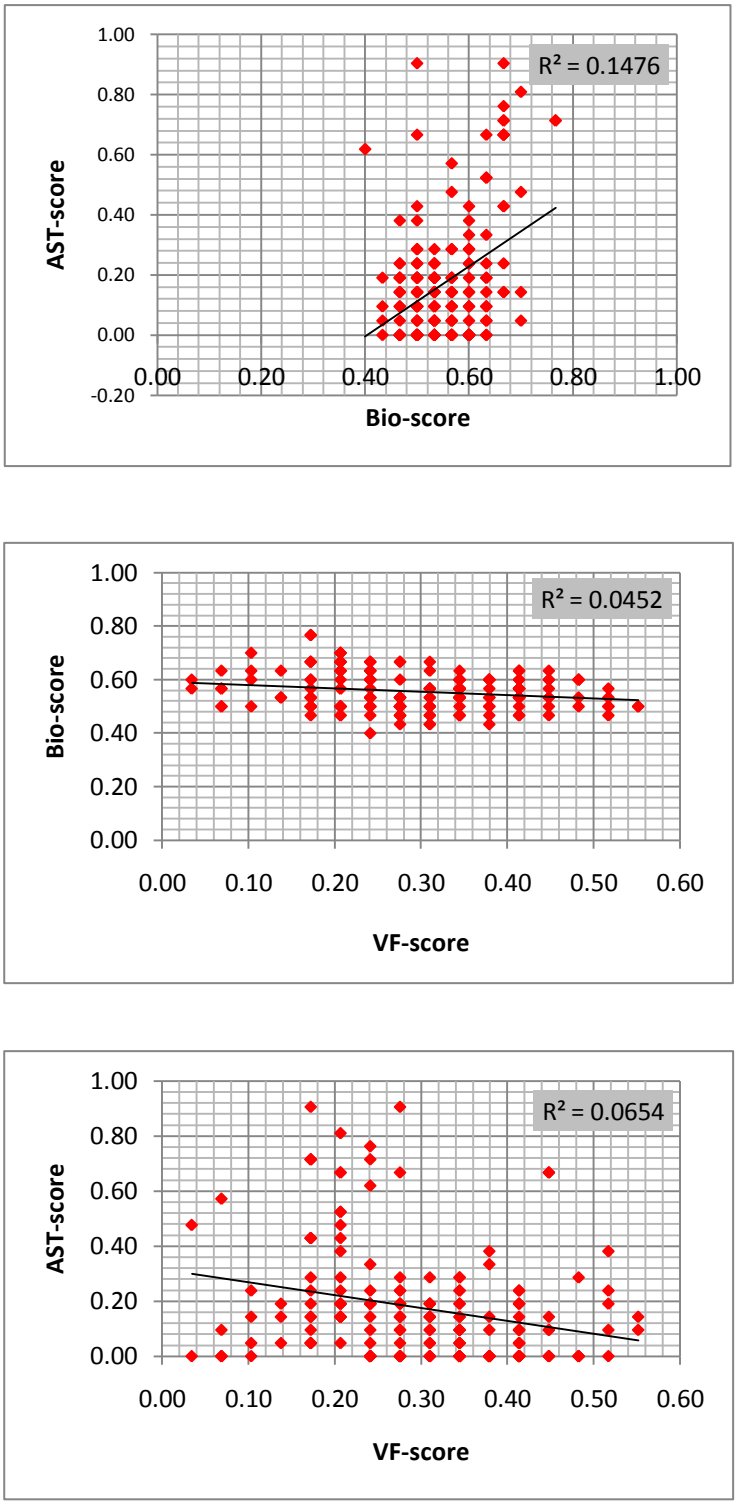


Figure 3.21 Pearson's correlation coefficient for the three possible relationships that can be drawn from phenotype characteristics using the aggregation score of each.

Table 3.21 Correlation between different characteristics and different STs

ST	Correlation coefficients R^2			statistical difference *P value*			correlation significance *P value*		
	Bio vs. VF	Bio vs. AST	AST vs. VF	Bio vs. VF	Bio vs. AST	AST vs. VF	Bio vs. VF	Bio vs. AST	AST vs. VF
ST10	0.0091	0.00	0.0352	—	—	—	—	—	—
ST14	0.1975	0.1123	0.1132	—	—	—	—	—	—
ST69	0.066	0.0041	0.01	—	—	—	—	—	—
ST73	0.0016	0.0018	0.0121	—	—	—	—	—	—
ST88	0.1837	0.0132	0.0366	—	—	—	—	—	—
ST95	**0.2279**	0.0325	0.0318	*0.04*	—	—	—	—	—
ST127	**0.267**	**0.1032**	0.0559	—	—	—	—	—	—
ST131	0.025	**0.3689**	0.0579	—	*<0.0001*	—	—	*< 0.01*	—
ST405	**0.0816**	**0.0608**	0.1062	—	—	—	—	—	—

P values (by Fisher's exact test) shown only where $P \leq 0.05$. Bold indicates positive correlation

3.15. Characterisation of the CTX-M-15 O25b ST131 clonal group and identification of factors predictive for this lineage.

Phenotypic and genotypic screening of various traits and their distribution among different STs revealed significant associations between ST131 and many either metabolic or virulence factors, which raises the necessity for further examination of ST131 in particular.

Of 24 isolates with an ESBL phenotype, 15 (62%) contained $bla_{CTX\text{-}M}$ alleles as demonstrated by PCR with the universal primers used (Figure 3.24). All of these isolates then yielded products of the expected size (approx. 415bp) with group 1-specific primers (Figure 3.25). Thirteen of the 15 isolates typed as group 1 CTX-M were of ST131.

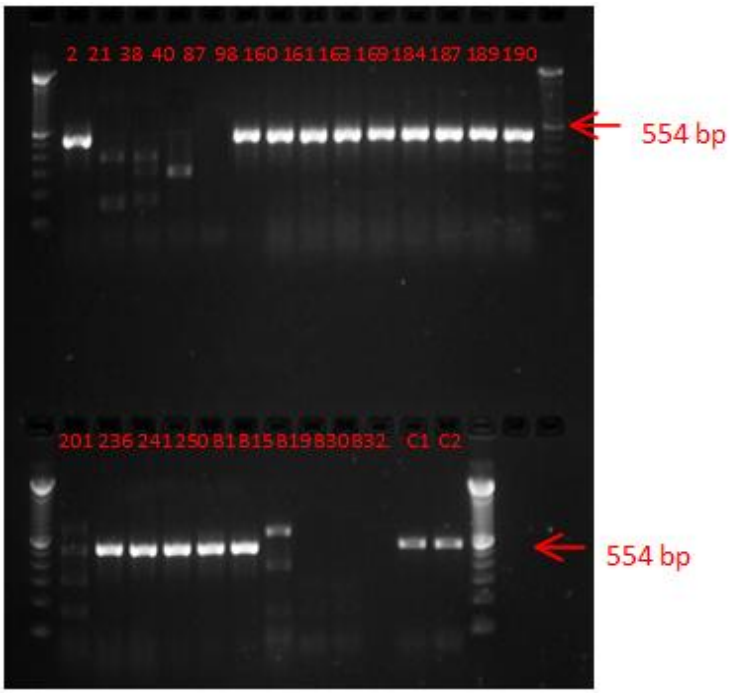


Figure 3.22 PCR screening for universal CTX-M primers (M1 & M2) , 100bp ladder in first and last lane of each row, positive isolates generated a product at 554bp compared to control strains C1 & C2 (known CTX-M positive isolates).

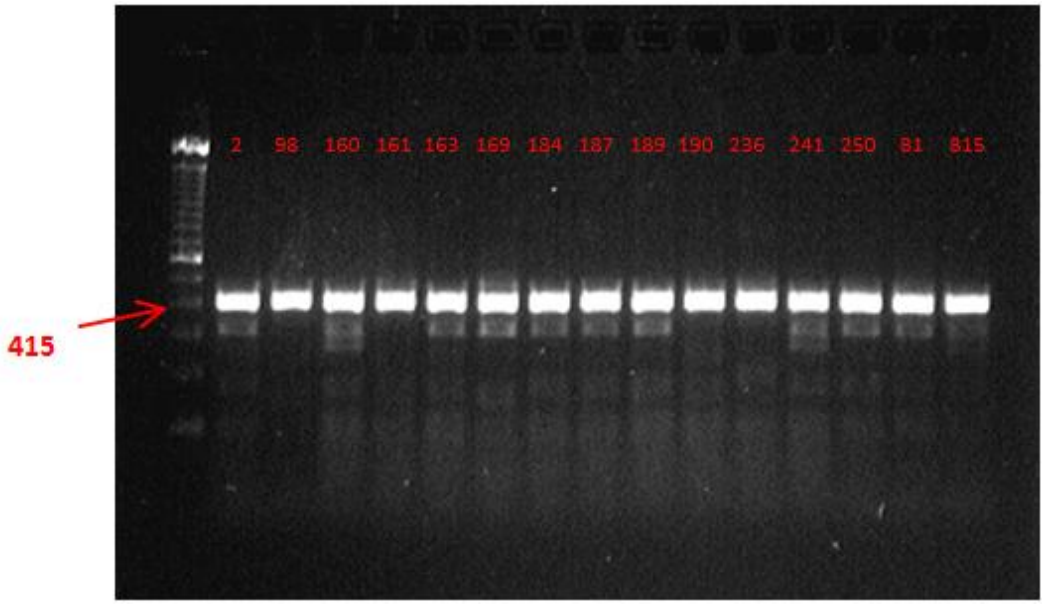


Figure 3.23 PCR screening for the group 1 CTX-M genes, 100bp ladder in first lane, positive isolates yielded a product at 415bp.

3.15.1. Prevalence of the CTX-M-15 O25b clone of ST131

All of the 13 ST131 typed as CTX-M (group 1) contained CTX-M-15 as demonstrated by the specific primers used. Of these, the CTX-M-15 O25b clone accounted for 10 isolates, which represents 29% of the total ST131 isolates (Figure 3.26).

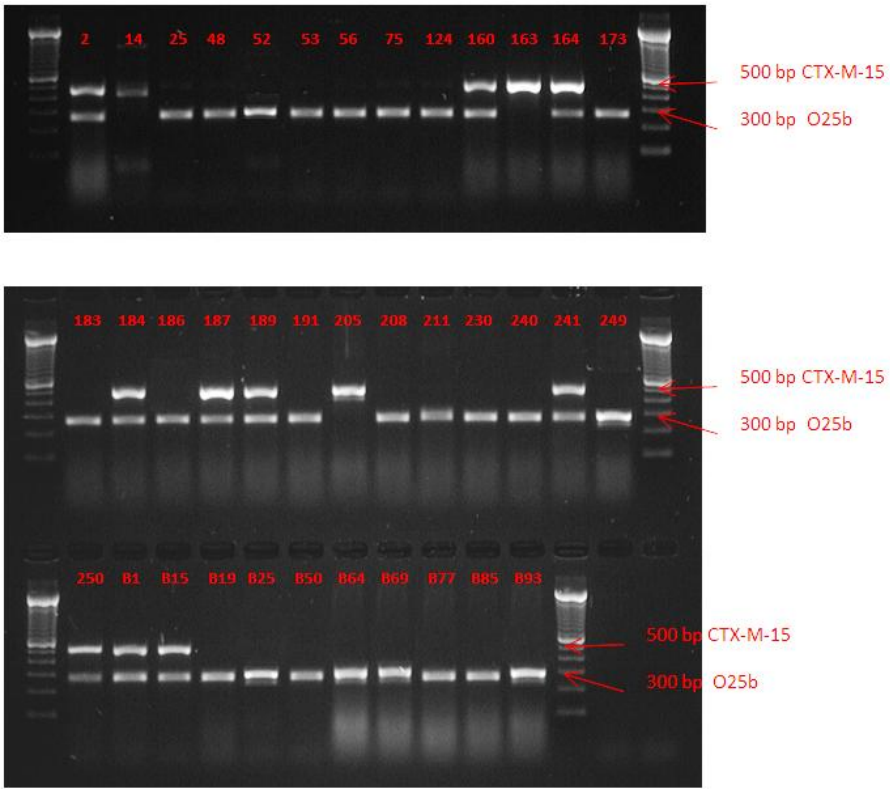


Figure 3.24 Multiplex PCR screening for O25b-ST131 clone producing CTX-M-15. 100bp ladder in first and last lane of each row, positive isolates yielded products at 500bp and 300bp

A nine-digit numerical code (three tests per digit) was generated using the results of 27 biochemical tests. A total of 149 different profiles were detected, of which 19 were solely associated with ST131 isolates. Figure 3.27 shows the distribution of ST131 biochemical profiles against profiles of other STs. Interestingly ST131 was divided into 2 groups, susceptible and multidrug resistant, where multidrug resistant isolates and those of CTX-M-15 O25b were both significantly associated with high bio-score profiles.

Cluster analysis of a simple matching coefficient based on the biochemical profiles was used to generate a dendrogram describing the association of ST131 and antibiogram profiles. This revealed a large ST131 cluster defined at 88% similarity level and comprised of 22 (60%) isolates that were characterised as having a multiple resistance antibiogram (Figure 3.28).

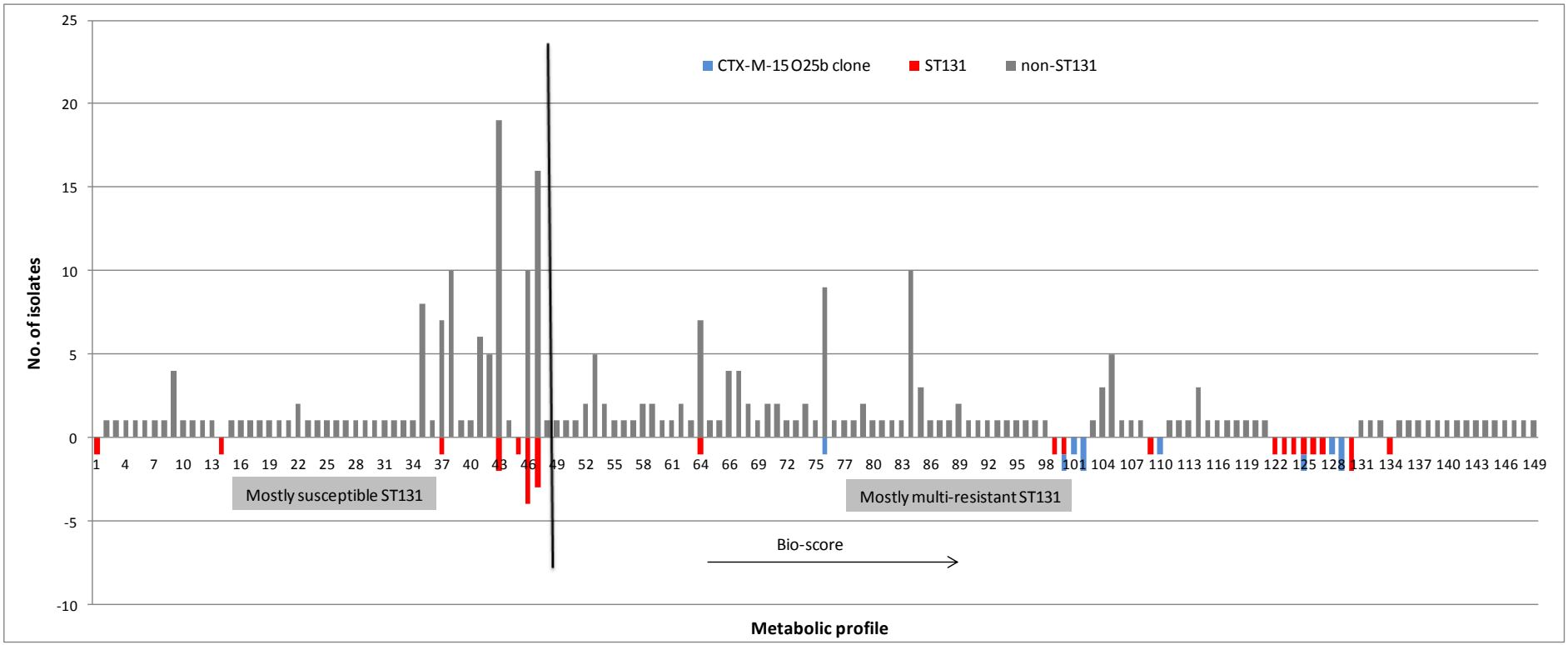


Figure 3.25 Distribution of metabolic profiles of ST131 isolates against other UPEC STs. ST131 displayed at the negative side of x axis (left of the black line crossing the x axis) were predominantly susceptible and those on the right side, resistant (as projected later by cluster analysis)

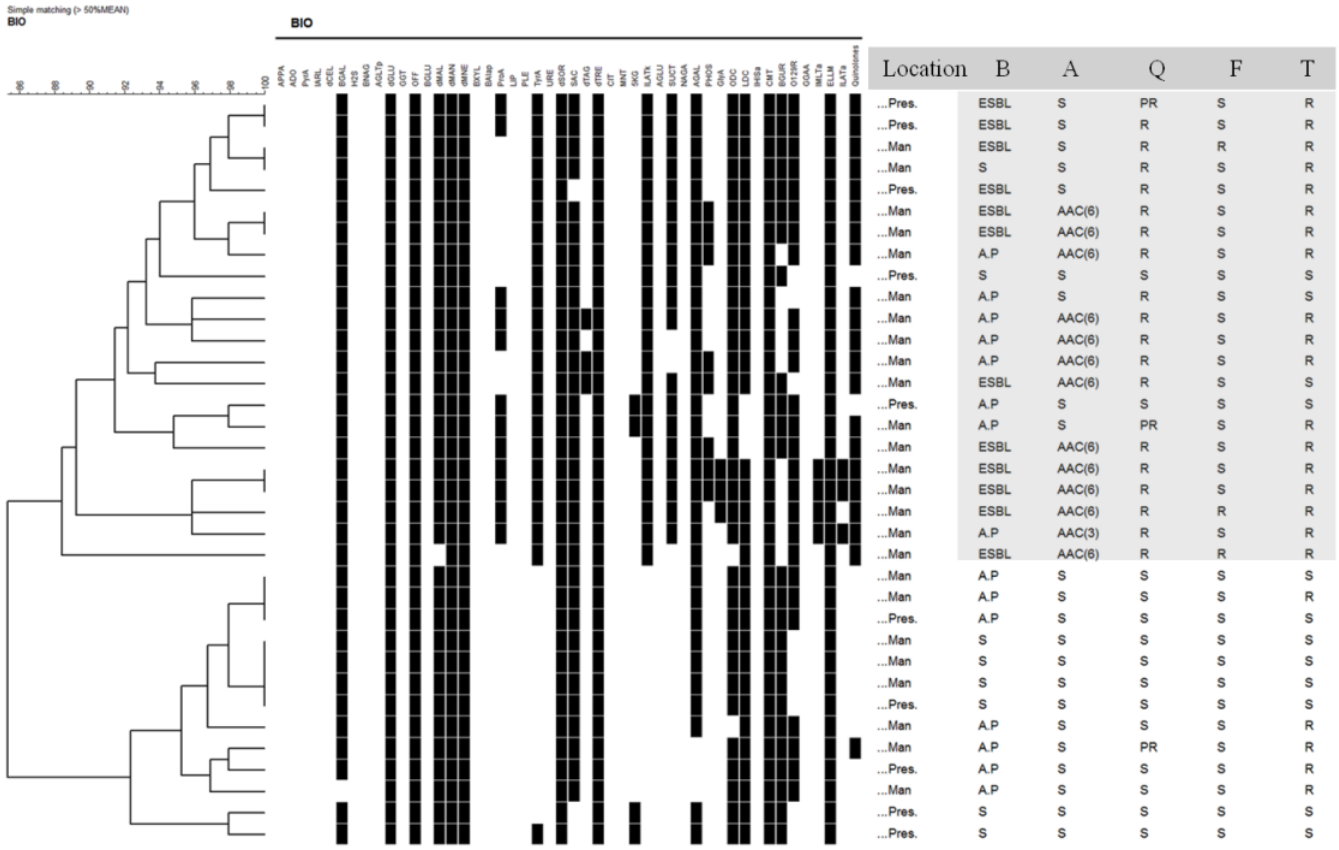


Figure 3.26 Biotype profiling of ST131 isolates, cluster analysis based on metabolic profiles. Codes above the column correspond to substrate code, predicted mechanism of resistance: B; β-lactam, A; aminoglycosides, Q; quinolones, F; furanes, T; trimethoprim, A.P acquired penicillinase, R; resistant, S; susceptible, PR; partial resistance (resistant to nalidixic acid only).

To investigate whether the observed distinctive characteristics of ST131 are general characteristics of the clone or they are only the result of O25b-CTX-M-15 ST131 clonal spread, a comparison analysis of incidence of these characteristics in O25b-CTX-M-15 ST131 and non-O25b-CTX-M-15 ST131 compared with non-ST131 isolates was carried out. Although most of the characteristics are general characteristics of the ST131 clone, some biochemical traits were significantly associated with O25b-CTX-M-15 ST131 such as *TyrA, ILATK, PHOS, SUCT* ($P<0.005$). Notably, other than resistance to fluoroquinolones, there were no significant differences in antibiotic resistance between non-O25b-CTX-M-15 ST131 and non-ST131 isolates (Table 3.22). However, statistical analysis of the distribution of different virulence factors according to quinolone and fluoroquinolone resistance among the ST131 clone showed that other than *ibeA*, which is significantly associated with susceptible isolates, no significant differences were observed in the incidence of other virulence factors (Table 3.23).

Table 3.22 Characteristics of *E. coli* ST131 isolates compared to non-ST131 isolates

	ST131		*P value*	Non-ST131	*P value*
	O25b-CTX-M-15 (11)	Non-O25b-CTX-M-15 (26)		(263)	
BGAL	11	25	–	261	–
OFF	11	25	–	254	–
dMAL	10	26	–	246	–
ProA	4	9	–	15	*0.0001*
TyrA	11	13	*0.0032*	138	–
SAC	10	24	–	133	*0.0001*
dTAG	1	2	–	16	–
5KG	0	4	–	119	*0.0001*
ILATk	11	12	*0.0022*	97	*0.0042*
SUCT	10	10	*0.0043*	108	–
AGAL	11	22	–	175	*0.0042*
PHOS	7	1	*0.0002*	28	–
GlyA	2	1	–	4	–
ODC	10	25	–	192	*0.0034*
LDC	10	24	–	246	–
BGUR	7	20	–	248	*0.0001*
O129R	10	18	–	147	–
IMLTa	2	2	–	3	–
ILATa	2	1	–	0	*0.0018*
papAH	2	1	–	114	*0.0001*
papC	2	0	–	122	*0.0001*
papEF	2	1	–	123	*0.0001*
papG II,III	1	1	–	116	*0.0001*
allele-II	2	0	–	79	*0.0007*
allele-III	0	1	–	42	–
afa/draBC	5	11	–	16	*0.0001*
nfaE	0	1	–	0	–
fimH	11	25	–	257	–
hlyA	2	1	–	71	–
cnfI	2	1	–	75	*0.0082*
fyuA	11	25	–	223	–
iutA	11	21	–	141	*0.0001*
kpsMT II	4	17	–	152	–
kpsMT k1	0	1	–	52	*0.0096*
kpsMT K5	4	16	–	88	–
ibeA	0	9	*0.0357*	22	*0.007*
traT	9	23	–	179	–
PAI	11	23	–	143	*0.0001*
ESBL	10	2	*0.0001*	*12*	*0.0001*
aac(6)	8	5	*0.0062*	*10*	*0.0001*
Fluoroquinolone	10	9	*0.003*	*15*	*0.0001*
Furantoin	2	1	–	0	–
Trimethoprim	9	14	–	88	*0.001*
B2	7	17	–	53	*0.0001*
D	4	9	–	162	*0.0039*

P values (by Fisher's exact test) are shown where $P < 0.05$ for virulence factors and phylogenetic groups and $P < 0.005$ for metabolic reactions and antibiotic resistant.

Table 3.23 Distribution of VFs according to Quinolone and Fluoroquinolone susceptibility among the ST131 clone

	Quinolone and Fluoroquinolone susceptibility of ST131 isolates					
	Q-R (22)	Q-S (15)	*P* value	FQ-R (19)	FQ-S (18)	*P* value
papAH	2	1	*NS*	2	1	*NS*
papC	2	0	*NS*	2	0	*NS*
papEF	2	1	*NS*	2	1	*NS*
papG II,III	1	1	*NS*	1	1	*NS*
allele-II	2	0	*NS*	2	0	*NS*
allele-III	0	1	*NS*	0	1	*NS*
sfa/focDE	0	0	*NS*	0	0	*NS*
afa/draBC	8	8	*NS*	6	10	*NS*
nfaE	0	1	*NS*	0	1	*NS*
fimH	22	14	*NS*	19	17	*NS*
hlyA	2	1	*NS*	2	1	*NS*
cnfI	2	1	*NS*	2	1	*NS*
fyuA	22	14	*NS*	19	17	*NS*
iutA	21	11	*NS*	19	13	*NS*
kpsMT II	9	12	*NS*	9	12	*NS*
kpsMT k1	1	0	*NS*	0	1	*NS*
kpsMT K5	9	11	*NS*	9	11	*NS*
ibeA	1	8	***0.0012***	0	9	***0.0004***
cvaC	0	0	*NS*	0	0	*NS*
traT	17	15	*NS*	14	18	*NS*
PAI	21	13	*NS*	18	16	*NS*
VF score	0.22	0.25		0.22	0.24	

P values (by Fisher's exact test) are shown where $P < 0.005$; *NS:* not significant

3.15.2. Mechanisms of fluoroquinolone resistance in ST 131 UPEC

Of the 37 tested ST131 isolates, 22 (57%) showed resistance to quinolones (nalidixic acid) from which 19 (90%) expressed additional resistance to fluoroquinolones (ciprofloxacin; Table 3.24). Among the quinolone resistant isolates, ESBL production was detected in 11 (58%) isolates.

Table 3.24 Characterisation of quinolone resistance mechanisms found in ST 131 UPEC isolates

Quinolone profile (no)	ESBL (%)	QRDR				*aac(6')-lb-cr* (%)	*qnr* genes
		gyrA		*parC*			
		(S)83(L) (%)	(D)87(N) (%)	(S)80(N) (%)	(E)84(V) (%)		
Q-S (15)	0	0	0	0	0	0	0
Q-R (3)	0	3(100)	0	0	0	0	0
FQ-R (19)	11(58)	19 (100)	19 (100)	19 (100)	19 (100)	9 (47)	0

Q-S: quinolone susceptible; Q-R: resistant to quinolones alone; FQ-R: resistant to fluoroquinolones

All ciprofloxacin resistant isolates had multiple mutations in both *gyrA* and *parC* genes, while the three nalidixic acid resistant isolates that failed to show resistance to ciprofloxacin all had single *gyrA* mutation at codon 83 (*Ser—Leu*).

On the other hand only nine isolates carried the *aac (6')-lb-cr* gene representing 47% of ciprofloxacin resistant isolates and 81% of the ciprofloxacin resistant ESBL producing isolates (Figures 3.29 and 3.30). Six of the nine *aac (6)-lb-cr* positive isolates that were resistant by mutation in both *gyrA* and *parC* belonged to the CTX-M15-O25b ST131 clone. Of the 20 isolates tested, none carried any of the *qnr* genes.

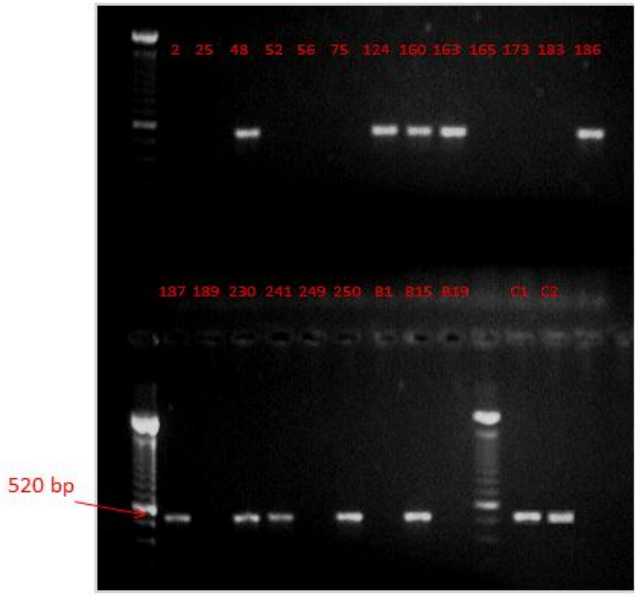


Figure 3.27 PCR screening for the *aac (6')-lb* gene100bp ladder in first lane of each row, positive isolates yielded a product at 520bp compared to control strains C1 & C2 (known positives)

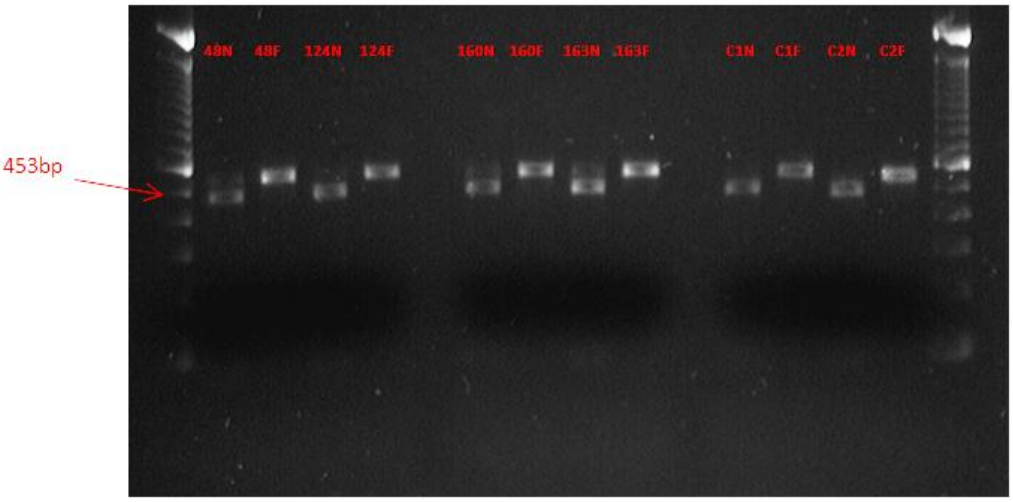


Figure 3.28 RFLP digestion of 4 isolates positive to *aac(6')-lb* native gene. Digestion with *NdeI* producing two fragments (453bp and 66bp) indicates modified variant *aac(6')-lb-cr*, two positive controls were run in parallel with each RFLP digestion

3.16. Multivariate logistic regression analysis

Stepwise multivariate logistic regression analysis was used to identify independent biochemical test or antibiotic susceptibility or virulence traits as predictors of ST131 and, among the ST131 clonal group, to predict the CTX-M-15/O25b sub-clonal group with $P<0.05$ considered to be the significance threshold.

In the first model**,** resistance to fluoroquinolones was the strongest predictor of ST131 ($P<0.0001$; OR 42) followed by carriage of three virulence factors PAI, *ibeA*, *afa/draBC* and finally lack of *sfa/focDE*, as shown in Table 3.25. The overall accuracy of this model to predict ST131 is 93% (positive predictive value 57%, negative predictive value 98.5%).

In the second model, carriage of ESBL enzymes with ($p<0.0001$; OR 204) appeared to be the only predictor of CTX-M-15 producing O25b-ST131 strains ($P= 0.0002$; OR, 120) within ST131. The overall accuracy of this model is 92% (positive predictive value 91%; negative predictive value 92%)

Table 3.25 Factors predictive of ST131/O25b-ST131 as identified using multivariate logistic regression analysis

Prediction factors					
	variables	*P* value	OR	95.0% C.I.for OR	
				Lower	Upper
1st model	CIP	0.0000	42.9	10.7	172.0
	sfafocDE	0.9900	0.0	0.0	0.0
	afadraBC	0.0001	13.3	3.7	48.1
	ibeA	0.0003	11.8	3.1	44.3
	PAI	0.0000	23.6	5.2	106.6
2nd model	ESBL	0.0002	120.0	9.7	1478.4

	Observed		Predicted		
			ST131		
			N	P	Percentage Correct
1st model	ST131	N	255	4	98.46
		P	16	21	56.76
	Overall Percentage				93.24
			CTX-M-15 O25b		
2nd model	CTX-M-15 O25b	N	24	2	92.31
		P	1	10	90.91
	Overall Percentage				91.89

OR: odds ratio; CIP: resistance to ciprofloxacin; N: negative; P: positive.

Discussion

4. Discussion

4.1. Multilocus sequence typing and phylogenetic analysis

UPEC have long been recognised as distinct clones of *E. coli*, which exhibit specific characteristics such as virulence associated traits, distinctive O antigens, genotypes and multidrug resistance (Zhang *et al.* 2000). Molecular epidemiological studies have revealed a clonal structure characterising *E. coli* and clarifying the genetic linkages between different lineages (Goldberg, Gillespie & Singer 2006; Tartof *et al.* 2005).

However, despite the increased use of MLST as a trusted epidemiological tool, the population structure of UPEC has been poorly studied using this technique, as most of the studies conducted were limited either by a bias towards certain characteristics, such as antimicrobial resistance and serogroup, or included a limited number of strains, which gave a false impression of overrepresentation of certain STs.

The house keeping genes selected in any MLST scheme should ideally encode proteins that are under stabilising selection for conservation of metabolic function (Maiden 2006). Within the current study, for the scheme used, the ratios of non-synonymous (*dn*) to synonymous (*ds*) changes of all target genes were less than 1, which is normally taken to signify that the population is subject to stabilizing selection, validating their suitability as epidemiological markers. Although, dN/dS has been used widely to analyse MLST data derived from bacterial populations, recent reports suggested that it may be impossible to infer selective pressure from dN/dS obtained from a single population (Kryazhimskiy & Plotkin 2008; Rocha *et al.* 2006).

However, the degree of sequence diversity within the housekeeping genes used in the current study was relatively high with polymorphic sites ranging from 5.0%-16.2%,

which agrees with various previous phylogenetic analyses of *E. coli* that show extensive allelic variation (Reid *et al.* 2000; Wirth *et al.* 2006).

The global availability of MLST databases enables the characterization of bacterial populations on a much wider scale. MLST has been used to characterize different *E. coli* pathotypes (Adiri *et al.* 2003; Noller *et al.* 2003; Petersen *et al.* 2009; Tartof *et al.* 2005). Although sequence type-pathotype associations do exist, allele distribution tells a different story with alleles shared by different STs and furthermore by different pathotypes (Wirth *et al.* 2006). In this study, although the majority of the identified UPEC associated alleles were found in both the ExPEC and EC databases, a degree of difference in prevalence was observed. The number of alleles with a significant difference in prevalence within the three groups was fewer when comparing the ExPEC database to the UPEC one (n=3), than in the comparison between EC and UPEC isolates (n=48). This was not unexpected given that UPEC are part of the ExPEC group. when three alleles that showed a significant difference in incidence between UPEC and ExPEC were compared, two showed a significant association with UPEC compared to ExPEC (*fum*C39, *mdh*17) and tracing these alleles to their possible STs showed an association with successful UPEC STs; *fum*C39 with ST10 and *mdh*17 with ST14, ST73 and ST95. Whether this finding holds any epidemiological significance towards understanding their success as UPEC will require further investigation.

MLST analysis of the studied population revealed a consistent profile of STs that occurred repeatedly in the current collections. It consisted primarily of ST73 (16%) followed by ST131 (13.3%), ST69 (9%), ST95 (6.3%), ST10 (4.3%), ST127 (3.6%), ST14 (2.6%) and ST405 (1.6%). ST131 and ST69 (CGA) have previously been extensively reported for their role as antimicrobial resistant clones that spread universally (Johnson *et al.* 2009; Nicolas-Chanoine *et al.* 2008; Tartof *et al.* 2005) and

recent reports have associated ST 405 with the global spread of CTX-M-15 and other ESBL enzymes (Jones *et al.* 2008; Nicolas-Chanoine *et al.* 2008) and ST95 with avian pathogenic *E. coli* (Mora *et al.* 2009). ST73 was previously reported as an important and diverse clone within the ECOR group B2 and was associated with UPEC (Zdziarski *et al.* 2008) and formed the main clone detected in the present collection (16%). Other successful STs (ST127 and ST80) in the current panel have never been reported as remarkable pathogens.

However, eBURST analysis of allelic profiles reinforces the status of ST73 and with its related STs it forms a large clonal complex that involves nine STs and accounts for 20% of the studied population. Spratt and colleagues suggested that the relative age of a CC can be estimated by how many SLVs it has and the presence of related subgroups (Feil *et al.* 2004). Using these criteria, ST73 is considered to be the oldest clonal complex in the population. Using the same characterisation, CC-ST405 and CC-ST69 can be considered as relatively newly evolved clonal complexes. The failure to cluster with any other ST in the population and the formation of only conservative clonal complexes when expanding the analysis to involve other ExPEC and EPEC isolates in the MLST database indicates that ST131 and ST127 have also emerged as newly evolved STs.

UPGMA analysis resolved the population into two large clusters, which correlates well with previously reported results using other molecular methods (Garcia-Martinez *et al.* 1996). However, the evolutionary divergence provided by clonalframe analysis emphasised the clonal distribution generated by UPGMA analysis, suggesting a well defined structure within the *E. coli* population, rather than vague structures often generated by recombination sensitive phylogenetic methods such as the neighbour-joining method (Jaureguy *et al.* 2008). Whether this clear-cut segregation within the

population reflected specific phenotypic characteristics of each cluster is not clearly established. Although no significant association between these clusters and biochemical characteristics was noted, similar to the situation mentioned in an earlier study (Garcia-Martinez *et al.* 1996), significant correlations were detected between some phenotypic characteristics and frequently detected STs. The natural over representation of these STs within each cluster could explain the correlation reported in the earlier study (Garcia-Martinez *et al.* 1996).

4.2. Pulse-field gel electrophoresis

To ensure the absence of unrecognised localised point-source outbreaks that may lead to overrepresentation of specific clonal groups in the current isolate collection, the clonal background of representative isolates of the common STs were assessed using PFGE. The profile similarity threshold approximately corresponds to a specific number of band differences, which in turn may correspond to a specific number of genetic events (Avery *et al.* 2002). Choosing the threshold value that accurately identifies the similarity of isolates is an area of continuing disagreement. Some investigators consider that isolates differing by three or less bands should be regarded as related strains (Louie *et al.* 1999), whereas others suggest that those with three or more band differences in their PFGE profiles should be considered unique strains (Tenover *et al.* 1995). Using the published Tenover *et al*. guidelines (Louie *et al.* 1999; Tenover *et al.* 1995), in the current study strains were defined as having a clonal relationship if they possessed 85% or greater similarity, corresponding to the occurrence of a single genetic event. Of the nine tested isolates belonging to ST73, only two isolates exhibited similar PFGE profiles, while the 10 isolates of ST69 exhibited eight different PFGE profiles. This

diversity in PFGE profiles indicates the absence of overrepresentation of any clone as a result of focal outbreak.

Although its primary function is to identify outbreaks involving a single clone, PFGE was able to cluster clonal groups. Each group exhibited a PFGE profile with similarity values between 60% and 65%, which is consistent with previous findings involving isolates of ST131, the O15:K52:H1 clonal group and CGA where PFGE was able to cluster isolates of each clonal group and their PFGE profile similarity values ranged between 66% to 74% (Johnson *et al.* 2009).

Interestingly, despite the considerable similarity of PFGE profiles observed among isolates of ST131 included in the current study, remarkable genetic similarity was detected that defined Preston isolates within the ST131 clone, suggesting ongoing sub-clonal evolution. This could either be due to abnormally slow accumulation of PFGE changes among the Preston ST131 isolates or abnormally rapid (point source-like) dissemination of the ST131 sub-clone, or both (Blackwood *et al.* 1997).

4.3. Virulence profiling

The ability of different strains of *E. coli* to cause different clinical syndrome is attributed to the cumulative impact of one or several key properties or virulence factors carried by these strains. For instance, carriage of genes for the K1 capsule allows evasion of neonatal host humoral immunity, whereas strains carrying genes for pyelonephritis associated pili (PAP) adhere well to urinary tract epithelial cells (Johnson 1991).

Although previous urovirulence associated genotypic studies have been conducted mainly on bacteraemic UTI associated strains, the prevalence of virulence factors

screened in this study was in general agreement with reported findings (Johnson *et al.* 2005d; Johnson & Stell 2000; Ramos *et al.* 2010).

The high prevalence of *fyuA* observed among *E. coli* isolates confirms previously reported findings involving isolates collected from patients with cystitis, pyelonephritis and prostatitis and 50 isolates from the stools of healthy adults, among which *fyuA* was significantly more prevalent in disease causing isolates than among faecal control strains (Kanamaru *et al.* 2006a). The *traT* gene was also found to be considerably prevalent in the UPEC isolates in the current collection. Similar findings were previously reported, though the prevalence of *traT* in UPEC compared to faecal isolates was not always significant. Experimental evaluation of *traT* as a potent urovirulence associated VF, is needed (Johnson *et al.* 2005d; Moreno *et al.* 2008; Ramos *et al.* 2010).

Several genes associated with the acquisition and development of UTIs are encoded on gene blocks known as pathogenicity islands (PAIs) that provide a mechanism for combined horizontal transfer of VF genes (Oelschlaeger *et al.* 2002).

The PAI investigated in the current study was previously identified within the chromosome of highly virulent uropathogenic *E. coli* CFT073 and appears to be found significantly more often in strains associated with UTI than in faecal strains (Kao *et al.* 1997). The high prevalence of the PAI marker observed in this study is consistent with previous reports involving other clinical isolates (Guyer, Kao & Mobley 1998; Johnson & Stell 2000).

The low prevalence or complete absence of some VF genes, for example *cdtB*, *bmaE*, *gapfD* and *nfaE*, suggested that production of the encoded proteins is not an important virulence property of UPEC strains. Although cytolethal toxin production associated the gene *cdtB* is primarily associated with enteric *E. coli*, previous report suggest an

association of *cdtB* with particular clones of bacteraemia associated *E. coli* (Johnson & Stell 2000).

Among the adhesin genes, the *pap* gene family is normally found among strains causing UTI and these are significantly associated with pyelonephritis (Johnson 1991). Previous studies have reported that the prevalence of different *pap* genes ranged from 50% up to 79% (Johnson *et al.* 2001; Leflon-Guibout *et al.* 2008). Notably, in the current study the prevalence of different *pap* genes was around 40%, which is relatively low and could be related to the increase in the prevalence of clones characterised by absence of *pap* genes, such as the ST131 clone in the UPEC population in the North West region of England.

Association of specific characteristics with particular genomic backgrounds can be retained over the short term predominantly by vertical transmission (Jaureguy *et al.* 2008). However the broad range of VF genes screened here allowed the generation of VF patterns that were significantly associated with different STs.

In the ST73 clone, a major clone within the studied population, only half of the isolates were able to express P fimbriae, which are required for colonisation and invasion of the upper urinary tract. Generally, more than 80% of UPEC strains express P fimbriae and the proportion of strains that express P fimbriae varies considerably from a high of 70% among isolates from patients with pyelonephritis to 36% among cystitis isolates and 24 % among those from asymptomatic bacteriuria (ABU) (Johnson 1991).

Interestingly, over 60% of the ST73 isolates expressed F1C fimbriae that mediate specific adherence to the collecting ducts and the distal tubules of the human kidney, as well as to renal tubulues cells (Antao, Wieler & Ewers 2009). This compares to previously reported figures of 14 to 30 % among general UPEC strains (Zdziarski *et al.*

2008), most of which were associated with pyelonephritis (Johnson *et al.* 2005d). Moreover, among the studied STs, F1C fimbriae were significantly associated with ST73 isolates, which suggests that F1C fimbriae may play an important role in the success of the ST73 clone.

For the screened toxins, about 69% of the ST73 clone members carried genes encoding haemolysin (*hlyA*). All of these isolates simultaneously encode cytotoxic necrotizing factor type1 (CNF1). For general UPEC isolates, previous reports estimated prevalence of haemolysin production from as low as 20% among isolates from ABU patients to 40% among isolates from cystitis patients and up to 50% among isolates from pyelonephritis patients (Johnson 1991; Kerenyi *et al.* 2005). Indeed, epidemiological studies have consistently shown that UPEC strains that make CNF1 also produce Hly (Antao *et al.* 2009). Moreover, Landraud *et al.* provided a genetic explanation for the association of CNF1 and Hly production by UPEC; they demonstrated that a cnf_1 gene and an *hly* operon in the prototypic UPEC strain J96 are not only co-transcribed, but are also co-regulated (Landraud *et al.* 2003).

The contribution of haemolysin and CNF1 production to urovirulence of ST73 is supported by the results of epidemiological studies associating the severity of UTIs with toxin production by UPEC isolates (Smith *et al.* 2008).

Isolates of ST95 have frequently been described as Avian pathogenic *E. coli* (APEC) responsible for avian colibacillosis in domesticated and wild birds (Mora *et al.* 2009) and have also been reported to cause neonatal meningitis in human (Bert *et al.* 2010). In the current study, all ST95 isolates harboured the K1 capsular antigen. However, the association of ST95 with K1 capsular antigen strongly linked it to NMEC strains and pathogenicity.

Most ST95 clone members harboured P fimbriae and, to a lesser extent, other types of adhesins, which agreed with previous reports on ST95 isolates collected from patients with bacteraemia (Bert *et al.* 2010). Notably, over 40% of ST95 isolates harboured the *cvaC* gene, which encodes production of and resistance to colicin V and has been associated with pyelonephritis (Johnson 1991). However, among the screened STs harbouring the *cvaC* gene, it appears to be significantly associated with ST95 in agreement with previous findings that show the *cvaC* gene exclusively occurred in ST95 isolates from bacteraemic patients (Bert *et al.* 2010). The unique virulence profiles associated with ST95, which consist of genes encoding significant adherence elements, production of *cvaC* and syntheses of capsule might explain its association with invasive UTIs and could explain its status as a successful UPEC clone.

Only a few reports mention ST127, either as a UPEC (Lau *et al.* 2008b) or as ExPEC with zoonotic potential (Johnson *et al.* 2008). Johnson and colleagues, in their study of the virulence and phylogenetic background of serogroup O6 isolates from humans and animals, revealed a strong association of ST127 and serogroup O6 and most ST127 isolates show a significant association with Pap elements especially *papG* allele III, S fimbriae and cytotoxin (*cnf1*) production (Johnson *et al.* 2008).

In the current study, ST127 clone members showed great similarity in VF profiles to those reported previously. Moreover, all ST127 isolates harboured capsular antigen K5 that provides a non-immunogenic shield protecting the bacteria. As a result, it is associated with *E. coli* causing serious invasive infections (Herias *et al.* 1997). Collectively, the VF associated with ST127 proposes a clone with potent virulence capabilities. ST127, which is suggested by phylogenetic analysis to be a newly evolved clone, holds the highest virulence score of all lineages in the current collection. This may permit its survival in the population long enough to gain antibiotic resistance.

Isolates of ST10 have frequently been identified among faecal isolates from healthy individuals and community acquired enterotoxigenic strains (Valverde *et al.* 2009), which could explain the lack of ExPEC virulence traits among ST10 clone members in the current study.

The emergent clonal group ST69 corresponds to *E. coli* CGA, which has been associated with outbreaks of trimethoprim-sulfamethoxazole resistant isolates in different communities (Johnson *et al.* 2009; Manges *et al.* 2008; Tartof *et al.* 2005). ST69 has been also associated with bacteraemia of non-urinary tract origin (Manges *et al.* 2006) and community acquired pneumonia in a renal transplant recipient (Johnson & Russo 2002). However, VF profiles observed among the ST69 clone members in the current collection only showed lower UTI potential, as only half of the isolates possessed P fimbriae and almost lack any other adhesins or toxins associated with upper UTI. Most notably, compared to other STs isolates of ST69 were negatively associated with the PAI gene. Similar results were reported by Johnson and colleagues where none of the CGA isolates harboured the pathogenic island marker (Johnson *et al.* 2009).

ST131 is best known for its association with production of ESBL resistance, especially the emergence of CTX-M-15 ESBL (Clermont *et al.* 2008; Coque *et al.* 2008; Lau *et al.* 2008a; Nicolas-Chanoine *et al.* 2008). As resistance comes at a cost to fitness, it is not surprising that the multidrug resistant ST131 clone is defined by a low virulence score and distinctive VF profiles that lack fimbrial adhesins, but significantly concentrated non-fimbrial adhesins of the Dr family and the gene encoding the invasion of brain endothelium phenotype (*ibeA*) that is responsible for neonatal meningitis in humans (Germon *et al.* 2005).

In general, ST131 VF profiles were consistent with those of organisms causing lower UTI and agreed with previous reports (Bert *et al.* 2010; Johnson *et al.* 2010b; Johnson *et al.* 2009). Although biofilm formation was not investigated in the current study, some of ST131 VFs may indicate potential biofilm production. Previous reports show some associations between biofilm-formation potential and some virulence-associated genes (Kanamaru *et al.* 2006b; Martinez-Medina *et al.* 2009; Ong *et al.* 2008). For example, the adhesin-coding gene *sfa/foc*DE and the *ibeA* gene, required for invasion in meningitis/sepsis-associated *E. coli* (MNEC), were more frequently detected amongst biofilm producers. However, consistent with results presented here, several studies have suggested that multiply resistant strains are usually less virulent than susceptible strains (Houdouin *et al.* 2006; Johnson *et al.* 2003a; Moreno *et al.* 2006).

The phylogenetic distribution of VFs and association of phylogenetic group B2 and D with extraintestinal infections are well established and results from the current study can be linked to these previous reports, strengthening the virulence status of B2 and D phylogenetic groups in UTIs (Moreno *et al.* 2009; Ramos *et al.* 2010). Despite the significant association of some virulence factors with different phylogenetic groups in the current study, the relationships were in general reflecting the association of virulence factors with over-represented STs among phylogenetic groups in the current strain collection, rather than the phylogenetic group itself. For instance, the high level of Pap elements detected among isolates of phylogenetic group D are in fact caused by the strong association of different Pap elements with ST73 and ST69. Additionally, within clones that consisted of isolates of different phylogenetic groups, the phylogenetic group assignment of an isolate did not affect its virulence profile. For example, no significant difference was observed in harbouring different Pap elements among ST73 isolates that belonged to phylogenetic group B2 or D.

On the other hand, the significant differences in the prevalence of some VFs, such as *fyuA* and PAI, among phylogenetic groups B2 and D compared to groups A and B1, may provide evidence of the probable role of phylogenetic background in determining the virulence of a strain. Furthermore, Moreno *et al*. (2008), in a study of urovirulence characteristics of faecal *E. coli*, showed that both groups B2 and D tend to dominate faecal *E. coli* populations leading to low clonal diversity, suggesting the presence of shared specific bacterial factors that contribute to fitness within the intestine (Moreno *et al.* 2009).

Since the theory that resistance comes at a cost to fitness was first proposed, several studies examining the relationship of antibiotic resistance with urovirulence potential have been published. Although they clearly show a significantly reduced prevalence of urovirulence genes among isolates resistant to quinolones and fluoroquinolones, the relationship of other antibiotic families such as trimethoprim/sulfamethoxazole, Beta-lactams and extended spectrum cephalosporins with urovirulence traits is still under debate (Horcajada *et al.* 2005; Houdouin *et al.* 2006; Johnson *et al.* 2005b; Johnson *et al.* 2003b; Moreno *et al.* 2006).

However, antibiotic resistant strains in the current study exhibited generally low virulence capability compared to susceptible strains and, consistent with previous findings (Moreno *et al.* 2006; Rijavec *et al.* 2008), resistance to quinolones and fluoroquinolones was associated with reduced incidence of different virulence traits. Significant associations were observed between quinolone resistance and low incidence of *papAH*, *papG* allele III, *sfa/focDE* and *cnf1*, which agrees with previous studies reporting quinolone resistant UPEC isolates having a lower prevalence of *papGIII* and *hlyA* (Starcic Erjavec *et al.* 2007), *hylA* and *cnf1* (Horcajada *et al.* 2005; Moreno *et al.* 2005) and *papA*, *papGIII*, *sfa/focDE*, *hlyA* and *cnf1* (Moreno *et al.* 2006) than

susceptible isolates. Fluoroquinolone resistant isolates were significantly associated with a low incidence of *papA* and *sfa/focDE* in addition to *papC*, which strengthens previous findings (Horcajada *et al.* 2005; Johnson *et al.* 2005b; Moreno *et al.* 2006; Piatti *et al.* 2008).

Several mechanisms have been proposed to explain this phenomenon. One proposal is that exposure to antibiotics may induce the loss of VFs. Soto *et al.* (2006) reported that UPEC strains exposed to sub-inhibitory concentrations of quinolones showed partial or total loss of PAIs containing VF genes (Soto, Jimenez de Anta & Vila 2006). However, Johnson *et al.* (2005) disagreed with the previous proposal and provided an alternative explanation, that the low level carriage of virulence determinants in quinolone and fluoroquinolone resistant *E. coli* clinical isolates was more likely due to the importation of resistant strains from an as-yet-undefined reservoir (for example, food animals) of isolates with low-virulence that are subjected to a high degree of selective pressure for development of resistance (Johnson *et al.* 2005a).

The present study provides an additional dimension to the latter explanation; although fluoroquinolone resistance was significantly associated with ST131, isolates of this clone showed no significant difference in the prevalence of most virulence determinants between fluoroquinolone resistant and susceptible isolates. This may indicate that the low prevalence of virulence determinants in resistant strains is more related to the characteristics of the ST/lineage than the acquisition of fluoroquinolone resistance.

Resistance to trimethoprim was associated with a reduced incidence of different virulence factors, including *papG* allele *III*, *sfa/focDE*, *sfa/draBC*, *hlyA* and *cnf1*, in keeping with the report of Moreno *et al* (Moreno *et al.* 2005), but not with some other reports (Houdouin *et al.* 2006; Johnson *et al.* 2005b; Vila *et al.* 2002). The virulence

differences noted in the current study could be due to confounding effects of concurrent quinolone and fluoroquinolone resistance.

On the other hand, it is worth noting that the quinolone, fluoroquinolone and trimethoprim resistant strains showed a high incidence of aerobactin (*iutA*) and serum resistance (*traT*) encoding genes, compared to susceptible strains. Although prevalence of aerobactin encoding genes were previously reported to be a s high as 91% (Rijavec *et al.* 2008) and 78% (Moreno *et al.* 2005), no significant differences were reported between resistant and susceptible strains in previous isolate collections. As iron acquisition systems are important VFs for bacterial passage to, and survival in, the bloodstream (Johnson & Stell 2000), the high level of aerobactin observed among resistant strains in the current study could be due to the over-representation of the ST131 clone, which has frequently been associated with bacteraemia.

It was clear that the relationship between virulence properties and antibiotic resistance in *E. coli*, is a complex phenomenon that could be affected by the phylogenetic background and ST and reflecting their interplay.

4.4. Metabolic profiling

Metabolic reactions have been conventionally used for the classification of bacteria into families and species. Further subdivision of bacteria below species level into subgroups such as serotypes, phage types or biotypes is crucial for epidemiological investigations. However the usefulness of biochemical reactions as a typing system is greatly dependent on the variability of these markers within the species and the reproducibility and discriminative power of the biochemical tests used (Katouli, Kuhn & Mollby 1990). Kuhn (1985) developed a biochemical typing method in which the kinetics of several biochemical reactions creates a distinctive fingerprint for bacterial strains under

investigation (Kuhn 1985). Recent advances in biochemical testing involving automated systems have overcome the low reproducibility of conventional biochemical tests. Additionally, in epidemiological investigations, biotyping systems are often coupled with different typing methods to improve their discriminative power (Katouli *et al.* 1990). The VITEK 2.0 compact system was used in the current study, as previous evaluation studies proved that the VITEK 2.0 compact system, through its newly developed fluorescence based technology, generates reliable and reproducible results, in addition to being less time and labour consuming (Funke *et al.* 1998; Gavin *et al.* 2002).

Uropathogenic isolates show similar biochemical profiles to those of the general *E. coli* population with one exception; α-galactosidase was observed in 69% of the UPEC isolates compared to 99% of the general *E. coli* population, as reported by other groups (Ewing 1986; Farmer *et al.* 1985; Leclercq *et al.* 2001). The reason for the discrepancy between the relatively low incidence of α-galactosidase among UPEC reported in the current study compared to the almost 100% incidence among general *E. coli* populations is not clear. However, the environment from which the *E. coli* strains were collected could play a role in the metabolic activity expressed by *E. coli*. Urine with its low pH and low glucose concentration could present a different environmental selective pressure on genes involved in metabolic activity. In this case transcription of the *melA* gene encoding α-galactosidase production is controlled by the cAMP-CRP regulatory circuit, where the presence or absence of glucose determines the concentration of cAMP-CRP that in turn activates expression of the *melA* gene (Liljestrom & Liljestrom 1987).

Some inferences could also be drawn regarding the correlation between biochemical tests with STs. Most of the significant correlations were confined to three types of

metabolic reactions: peptidase (proline and tyrosine arylamidase); decarboxylase (ornithine and lysine decarboxylase); and alkalinisation (L-lactate and succinate). Proline and tyrosine arylamidase are specific peptidase enzymes that hydrolysis proline and tyrosine rich proteins (Fukasawa *et al.* 1982). Similarly ornithine and lysine decarboxylase are also involved in protein metabolism.

The other metabolic reactions such as alkalinisation of lactate and succinate observed with an elevated incidence could be related to bacterial efforts to relieve acid stress exerted by amino acid metabolism. In a study of fitness of *E. coli* during UTI, Alter *et al.* found, by examining expression of UPEC cytoplasmic protein during growth in human urine, that *E. coli* scavenges amino acids and peptides and that disruption of peptide import in UPEC significantly compromised fitness during infection. This suggested that short peptides taken up by UPEC are degraded to amino acids that are catabolised and used as intermediates for the TCA cycle and a substrate for the gluconeogenesis pathway (Alteri, Smith & Mobley 2009).

Although, little is known about the contribution of bacterial metabolic activity to pathogenesis, several reports have shown that certain metabolic enzymes may play a role to enhance virulence of several microorganisms (Pancholi & Chhatwal 2003). Recent studies showed that the ability of UPEC to catabolise the amino acid D-serine during UTI not only supports bacterial growth but also acts as a signalling mechanism to trigger virulence gene expression (Anfora *et al.* 2007; Roesch *et al.* 2003).

Interestingly, in the assessment of the metabolic activity of different STs, the ST131 clone members showed a high metabolic capacity compared to other STs, which may compensate for the low virulence capacity and explain the virulence reported for members of this ST. In contrast ST127, showed the lowest metabolic capacity but held the highest VF-score among the commonly detected STs.

Reconsidering the metabolic activity as an indicator of bacterial fitness and an important factor in the pathogenesis of UPEC led to exploration of its interplay with virulence score and antibiotic resistance score. Although no significant association was observed between the three characteristics in general UPEC, re-examining the correlation at ST level revealed an appreciable negative correlation between metabolic score and the antibiotic resistance score of members of the ST131 clone.

The remarkable association of ST131 with several biochemical tests observed in the current study revealed distinctive biochemical profiles. Moreover, in cluster analysis based on biochemical profiles over half of the ST131 isolates shared 70% similarity, forming a single large cluster that was characterised by multiple antibiotic resistance profiles, compared to the rest of the ST131 isolates. This suggests ongoing subclonal spread of ST131. The association of clonal groups to specific biochemical profiles was previously reported for the O157:H7 clone (Leclercq *et al.* 2001) and the O15:K52:H1 and O25:K4 ST131 clones (Cagnacci *et al.* 2008), using API 20E tests.

Although, sequence types such ST73 and ST69 shows only few significant associations with individual biochemical tests, each ST tended to form a few small clusters indicating association with a small number of distinctive biochemical profiles, with no significant association to antibiotic resistance.

4.5. Antibiogram profiling

Analysis of antimicrobial resistance patterns revealed high-level resistance to β-lactam antibiotics in 55% of the tested isolates and over 38% were resistant to trimethoprim, followed by the quinolones, which collectively accounted for 15% of the isolates. These results were expected as they are commonly used first line antibiotics for the

treatment of UTI and they agree with those reported following Health Protection Agency surveillance (Johnson 2005).

The observed low prevalence of antibiotic resistant isolates among the Preston collection compared isolates from Manchester was expected as the Preston collection largely consisted of isolates collected from patients with community acquired UTI, where antibiotic resistance is less common.

Evidence of an association between *E. coli* clones and certain antibiotic resistance patterns has been previously reported. In an intercontinental study Nicolas-Chanoine *et al* (2008) described broad dissemination of the CTX-M15 clonal group with MLST profile ST131 that exhibited multiple antibiotic resistances including to ciprofloxacin (95%) and co-trimoxazole (50%) (Nicolas-Chanoine *et al.* 2008). Similarly, after examining 43 CTX-M-15 producing *E. coli* from different countries, Coque and colleagues (2008) showed that clonal complexes ST131 and ST405 widely associated with the CTX-M-15 enzyme (Coque *et al.* 2008). Consistent with the strong correlation between ST131 and resistance to ciprofloxacin and trimethoprim found in the previous studies, the ST131 clone reported in the current study was characterised by multiple antibiotic resistance patterns and showed a significant association with all antibiotic families tested.

The significant association of ST131 with multiple resistance profiles has been previously reported (Coque *et al.* 2008; Lau *et al.* 2008a). In the latter of these studies isolates, were collected from the same area covered in the current study and findings suggests that it is a result of co-selection of resistance determinants as a result of pre-exposure to antibiotic treatment.

The worldwide CGA isolates characteristically exhibit multidrug resistance and are greatly associated with resistance to co-trimoxazole. In 2005, Tartof *et al*. significantly

associated CGA isolates with ST69, which could explain the high level of resistance to trimethoprim among ST69 isolates tested in the current study (Tartof *et al.* 2005).

In the present study, the ST clones were compared against patient demographic data, and the main correlation was seen for isolates of ST69 showed a significant association with recovery from females. Although it was first reported among women (Johnson *et al.* 2002), CGA association to certain gender groups is not well established (Johnson *et al.* 2005c).

Similarly ST131 was previously recovered from hospital patients as well as from the community (Peirano & Pitout 2010). The high prevalence of ST131 among community acquired infection group reported in the current study could be related to the previously reported dissemination of the ST131 community onset UPEC clone (Nicolas-Chanoine *et al.* 2008).

This correlation between phenotypic characteristics and STs gives rise to the question of how much horizontal transfer contributes to the evolution of these clusters. Generally, the extent and the significance of recombination in bacterial populations has always been a matter of debate (Feil *et al.* 2001). In some studies, the high levels of linkage disequilibrium between alleles and the existence of a clonal population structure, suggest that recombination rarely occurs in house keeping genes (Feil *et al.* 2001; Perez-Losada *et al.* 2006), while others propose a relative contribution of recombination to clonal diversification (Lacher *et al.* 2007). Results from the current study suggest an early role of recombination in the diversity of *E. coli* populations, whereas recurrent mutations are responsible for the diversity within clonal groups.

In contrast, the mosaic structure observed when phylogenetic groups were assigned to different isolates indicates a much more complex population, in congruence with

previous study (Wirth *et al.* 2006) that revealed the presence of hybrid groups within the ABD lineage classification system. Data recorded here shows that 10 STs accounting for 46% of the studied population, consisted of mixed phylogenetic groups, which could indicate recent genetic exchange of phylogenetic group markers between STs.

4.6. ST131 and CTX-M-15 producing O25b ST131 subclone

Compared to other successful STs, the ST131 clone always displayed distinctive phenotypic and genotypic characteristics that set it apart from other STs. ST131 accounts for 12% of the total UPEC, 50% of ESBL producing isolates and 79% of fluoroquinolone resistant isolates. Similar high prevalence values were previously reported (Johnson *et al.* 2010a) suggesting ongoing spread of ST131 as a major multidrug resistant extraintestinal pathogen.

However, recent dissemination of ST131 lineage isolates associated with CTX-M-15 and serogroup O25 has been widely reported (Coque *et al.* 2008; Nicolas-Chanoine *et al.* 2008; Yumuk *et al.* 2008). These isolates are usually fluoroquinolone resistant and unlike most historical antimicrobial resistant strains, this clone derives from the most virulent phylogenetic group B2 (Blanco *et al.* 2009). Analysis of ST131 isolates in the current collection revealed that CTX-M-15 producing O25b ST131 subclone accounted for 11 (30%) of the total ST131 isolates.

Several factors may play a role in the success of the ST131 clone; Peirano and Pitout (2010) suggest a combination of phylogenetic group B2 and certain virulence factors and fluoroquinolone resistance as important factors in the success of the ST131 clone in causing UTI, whereas the acquisition of CTX-M-15 carrying plasmids have facilitated the rapid global spread of the clone (Peirano & Pitout).

Although, ST131 isolates in the present study were significantly more likely to be from group B2 than non-ST131 isolates, they did not exhibit the classical virulence traits associated with phylogenetic group B2, such as P and S adhesins, Hly and CNF1 toxins and K1 invasive antigen, which is in agreement with similar findings (Blanco *et al.* 2009; Clermont *et al.* 2008).

In the current study, much of the virulence potential of ST131 isolates was due to a few virulence traits that may not explain their success as ExPEC. Johnson *et al.* (2010) suggest that undefined group B2 associated factors provide fitness advantages to ST131, independent of its virulence traits (Johnson *et al.* 2010a). However, there was no significant difference in phylogenetic background between the successful O25b-CTX-M-15 clone and the rest of the ST131 isolates.

On the other hand, apart from resistance to fluoroquinolones, most of the significant associations of ST131 with antibiotic resistance were largely caused by the O25b-CTX-M-15 subclone. CTX-M–encoding genes are usually carried by closely related IncFII plasmids embedded in class 1 integron containing additional resistance genes conferring resistance to β-lactams, fluoroquinolones, aminoglycosides, and trimethoprim-sulfamethoxazole (Peirano & Pitout 2010).

The identification of IncF plasmids carrying the FII and FIA replicons and association with the blaCTX-M-15 gene was demonstrated in O25-ST131 in the UK and other countries (Coque *et al.* 2008; Karisik *et al.* 2006) and Woodford *et al.* (2009) sequenced IncFII plasmids carried by an isolate of the CTX-M ESBL producing ST131 clone demonstrating that the IncFII plasmids harbouring $bla_{CTX\text{-}M\text{-}15}$, $bla_{OXA\text{-}1}$, *tetA*, *aac*(6')-*lb-cr* and *aac*(3)-*II* have played a key role in the rapid spread of CTX-M-15 β-lactamases in *E. coli* (Woodford *et al.* 2009). These plasmids might carry important,

but still unknown, VFs that have supported the spread of the epidemic clone O25b:H4-ST131 (Blanco *et al.* 2009).

Moreover, the exclusive association of *aac(6')-lb-cr* with ESBL producing ciprofloxacin resistant *E. coli* observed in the current study, strongly supports the role of plasmids in the resistance capacities of ST131 isolates. However, the high level of fluoroquinolone resistance among non-O25b-CTX-M-15-ST131 isolates suggests the involvement of a different mechanism. Interestingly, all fluoroquinolone resistant ST131 isolates showed multiple mutations in both *gyrA* and *parC* genes, a pattern of mutation frequently reported (Drlica & Zhao 1997; Jacoby 2005) and associated with high-level fluoroquinolone resistance. Thus, it appears that resistance to fluoroquinolones in ST131 is due to mutations of the chromosomal targets of these drugs, as opposed to acquisition of plasmid mediated resistance determinants. As recently demonstrated by (Cerquetti *et al.* 2010).

The high prevalence of isolates of the ST131 clone among fluoroquinolone resistant *E. coli* (Cerquetti *et al.* 2010; Uchida *et al.* 2010) and the acquisition of β-lactamase enzymes other than CTX-M-15 by ST131 in different geographical regions, suggested that spread of ST131 *E. coli* as a clone precedes the acquisition of antibiotic resistance (Cerquetti *et al.* 2010), which was supported by the prevalence of antibiotic susceptible ST131 reported in the current study. Although previous reports suggested that the strong linkage between this clone and quinolone resistance could be promoted by the high prevalence of *aac(6)-lb-cr* among isolates of this clone (Jones *et al.* 2008), the fluoroquinolone resistance reported in the current study occurs in ST131 UPEC as a result of accumulated point mutations in both *gyrA* and *parC* genes despite their acquisition of any plasmid mediated resistance.

The multiple antibiotic resistance that characterised ST131 can not completely explain the success of the clone as an ExPEC, especially with its lack of ExPEC key virulence capabilities. The introduction of bacterial metabolism as an important factor in the pathogenesis process provides new perspectives in understanding the basis of bacterial pathogenesis (Le Bouguenec & Schouler 2010).

The high aggregative bio-scores among ST131 isolates compared to other commonly detected clonal groups and the association of the O25b-CTX-M-15-ST131 clone with distinctive biochemical profiles that share 88% similarity, suggested that biochemical capabilities could play an important role in the success of ST131 as an ExPEC clone by facilitating the adaptation of strains to their host environment by modifying catabolic pathways, and taking advantage of available nutrients present in the urine such as amino acids and small peptides. This may subsequently promote colonisation of the host and increase the pathogenic potential of the bacteria(Le Bouguenec & Schouler 2010).

The main conclusion of these previous investigations is that the ST131 clonal group in general, and O25b-ST131 CTX-M-15 producing strains in particular, exhibit unique characteristics and that no single factor is responsible for their success as ExPEC. Multivariate logistic regression analysis was used to determine which factors collectively most likely describe the ST131 clone and the O25b-CTX-M-15 ST131 sub-clone. In the first model, ST131 is best described by its fluoroquinolone resistance and possession of PAI, *ibeA*, and Dr antigen-specific adhesins, whereas in the second model, the O25b-CTX-M-15 ST131 sub-clone was only differentiated from the rest of ST131 clone members by the production of an ESBL enzyme.

The report by Diard *et al.* (2010) that ExPEC PAIs play an important role in intestinal colonisation (Diard *et al.* 2010) and demonstration by Moreno *et al.* (2009) that both

groups B2 and D tend to dominate faecal *E. coli* populations prior to extraintestinal infections, suggest the presence of shared specific bacterial factors that contribute to fitness within the intestine (Moreno *et al.* 2009). All these observations in addition to the detection of *E. coli* belong to the ST131 clone in the faeces of healthy individuals (Leflon-Guibout *et al.* 2008) indicate that ST131 isolates colonise the intestine were they compete with and dominate the faecal *E. coli* population increasing the chance of transmission to the urinary tract. On the other hand, Diard *et al*. (2010) showed that PAIs diminish fitness of their carrier during growth in urine (Diard *et al.* 2010).

However, maintenance of fitness requires many properties, among which metabolic competence is of the utmost importance as is resistance to different stresses or urine flow, such as through biofilm formation. Although biofilm formation was not examined in the current study, some observation, such as the association of ST131 with the *ibeA* gene and the high incidence among catheterised patients suggest that biofilm could facilitate ST131 colonisation. The fact that the O25b-CTX-M-15 ST131 sub-clone was only differentiate from the rest of ST131 clone by its acquisition of CTX-M-15 ESBL enzyme indicates it's recent diversion from the main clone.

4.7. Summary

In conclusion, the use of MLST to characterise the UPEC population circulating in the North West region of England between June 2007 and June 2009, has proven to be a powerful epidemiological tool that has significantly contributed to our understanding of the various phylogenetic and phenotypic characteristics, previously associated with UPEC strains.

The first part of the current study described the population structure of a defined UPEC population, based on MLST and phylogenetic analysis and the facilitated their distribution into phylogenetic groups and clonal groups in an unbiased way. The findings of this part of the current study can be summarised as follows:

- The findings contradict the results of studies using other, lower resolution, methods that tend to mask the real correlations between strains in the population by over representation of certain clones in a group.
- A consistent profile of STs repeatedly appeared in the population over an extended time period.
- ST73 was the most commonly detected lineage, followed by ST131 and eBURST analysis of allelic profiles reinforced the status of ST73 in the population and identified it as the founder of the largest CC.
- Phylogenetic analysis in current study provided evidence that ST131 and ST127 have emerged as newly evolved STs.
- PFGE profiles observed among isolates of ST131 included in the current study suggest that sub-clonal evolution has already taken place, as demonstrated by a sub-cluster of ST131 isolates collected in Preston.

- Although many of the detected STs were previously been reported, other successful STs such ST127 and ST80 presented in this study have never been reported as remarkable pathogens.

The second part of the study determined the metabolic and the virulence capacity of the UPEC population, in addition to their antibiotic resistance profiles, using different phenotypic and genotypic methods and analysed the prevalence of these traits in relation to the defined ST clones. Several correlations were identified between the successful STs and key traits and these patterns could help to explain why isolates from certain STs are successful ExPEC clones.

- Despite a low metabolic capacity, ST127, which was suggested by phylogenetic analysis to be a newly evolved clone, holds the highest virulence score of all lineages in the current collection. This may permit survival of ST127 isolates in the population long enough for them to gain antibiotic resistance markers.
- As resistance comes at a cost to fitness, it is not surprising that the multidrug resistant ST131 clone was defined by a low virulence score and distinctive VF profiles lacking fimbrial adhesins. This, however, was balanced by significant concentration of non-fimbrial adhesins of the Dr family and the gene encoding the invasion of brain endothelium phenotype (*ibeA*) that is responsible for neonatal meningitis in humans.
- ST131 revealed distinctive biochemical profiles and was characterised by a high metabolic capacity compared to other STs, which may compensate for the low virulence capacity and explain the fitness for causing infection reported for members of this ST.

The third part of the study concentrated on analysing characteristics of members of the ST131 clone and assessment of the importance of all of the phenotypic and genotypic characteristics that showed a significant association with ST131 isolates. This was carried out to determine the possible role in the success of the isolates from the ST131 clone and the O25b-CTX-M-15 ST131 sub-clone as pathogenic ExPEC strains.

- The introduction of bacterial metabolism as an important factor in the pathogenesis process provided a new perspective on understanding the success of ST131 by facilitating the adaptation of strains to their host environment using modified catabolic pathways, allowing isolates to take advantage of available nutrients present in the urine, such as amino acids and small peptides.
- ST131 is best described by its fluoroquinolone resistance and possession of PAI, *ibeA*, and Dr antigen-specific adhesins, whereas the O25b-CTX-M-15 ST131 sub-clone was only differentiated from the rest of the ST131 clone members by the production of an ESBL enzyme.

4.8. Additional further work

In addition to the points mentioned above, further analyses may increase the impact of the work described in this thesis. It would be useful to study the association of different STs to typing using other phenotypic techniques such as serotyping, as this may lead to a better understanding of the phylogeny of UPEC.

The high metabolic capacity associated with ST131 in the absence of any clear virulence advantage over other STs gives rise to the question of the contribution of bacterial metabolic activity to ST131 pathogenesis. Further studies are warranted involving screening for expression of genes involved in metabolic activity associated with increased virulence of UPEC isolates. This could be carried out using techniques such as microarray or 'RNAseq'. During such experiments, tested isolates should be grown in conditions similar to that experienced by UPEC during urinary tract infection so RNA would be extracted from isolates used in animal models of infection and following growth in urine.

It would be useful to further study some of the other STs, such as ST73, to fully understand its virulence potential and factors associated with its domination of the UPEC population. This may contribute to a better understanding of the potential for emergence of other STs such as may be the case ST127, which could be a significant future UPEC lineage.

Inclusion of more clinical data associated with isolates presented in this study could have provided a deeper insight in to the pathogenesis of each ST.

References

5. References

Achtman M., Mercer A., Kusecek B., Pohl A., Heuzenroeder M., Aaronson W., Sutton A. & Silver R.P. (1983) Six widespread bacterial clones among Escherichia coli K1 isolates. *Infect Immun* **39**, 315-335.

Achtman M., Morelli G., Zhu P., Wirth T., Diehl I., Kusecek B., Vogler A.J., Wagner D.M., Allender C.J., Easterday W.R., Chenal-Francisque V., Worsham P., Thomson N.R., Parkhill J., Lindler L.E., Carniel E. & Keim P. (2004) Microevolution and history of the plague bacillus, Yersinia pestis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17837-17842.

Achtman M. & Pluschke G. (1986) Clonal analysis of descent and virulence among selected Escherichia coli. *Annu Rev Microbiol* **40**, 185-210.

Adiri R.S., Gophna U. & Ron E.Z. (2003) Multilocus sequence typing (MLST) of Escherichia coli O78 strains. *FEMS Microbiol Lett* **222**, 199-203.

Aguilera A., Rothstein R., Michel B., Baharoglu Z. & Lestini R. (2007) Genetics of Recombination in the Model Bacterium Escherichia Coli. In: *Molecular Genetics of Recombination*, Vol. 17, pp. 1-26. Springer Berlin / Heidelberg.

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 Pathog* **5**, e1000448.

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. *Infect Immun* **75**, 5298-5304.

Anonymous (2004) *Investigation of urine* Health Protection Agency, UK.

Anonymous (2006) Management of suspected bacterial urinary tract infection in adults: A national clinical guideline. In: *Scottish Intercollegiate Guidelines Network*. Scottish Intercollegiate Guidelines Network.

Anonymous (2007) *Antimicrobial resistance surveillance in Europe 2007*. European Centre for Disease Prevention and Control.

Anonymous (2009) *Antimicrobial resistance surveillance in Europe 2009*. European Centre for Disease Prevention and Control.

Antao E.M., Wieler L.H. & Ewers C. (2009) Adhesive threads of extraintestinal pathogenic Escherichia coli. *Gut Pathog* **1**, 22.

Arthur M., Arbeit R.D., Kim C., Beltran P., Crowe H., Steinbach S., Campanelli C., Wilson R.A., Selander R.K. & Goldstein R. (1990) Restriction fragment length

polymorphisms among uropathogenic Escherichia coli isolates: pap-related sequences compared with rrn operons. *Infect Immun* **58**, 471-479.

Avery S.M., Liebana E., Reid C.A., Woodward M.J. & Buncic S. (2002) Combined use of two genetic fingerprinting methods, pulsed-field gel electrophoresis and ribotyping, for characterization of Escherichia coli O157 isolates from food animals, retail meats, and cases of human disease. *J Clin Microbiol* **40**, 2806-2812.

Ayling R.D., Woodward M.J., Evans S. & Newell D.G. (1996) Restriction fragment length polymorphism of polymerase chain reaction products applied to the differentiation of poultry campylobacters for epidemiological investigations. *Research in Veterinary Science* **60**, 168-172.

Ball P. (2000) Quinolone generations: natural history or natural selection? *J Antimicrob Chemother* **46 Suppl T1**, 17-24.

Baquero M.R., Galan J.C., del Carmen Turrientes M., Canton R., Coque T.M., Martinez J.L. & Baquero F. (2005) Increased mutation frequencies in Escherichia coli isolates harboring extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* **49**, 4754-4756.

Bebora L.C., Yamamoto H., Sang W.K., Oundo J.O., Khamala J., Mwangudza A. & Mukundi P.W. (1994) Plasmid profiles of E. coli, Salmonella and shigella organisms isolated in Kenya. *East African Medical Journal* **71**, 619-623.

Ben-Darif E., Jury F., De Pinna E., Threlfall E.J., Bolton F.J., Fox A.J. & Upton M. (2010) Development of a multiplex primer extension assay for rapid detection of salmonella isolates of diverse serotypes. *Journal of Clinical Microbiology* **48**, 1055-1060.

Bender J.B., Hedberg C.W., Besser J.M., Boxrud D.J., MacDonald K.L. & Osterholm M.T. (1997) Surveillance by molecular subtype for Escherichia coli O157:H7 infections in Minnesota by molecular subtyping. *N Engl J Med* **337**, 388-394.

Bert F., Johnson J.R., Ouattara B., Leflon-Guibout V., Johnston B., Marcon E., Valla D., Moreau R. & Nicolas-Chanoine M.H. (2010) Genetic diversity and virulence profiles of Escherichia coli isolates causing spontaneous bacterial peritonitis and bacteremia in patients with cirrhosis. *J Clin Microbiol* **48**, 2709-2714.

Best E.L., Fox A.J., Frost J.A. & Bolton F.J. (2004) Identification of Campylobacter jejuni multilocus sequence type ST-21 clonal complex by single-nucleotide polymorphism analysis. *Journal of Clinical Microbiology* **42**, 2836-2839.

Beutin L., Kaulfuss S., Herold S., Oswald E. & Schmidt H. (2005) Genetic analysis of enteropathogenic and enterohemorrhagic Escherichia coli serogroup O103 strains by molecular typing of virulence and housekeeping genes and pulsed-field gel electrophoresis. *J Clin Microbiol* **43**, 1552-1563.

Bingen E., Picard B., Brahimi N., Mathy S., Desjardins P., Elion J. & Denamur E. (1998) Phylogenetic analysis of Escherichia coli strains causing neonatal meningitis

suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J Infect Dis* **177**, 642-650.

Birkett C.I., Ludlam H.A., Woodford N., Brown D.F., Brown N.M., Roberts M.T., Milner N. & Curran M.D. (2007) Real-time TaqMan PCR for rapid detection and typing of genes encoding CTX-M extended-spectrum beta-lactamases. *J Med Microbiol* **56**, 52-55.

Birnboim H.C. & Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-1523.

Blackwood R.A., Rode C.K., Pierson C.L. & Bloch C.A. (1997) Pulsed-field gel electrophoresis genomic fingerprinting of hospital Escherichia coli bacteraemia isolates. *J Med Microbiol* **46**, 506-510.

Blahna M.T., Zalewski C.A., Reuer J., Kahlmeter G., Foxman B. & Marrs C.F. (2006) The role of horizontal gene transfer in the spread of trimethoprim-sulfamethoxazole resistance among uropathogenic Escherichia coli in Europe and Canada. *J Antimicrob Chemother* **57**, 666-672.

Blanco M., Alonso M.P., Nicolas-Chanoine M.H., Dahbi G., Mora A., Blanco J.E., Lopez C., Cortes P., Llagostera M., Leflon-Guibout V., Puentes B., Mamani R., Herrera A., Coira M.A., Garcia-Garrote F., Pita J.M. & Blanco J. (2009) Molecular epidemiology of Escherichia coli producing extended-spectrum {beta}-lactamases in Lugo (Spain): dissemination of clone O25b:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* **63**, 1135-1141.

Blattner F.R., Plunkett G., 3rd, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B. & Shao Y. (1997) The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1474.

Bouchet V., Huot H. & Goldstein R. (2008) Molecular genetic basis of ribotyping. *Clinical Microbiology Reviews* **21**, 262-273.

Brauner A., Boeufgras J.M., Jacobson S.H., Kaijser B., Kallenius G., Svenson S.B. & Wretlind B. (1987) The use of biochemical markers, serotype and fimbriation in the detection of Escherichia coli clones. *Journal of General Microbiology* **133**, 2825-2834.

Bryant D. & Moulton V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255-265.

Bush K. & Macielag M.J. (2010) New beta-lactam antibiotics and beta-lactamase inhibitors. *Expert Opin Ther Pat* **20**, 1277-1293.

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. *J Clin Microbiol* **46**, 2605-2612.

Canton R. & Coque T.M. (2006) The CTX-M beta-lactamase pandemic. *Curr Opin Microbiol* **9**, 466-475.

Cattoir V., Poirel L., Rotimi V., Soussy C.J. & Nordmann P. (2007) Multiplex PCR for detection of plasmid-mediated quinolone resistance qnr genes in ESBL-producing enterobacterial isolates. *J Antimicrob Chemother* **60**, 394-397.

Cerquetti M., Giufre M., Garcia-Fernandez A., Accogli M., Fortini D., Luzzi I. & Carattoli A. (2010) Ciprofloxacin-resistant, CTX-M-15-producing Escherichia coli ST131 clone in extraintestinal infections in Italy. *Clin Microbiol Infect.*

Chaslus-Dancla E., Pohl P., Meurisse M., Marin M. & Lafont J.P. (1991) High genetic homology between plasmids of human and animal origins conferring resistance to the aminoglycosides gentamicin and apramycin. *Antimicrob Agents Chemother* **35**, 590-593.

Chulain M.N., Morris D. & Cormican M. (2006) Enterobacterial repetitive intergenic consensus-polymerase chain reaction for typing of uropathogenic Escherichia coli is not what it seems. *Clin Infect Dis* **42**, 1805-1806.

Clermont O., Bonacorsi S. & Bingen E. (2000) Rapid and simple determination of the Escherichia coli phylogenetic group. *Appl Environ Microbiol* **66**, 4555-4558.

Clermont O., Cordevant C., Bonacorsi S., Marecat A., Lange M. & Bingen E. (2001) Automated ribotyping provides rapid phylogenetic subgroup affiliation of clinical extraintestinal pathogenic Escherichia coli strains. *Journal of Clinical Microbiology* **39**, 4549-4553.

Clermont O., Lavollay M., Vimont S., Deschamps C., Forestier C., Branger C., Denamur E. & Arlet G. (2008) The CTX-M-15-producing Escherichia coli diffusing clone belongs to a highly virulent B2 phylogenetic subgroup. *J Antimicrob Chemother* **61**, 1024-1028.

Cloninger C., Felton M., Paul B., Hirakawa Y. & Metzenberg S. (2008) Control of pH during plasmid preparation by alkaline lysis of Escherichia coli. *Anal Biochem* **378**, 224-225.

Coque T.M., Novais A., Carattoli A., Poirel L., Pitout J., Peixe L., Baquero F., Canton R. & Nordmann P. (2008) Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis* **14**, 195-200.

Denamur E., Bonacorsi S., Giraud A., Duriez P., Hilali F., Amorin C., Bingen E., Andremont A., Picard B., Taddei F. & Matic I. (2002) High frequency of mutator strains among human uropathogenic Escherichia coli isolates. *J Bacteriol* **184**, 605-609.

Diard M., Garry L., Selva M., Mosser T., Denamur E. & Matic I. (2010) Pathogenicity-associated islands in extraintestinal pathogenic Escherichia coli are fitness elements involved in intestinal colonization. *J Bacteriol* **192**, 4885-4893.

Didelot X. & Falush D. (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251-1266.

Diggle M.A., Bell C.M. & Clarke S.C. (2003) Nucleotide sequence-based typing of meningococci directly from clinical samples. *J Med Microbiol* **52**, 505-508.

Domingue G., Willshaw G.A., Smith H.R., Perry N., Radford D. & Cheasty T. (2003) DNA-based subtyping of verocytotoxin-producing Escherichia coli (VTEC) O128ab:H2 strains from human and raw meat sources. *Letters in Applied Microbiology* **37**, 433-437.

Doolittle W.F. (1999) Lateral genomics. *Trends Cell Biol* **9**, M5-8.

Drlica K. & Zhao X. (1997) DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev* **61**, 377-392.

Duffell E., Espie E., Nichols T., Adak G.K., De Valk H., Anderson K. & Stuart J.M. (2003) Investigation of an outbreak of E. coli O157 infections associated with a trip to France of schoolchildren from Somerset, England. *Euro Surveill* **8**, 81-86.

Dutta S., Rajendran K., Roy S., Chatterjee A., Dutta P., Nair G.B., Bhattacharya S.K. & Yoshida S.I. (2002) Shifting serotypes, plasmid profile analysis and antimicrobial resistance pattern of shigellae strains isolated from Kolkata, India during 1995-2000. *Epidemiol Infect* **129**, 235-243.

Ellington M.J., Livermore D.M., Pitt T.L., Hall L.M. & Woodford N. (2006) Mutators among CTX-M beta-lactamase-producing Escherichia coli and risk for the emergence of fosfomycin resistance. *J Antimicrob Chemother* **58**, 848-852.

Enright M.C. & Spratt B.G. (1999) Multilocus sequence typing. *Trends Microbiol* **7**, 482-487.

Ensor V.M., Shahid M., Evans J.T. & Hawkey P.M. (2006) Occurrence, prevalence and genetic environment of CTX-M beta-lactamases in Enterobacteriaceae from Indian hospitals. *J Antimicrob Chemother* **58**, 1260-1263.

Erb A., Sturmer T., Marre R. & Brenner H. (2007) Prevalence of antibiotic resistance in Escherichia coli: overview of geographical, temporal, and methodological variations. *Eur J Clin Microbiol Infect Dis* **26**, 83-90.

Ewing W.H.E., P. R. (1986) *(The genus Escherichia) Identification of Enterobacteriacae*. 4th edn. Minneapolis, Minn;Burgess.

Farmer J.J., 3rd, Davis B.R., Hickman-Brenner F.W., McWhorter A., Huntley-Carter G.P., Asbury M.A., Riddle C., Wathen-Grady H.G., Elias C., Fanning G.R. & et al. (1985) Biochemical identification of new species and biogroups of Enterobacteriaceae isolated from clinical specimens. *J Clin Microbiol* **21**, 46-76.

Feil E.J., Holmes E.C., Bessen D.E., Chan M.S., Day N.P., Enright M.C., Goldstein R., Hood D.W., Kalia A., Moore C.E., Zhou J. & Spratt B.G. (2001) Recombination within

natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A* **98**, 182-187.

Feil E.J., Li B.C., Aanensen D.M., Hanage W.P. & Spratt B.G. (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**, 1518-1530.

Felsenstein J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-376.

Filliol I., Motiwala A.S., Cavatore M., Qi W., HazbÃ³n M.H., Del Valle M.B., Fyfe J., GarcÃ-a-GarcÃ-a L., Rastogi N., Sola C., Zozio T., Guerrero M.I., LeÃ³n C.I., Crabtree J., Angiuoli S., Eisenach K.D., Durmaz R., Joloba M.L., RendÃ³n A., Sifuentes-Osornio J., Ponce De Leon A., Cave M.D., Fleischmann R., Whittam T.S. & Alland D. (2006) Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: Insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *Journal of Bacteriology* **188**, 759-772.

Fluit A.C., Jones M.E., Schmitz F.J., Acar J., Gupta R. & Verhoef J. (2000) Antimicrobial resistance among urinary tract infection (UTI) isolates in Europe: results from the SENTRY Antimicrobial Surveillance Program 1997. *Antonie Van Leeuwenhoek* **77**, 147-152.

Foley S.L., Lynne A.M. & Nayak R. (2009) Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. *Infect Genet Evol* **9**, 430-440.

Foxman B. (2007) Contributions of molecular epidemiology to the understanding of infectious disease transmission, pathogenesis, and evolution. *Ann Epidemiol* **17**, 148-156.

Foxman B., Manning S.D., Tallman P., Bauer R., Zhang L., Koopman J.S., Gillespie B., Sobel J.D. & Marrs C.F. (2002) Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. *Am J Epidemiol* **156**, 1133-1140.

Fukasawa K., Hiraoka B.Y., Fukasawa K.M. & Harada M. (1982) Arylamidase activities specific for proline, tyrosine, and basic amino acid residues in some oral bacteria. *J Dent Res* **61**, 818-820.

Fukigai S., Alba J., Kimura S., Iida T., Nishikura N., Ishii Y. & Yamaguchi K. (2007) Nosocomial outbreak of genetically related IMP-1 beta-lactamase-producing Klebsiella pneumoniae in a general hospital in Japan. *Int J Antimicrob Agents* **29**, 306-310.

Funke G., Monnet D., deBernardis C., von Graevenitz A. & Freney J. (1998) Evaluation of the VITEK 2 system for rapid identification of medically relevant gram-negative rods. *J Clin Microbiol* **36**, 1948-1952.

Garcia-Martinez J., Martinez-Murcia A.J., Rodriguez-Valera F. & Zorraquino A. (1996) Molecular evidence supporting the existence of two major groups in uropathogenic Escherichia coli. *FEMS Immunol Med Microbiol* **14**, 231-244.

Garrity G.M. (2005) *Bergeys Manual of Systematic Bacteriology*. Springer Science and Business Media Inc., New York.

Gavin P.J., Warren J.R., Obias A.A., Collins S.M. & Peterson L.R. (2002) Evaluation of the Vitek 2 system for rapid identification of clinical isolates of gram-negative bacilli and members of the family Streptococcaceae. *Eur J Clin Microbiol Infect Dis* **21**, 869-874.

Germon P., Chen Y.H., He L., Blanco J.E., Bree A., Schouler C., Huang S.H. & Moulin-Schouleur M. (2005) ibeA, a virulence factor of avian pathogenic Escherichia coli. *Microbiology* **151**, 1179-1186.

Godbout-DeLasalle F. & Higgins R. (1986) Biotyping of clinical isolates of Escherichia coli of animal origin, using the Analytab API 20E system. *Can J Vet Res* **50**, 418-421.

Goldberg T.L., Gillespie T.R. & Singer R.S. (2006) Optimization of analytical parameters for inferring relationships among Escherichia coli isolates from repetitive-element PCR by maximizing correspondence with multilocus sequence typing data. *Appl Environ Microbiol* **72**, 6049-6052.

Gordon K.A. & Jones R.N. (2003) Susceptibility patterns of orally administered antimicrobials among urinary tract infection pathogens from hospitalized patients in North America: comparison report to Europe and Latin America. Results from the SENTRY Antimicrobial Surveillance Program (2000). *Diagn Microbiol Infect Dis* **45**, 295-301.

Guttman D.S. (1997) Recombination and clonality in natural populations of Escherichia coil. *Trends in Ecology and Evolution* **12**, 16-21.

Guttman D.S. & Dykhuizen D.E. (1994) Clonal divergence in Escherichia coli as a result of recombination, not mutation. *Science* **266**, 1380-1383.

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 Catheter-associated bacteriuria and from fecal samples. *Infect Immun* **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. *Infect Immun* **70**, 4539-4546.

Hacker J. & Kaper J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641-679.

Hahm B.K., Maldonado Y., Schreiber E., Bhunia A.K. & Nakatsu C.H. (2003) Subtyping of foodborne and environmental isolates of Escherichia coli by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. *Journal of Microbiological Methods* **53**, 387-399.

Hall L.M. & Henderson-Begg S.K. (2006) Hypermutable bacteria isolated from humans--a critical analysis. *Microbiology* **152**, 2505-2514.

Hall R.M. & Collis C.M. (1998) Antibiotic resistance in gram-negative bacteria: the role of gene cassettes and integrons. *Drug Resist Updat* **1**, 109-119.

Haubold B. & Hudson R.R. (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. *Bioinformatics* **16**, 847-848.

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. *Infect Immun* **65**, 531-536.

Ho P.L., Wong R.C., Lo S.W., Chow K.H., Wong S.S. & Que T.L. (2010) Genetic identity of aminoglycoside-resistance genes in Escherichia coli isolates from human and animal sources. *J Med Microbiol* **59**, 702-707.

Hommais F., Pereira S., Acquaviva C., Escobar-PÃ¡ramo P. & Denamur E. (2005) Single-nucleotide polymorphism phylotyping of Escherichia coli. *Applied and Environmental Microbiology* **71**, 4784-4792.

Hooper D.C. (2001) Emerging mechanisms of fluoroquinolone resistance. *Emerg Infect Dis* **7**, 337-341.

Hopkins K.L. & Hilton A.C. (2000) Methods available for the sub-typing of Escherichia coli O157. *World Journal of Microbiology and Biotechnology* **16**, 741-748.
Horcajada J.P., Soto S., Gajewski A., Smithson A., Jimenez de Anta M.T., Mensa J., Vila J. & Johnson J.R. (2005) Quinolone-resistant uropathogenic Escherichia coli strains from phylogenetic group B2 have fewer virulence factors than their susceptible counterparts. *J Clin Microbiol* **43**, 2962-2964.

Houdouin V., Bonacorsi S., Bidet P., Bingen-Bidois M., Barraud D. & Bingen E. (2006) Phylogenetic background and carriage of pathogenicity island-like domains in relation to antibiotic resistance profiles among Escherichia coli urosepsis isolates. *J Antimicrob Chemother* **58**, 748-751.

Hulton C.S., Higgins C.F. & Sharp P.M. (1991) ERIC sequences: a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria. *Mol Microbiol* **5**, 825-834.

Huovinen P., Pulkkinen L., Helin H.L., Makila M. & Toivanen P. (1986) Emergence of trimethoprim resistance in relation to drug consumption in a Finnish hospital from 1971 through 1984. *Antimicrob Agents Chemother* **29**, 73-76.

Huson D.H. & Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.

Jacoby G.A. (2005) Mechanisms of resistance to quinolones. *Clin Infect Dis* **41 Suppl 2**, S120-126.

Jan N., Meshram S.U. & Kulkarni A. (2009) Plasmid profile analysis of multidrug resistant E. coli isolated from UTI patients of Nagpur City, India. *Romanian Biotechnological Letters* **14**, 4635-4640.

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

Johnson A. (2005) *Trends in Antimicrobial resistance in England and Wales*. Health Protection Agency, London.

Johnson A.P., Burns L., Woodford N., Threlfall E.J., Naidoo J., Cooke E.M. & George R.C. (1994) Gentamicin resistance in clinical isolates of Escherichia coli encoded by genes of veterinary origin. *J Med Microbiol* **40**, 221-226.

Johnson A.P., Malde M., Woodford N., Cunney R.J. & Smyth E.G. (1995) Urinary isolates of apramycin-resistant Escherichia coli and Klebsiella pneumoniae from Dublin. *Epidemiol Infect* **114**, 105-112.

Johnson J.R. (1991) Virulence factors in Escherichia coli urinary tract infection. *Clinical microbiology reviews* **4**, 80-128.

Johnson J.R., Gajewski A., Lesse A.J. & Russo T.A. (2003a) Extraintestinal pathogenic Escherichia coli as a cause of invasive nonurinary infections. *J Clin Microbiol* **41**, 5798-5802.

Johnson J.R., Johnston B., Clabots C., Kuskowski M.A. & Castanheira M. (2010a) Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. *Clin Infect Dis* **51**, 286-294.

Johnson J.R., Johnston B., Clabots C., Kuskowski M.A., Pendyala S., Debroy C., Nowicki B. & Rice J. (2010b) Escherichia coli sequence type ST131 as an emerging fluoroquinolone-resistant uropathogen among renal transplant recipients. *Antimicrob Agents Chemother* **54**, 546-550.

Johnson J.R., Johnston B., Clabots C.R., Kuskowski M.A., Roberts E. & DebRoy C. (2008) Virulence genotypes and phylogenetic background of Escherichia coli serogroup O6 isolates from humans, dogs, and cats. *J Clin Microbiol* **46**, 417-422.

Johnson J.R., Johnston B., Kuskowski M.A., Colodner R. & Raz R. (2005a) Spontaneous conversion to quinolone and fluoroquinolone resistance among wild-type Escherichia coli isolates in relation to phylogenetic background and virulence genotype. *Antimicrob Agents Chemother* **49**, 4739-4744.

Johnson J.R., Kuskowski M.A., O'Bryan T T., Colodner R. & Raz R. (2005b) Virulence genotype and phylogenetic origin in relation to antibiotic resistance profile among Escherichia coli urine sample isolates from Israeli women with acute uncomplicated cystitis. *Antimicrob Agents Chemother* **49**, 26-31.

Johnson J.R., Kuskowski M.A., Owens K., Gajewski A. & Winokur P.L. (2003b) Phylogenetic origin and virulence genotype in relation to resistance to fluoroquinolones and/or extended-spectrum cephalosporins and cephamycins among Escherichia coli isolates from animals and humans. *J Infect Dis* **188**, 759-768.

Johnson J.R., Manges A.R., O'Bryan T.T. & Riley L.W. (2002) A disseminated multidrug-resistant clonal group of uropathogenic Escherichia coli in pyelonephritis. *Lancet* **359**, 2249-2251.

Johnson J.R., Menard M., Johnston B., Kuskowski M.A., Nichol K. & Zhanel G.G. (2009) Epidemic clonal groups of Escherichia coli as a cause of antimicrobial-resistant urinary tract infections in Canada, 2002 to 2004. *Antimicrob Agents Chemother* **53**, 2733-2739.

Johnson J.R., Murray A.C., Kuskowski M.A., Schubert S., Prere M.F., Picard B., Colodner R. & Raz R. (2005c) Distribution and characteristics of Escherichia coli clonal group A. *Emerg Infect Dis* **11**, 141-145.

Johnson J.R., O'Bryan T.T., Kuskowski M. & Maslow J.N. (2001) Ongoing horizontal and vertical transmission of virulence genes and papA alleles among Escherichia coli blood isolates from patients with diverse-source bacteremia. *Infect Immun* **69**, 5363-5374.

Johnson J.R., Owens K., Gajewski A. & Kuskowski M.A. (2005d) Bacterial characteristics in relation to clinical source of Escherichia coli isolates from women with acute cystitis or pyelonephritis and uninfected women. *J Clin Microbiol* **43**, 6064-6072.

Johnson J.R. & Russo T.A. (2002) Uropathogenic Escherichia coli as agents of diverse non-urinary tract extraintestinal infections. *J Infect Dis* **186**, 859-864.

Johnson J.R. & Russo T.A. (2005) Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. *Int J Med Microbiol* **295**, 383-404.

Johnson J.R., Stapleton A.E., Russo T.A., Scheutz F., Brown J.J. & Maslow J.N. (1997) Characteristics and prevalence within serogroup O4 of a J96-like clonal group of uropathogenic Escherichia coli O4:H5 containing the class I and class III alleles of papG. *Infect Immun* **65**, 2153-2159.

Johnson J.R. & Stell A.L. (2000) Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* **181**, 261-272.

Jolley K.A., Feil E.J., Chan M.S. & Maiden M.C. (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**, 1230-1231.

Jones G.L., Warren R.E., Skidmore S.J., Davies V.A., Gibreel T. & Upton M. (2008) Prevalence and distribution of plasmid-mediated quinolone resistance genes in clinical isolates of Escherichia coli lacking extended-spectrum beta-lactamases. *J Antimicrob Chemother* **62**, 1245-1251.

Jordan R., Van Heerden E., Hugo C.J. & Piater L.A. (2009) Using current molecular techniques for rapid differentiation of Salmonella Typhi and Salmonella Typhimurium. *African Journal of Biotechnology* **8**, 1815-1818.

Kado C.I. & Liu S.T. (1981) Rapid procedure for detection and isolation of large and small plasmids. *J Bacteriol* **145**, 1365-1373.

Kanamaru S., Kurazono H., Nakano M., Terai A., Ogawa O. & Yamamoto S. (2006a) Subtyping of uropathogenic Escherichia coli according to the pathogenicity island encoding uropathogenic-specific protein: comparison with phylogenetic groups. *Int J Urol* **13**, 754-760.

Kanamaru S., Kurazono H., Terai A., Monden K., Kumon H., Mizunoe Y., Ogawa O. & Yamamoto S. (2006b) Increased biofilm formation in Escherichia coli isolated from acute prostatitis. *Int J Antimicrob Agents* **28 Suppl 1**, S21-25.

Kao J.S., Stucker D.M., Warren J.W. & Mobley H.L. (1997) Pathogenicity island sequences of pyelonephritogenic Escherichia coli CFT073 are associated with virulent uropathogenic strains. *Infect Immun* **65**, 2812-2820.

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

Karisik E., Ellington M.J., Pike R., Warren R.E., Livermore D.M. & Woodford N. (2006) Molecular characterization of plasmids encoding CTX-M-15 beta-lactamases from Escherichia coli strains in the United Kingdom. *J Antimicrob Chemother* **58**, 665-668.

Katouli M., Kuhn I. & Mollby R. (1990) Evaluation of the stability of biochemical phenotypes of Escherichia coli upon subculturing and storage. *J Gen Microbiol* **136**, 1681-1688.

Kerenyi M., Allison H.E., Batai I., Sonnevend A., Emody L., Plaveczky N. & Pal T. (2005) Occurrence of hlyA and sheA genes in extraintestinal Escherichia coli strains. *J Clin Microbiol* **43**, 2965-2968.

Kilic A., Bedir O., Kocak N., Levent B., Eyigun C.P., Tekbas O.F., Gorenek L., Baylan O. & Basustaoglu A.C. (2010) Analysis of an outbreak of Salmonella enteritidis by repetitive-sequence-based PCR and pulsed-field gel electrophoresis. *Intern Med* **49**, 31-36.

Kimura M. (1983) Rare variant alleles in the light of the neutral theory. *Mol Biol Evol* **1**, 84-93.

Kramer A., Schwebke I. & Kampf G. (2006) How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases* **6**.

Kryazhimskiy S. & Plotkin J.B. (2008) The population genetics of dN/dS. *PLoS Genet* **4**, e1000304.

Kuhn I. (1985) Biochemical fingerprinting of Escherichia coli: A simple method for epidemiological investigations. *Journal of Microbiological Methods* **3**, 159-170.

Kuhn I., Burman L.G., Haeggman S., Tullus K. & Murray B.E. (1995) Biochemical fingerprinting compared with ribotyping and pulsed-field gel electrophoresis of DNA for epidemiological typing of enterococci. *J Clin Microbiol* **33**, 2812-2817.

Kuhnert P., Boerlin P. & Frey J. (2000) Target genes for virulence assessment of Escherichia coli isolates from water, food and the environment. *FEMS Microbiol Rev* **24**, 107-117.

Labat F., Pradillon O., Garry L., Peuchmaur M., Fantin B. & Denamur E. (2005) Mutator phenotype confers advantage in Escherichia coli chronic urinary tract infection pathogenesis. *FEMS Immunol Med Microbiol* **44**, 317-321.

Lacher D.W., Steinsland H., Blank T.E., Donnenberg M.S. & Whittam T.S. (2007) Molecular evolution of typical enteropathogenic Escherichia coli: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. *J Bacteriol* **189**, 342-350.

Landraud L., Gibert M., Popoff M.R., Boquet P. & Gauthier M. (2003) Expression of cnf1 by Escherichia coli J96 involves a large upstream DNA region including the hlyCABD operon, and is regulated by the RfaH protein. *Mol Microbiol* **47**, 1653-1667.

Lau S.H., Cheesborough J., Kaufmann M.E., Woodford N., Dodgson A.R., Dodgson K.J., Bolton E.J., Fox A.J. & Upton M. (2010) Rapid identification of uropathogenic Escherichia coli of the O25:H4-ST131 clonal lineage using the DiversiLab repetitive sequence-based PCR system. *Clin Microbiol Infect* **16**, 232-237.

Lau S.H., Kaufmann M.E., Livermore D.M., Woodford N., Willshaw G.A., Cheasty T., Stamper K., Reddy S., Cheesbrough J., Bolton F.J., Fox A.J. & Upton M. (2008a) UK epidemic Escherichia coli strains A-E, with CTX-M-15 beta-lactamase, all belong to the international O25:H4-ST131 clone. *J Antimicrob Chemother* **62**, 1241-1244.

Lau S.H., Reddy S., Cheesbrough J., Bolton F.J., Willshaw G., Cheasty T., Fox A.J. & Upton M. (2008b) Major uropathogenic Escherichia coli strain isolated in the northwest of England identified by multilocus sequence typing. *J Clin Microbiol* **46**, 1076-1080.

Lavigne J.P., Marchandin H., Delmas J., Moreau J., Bouziges N., Lecaillon E., Cavalie L., Jean-Pierre H., Bonnet R. & Sotto A. (2007) CTX-M beta-lactamase-producing

Escherichia coli in French hospitals: prevalence, molecular epidemiology, and risk factors. *Journal of clinical microbiology* **45**, 620-626.

Lawrence J.G. & Ochman H. (1998) Molecular archaeology of the Escherichia coli genome. *Proc Natl Acad Sci U S A* **95**, 9413-9417.

Le Bouguenec C. & Schouler C. (2010) Sugar metabolism, an additional virulence factor in enterobacteria. *Int J Med Microbiol*.

LeClerc J.E., Li B., Payne W.L. & Cebula T.A. (1996) High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science* **274**, 1208-1211.

Leclercq A., Lambert B., Pierard D. & Mahillon J. (2001) Particular biochemical profiles for enterohemorrhagic Escherichia coli O157:H7 isolates on the ID 32E system. *J Clin Microbiol* **39**, 1161-1164.

LeCointre G., Rachdi L., Darlu P. & Denamur E. (1998) Escherichia coli molecular phylogeny using the incongruence length difference test. *Molecular Biology and Evolution* **15**, 1685-1695.

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. *J Clin Microbiol* **46**, 3900-3905.

Leflon-Guibout V., Jurand C., Bonacorsi S., Espinasse F., Guelfi M.C., Duportail F., Heym B., Bingen E. & Nicolas-Chanoine M.H. (2004) Emergence and spread of three clonally related virulent isolates of CTX-M-15-producing Escherichia coli with variable resistance to aminoglycosides and tetracycline in a French geriatric hospital. *Antimicrob Agents Chemother* **48**, 3736-3742.

Levinson G. & Gutman G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**, 203-221.

Liljestrom P.L. & Liljestrom P. (1987) Nucleotide sequence of the melA gene, coding for alpha-galactosidase in Escherichia coli K-12. *Nucleic Acids Res* **15**, 2213-2220.

Livermore D.M., Nichols T., Lamagni T.L., Potz N., Reynolds R. & Duckworth G. (2003) Ciprofloxacin-resistant Escherichia coli from bacteraemias in England; increasingly prevalent and mostly from men. *J Antimicrob Chemother* **52**, 1040-1042.

Lloyd A.L., Rasko D.A. & Mobley H.L. (2007) Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. *J Bacteriol* **189**, 3532-3546.

Louie M., Read S., Louie L., Ziebell K., Rahn K., Borczyk A. & Lior H. (1999) Molecular typing methods to investigate transmission of Escherichia coli O157:H7 from cattle to humans. *Epidemiol Infect* **123**, 17-24.

Mahan M.J. & Roth J.R. (1991) Ability of a bacterial chromosome segment to invert is dictated by included material rather than flanking sequence. *Genetics* **129**, 1021-1032.
Maiden M.C. (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**, 561-588.

Maiden M.C., Bygraves J.A., Feil E., Morelli G., Russell J.E., Urwin R., Zhang Q., Zhou J., Zurth K., Caugant D.A., Feavers I.M., Achtman M. & Spratt B.G. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140-3145.

Manges A.R., Dietrich P.S. & Riley L.W. (2004) Multidrug-resistant Escherichia coli clonal groups causing community-acquired pyelonephritis. *Clin Infect Dis* **38**, 329-334.

Manges A.R., Johnson J.R., Foxman B., O'Bryan T.T., Fullerton K.E. & Riley L.W. (2001) Widespread distribution of urinary tract infections caused by a multidrug-resistant Escherichia coli clonal group. *N Engl J Med* **345**, 1007-1013.

Manges A.R., Perdreau-Remington F., Solberg O. & Riley L.W. (2006) Multidrug-resistant Escherichia coli clonal groups causing community-acquired bloodstream infections. *J Infect* **53**, 25-29.

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. *Emerg Infect Dis* **14**, 1575-1583.

Marrs C.F., Zhang L. & Foxman B. (2005) Escherichia coli mediated urinary tract infections: are there distinct uropathogenic E. coli (UPEC) pathotypes? *FEMS Microbiol Lett* **252**, 183-190.

Martinez-Medina M., Naves P., Blanco J., Aldeguer X., Blanco J.E., Blanco M., Ponte C., Soriano F., Darfeuille-Michaud A. & Garcia-Gil L.J. (2009) Biofilm formation as a novel phenotypic feature of adherent-invasive Escherichia coli (AIEC). *BMC Microbiol* **9**, 202.

Martinez J.L. & Baquero F. (2000) Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother* **44**, 1771-1777.

Mayer L.W. (1988) Use of plasmid profiles in epidemiologic surveillance of disease outbreaks and in tracing the transmission of antibiotic resistance. *Clinical Microbiology Reviews* **1**, 228-243.

Miller K., O'Neill A.J. & Chopra I. (2004) Escherichia coli mutators present an enhanced risk for emergence of antibiotic resistance during urinary tract infections. *Antimicrob Agents Chemother* **48**, 23-29.

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. *Infect Immun* **68**, 5869-5880.

Mingeot-Leclercq M.P., Glupczynski Y. & Tulkens P.M. (1999) Aminoglycosides: activity and resistance. *Antimicrob Agents Chemother* **43**, 727-737.

Mohan V.P., Sharma K.B., Agarwal D.S., Purnima G. & Pillai P.K. (1995) Plasmid profile and phage type of Salmonella typhimurium strains encountered in different regions of India. *Comp Immunol Microbiol Infect Dis* **18**, 283-290.

Moore K.N., Day R.A. & Albers M. (2002) Pathogenesis of urinary tract infections: a review. *Journal of clinical nursing* **11**, 568-574.

Mora A., Lopez C., Dabhi G., Blanco M., Blanco J.E., Alonso M.P., Herrera A., Mamani R., Bonacorsi S., Moulin-Schouleur M. & Blanco J. (2009) Extraintestinal pathogenic Escherichia coli O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol* **9**, 132.

Moreno E., Andreu A., Pigrau C., Kuskowski M.A., Johnson J.R. & Prats G. (2008) Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. *J Clin Microbiol* **46**, 2529-2534.

Moreno E., Johnson J.R., Perez T., Prats G., Kuskowski M.A. & Andreu A. (2009) Structure and urovirulence characteristics of the fecal Escherichia coli population among healthy women. *Microbes Infect* **11**, 274-280.

Moreno E., Planells I., Prats G., Planes A.M., Moreno G. & Andreu A. (2005) Comparative study of Escherichia coli virulence determinants in strains causing urinary tract bacteremia versus strains causing pyelonephritis and other sources of bacteremia. *Diagn Microbiol Infect Dis* **53**, 93-99.

Moreno E., Prats G., Sabate M., Perez T., Johnson J.R. & Andreu A. (2006) Quinolone, fluoroquinolone and trimethoprim/sulfamethoxazole resistance in relation to virulence determinants and phylogenetic background among uropathogenic Escherichia coli. *J Antimicrob Chemother* **57**, 204-211.

Morrison D.A. (1996) Phylogenetic tree-building. *International Journal for Parasitology* **26**, 589-617.

Mugnaioli C., Luzzaro F., De Luca F., Brigante G., Perilli M., Amicosante G., Stefani S., Toniolo A. & Rossolini G.M. (2006) CTX-M-type extended-spectrum beta-lactamases in Italy: molecular epidemiology of an emerging countrywide problem. *Antimicrob Agents Chemother* **50**, 2700-2706.

Naber K.G., Bergman B., Bishop M.C., Bjerklund-Johansen T.E., Botto H., Lobel B., Jinenez Cruz F. & Selvaggi F.P. (2001) EAU guidelines for the management of urinary and male genital tract infections. Urinary Tract Infection (UTI) Working Group of the Health Care Office (HCO) of the European Association of Urology (EAU). *Eur Urol* **40**, 576-588.

Nei M. (1996) Phylogenetic analysis in molecular evolutionary genetics. In: *Annual Review of Genetics*, Vol. 30, pp. 371-403.

Nemoy L.L., Kotetishvili M., Tigno J., Keefer-Norris A., Harris A.D., Perencevich E.N., Johnson J.A., Torpey D., Sulakvelidze A., Morris J.G., Jr. & Stine O.C. (2005) Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing Escherichia coli isolates. *J Clin Microbiol* **43**, 1776-1781.

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

Noller A.C., McEllistrem M.C., Stine O.C., Morris J.G., Jr., Boxrud D.J., Dixon B. & Harrison L.H. (2003) Multilocus sequence typing reveals a lack of diversity among Escherichia coli O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol* **41**, 675-679.

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

Ochman H. & Selander R.K. (1984) Evidence for clonal population structure in Escherichia coli. *Proc Natl Acad Sci U S A* **81**, 198-201.

Octavia S. & Lan R. (2007) Single-nucleotide-polymorphism typing and genetic relationships of Salmonella enterica serovar Typhi isolates. *Journal of Clinical Microbiology* **45**, 3795-3801.

Oelschlaeger T.A., Dobrindt U. & Hacker J. (2002) Pathogenicity islands of uropathogenic E. coli and the evolution of virulence. *Int J Antimicrob Agents* **19**, 517-521.

Olive D.M. & Bean P. (1999) Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* **37**, 1661-1669.

Olsen J.E., Brown D.J., Baggesen D.L. & Bisgaard M. (1992) Biochemical and molecular characterization of Salmonella enterica serovar berta, and comparison of methods for typing. *Epidemiology and Infection* **108**, 243-260.

Olsen J.E., Brown D.J., Skov M.N. & Christensen J.P. (1993) Bacterial typing methods suitable for epidemiological analysis. Applications in investigations of salmonellosis among livestock. *Vet Q* **15**, 125-135.

Ong C.L., Ulett G.C., Mabbett A.N., Beatson S.A., Webb R.I., Monaghan W., Nimmo G.R., Looke D.F., McEwan A.G. & Schembri M.A. (2008) Identification of type 3 fimbriae in uropathogenic Escherichia coli reveals a role in biofilm formation. *J Bacteriol* **190**, 1054-1063.

Orskov F. & Orskov I. (1992) Escherichia coli serotyping and disease in man and animals. *Can J Microbiol* **38**, 699-704.

Paiva J.B., Cavallini J.S., Silva M.D., Almeida M.A., Ã,ngela H.I. & Berchieri J.A. Molecular differentiation of Salmonella Gallinarum and Salmonella Pullorum by RFLP of flic gene from Brazilian isolates. *Revista Brasileira de Ciencia Avicola* **11**, 271-276.

Pancholi V. & Chhatwal G.S. (2003) Housekeeping enzymes as virulence factors for pathogens. *Int J Med Microbiol* **293**, 391-401.

Parsek M.R. & Singh P.K. (2003) Bacterial biofilms: an emerging link to disease pathogenesis. *Annu Rev Microbiol* **57**, 677-701.

Parveen S., Hodge N.C., Stall R.E., Farrah S.R. & Tamplin M.L. (2001) Phenotypic and genotypic characterization of human and nonhuman Escherichia coli. *Water Res* **35**, 379-386.

Parveen S., Portier K.M., Robinson K., Edmiston L. & Tamplin M.L. (1999) Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. *Applied and Environmental Microbiology* **65**, 3142-3147.

Pearson T., Busch J.D., Ravel J., Read T.D., Rhoton S.D., U'Ren J.M., Simonson T.S., Kachur S.M., Leadem R.R., Cardon M.L., Van Ert M.N., Huynh L.Y., Fraser C.M. & Keim P. (2004) Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13536-13541.

Peirano G. & Pitout J.D. (2010) Molecular epidemiology of Escherichia coli producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4. *Int J Antimicrob Agents* **35**, 316-321.

Perez-Losada M., Browne E.B., Madsen A., Wirth T., Viscidi R.P. & Crandall K.A. (2006) Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* **6**, 97-112.

Petersen A.M., Nielsen E.M., Litrup E., Brynskov J., Mirsepasi H. & Krogfelt K.A. (2009) A phylogenetic group of Escherichia coli associated with active left-sided inflammatory bowel disease. *BMC Microbiol* **9**, 171.

Philippon A., Labia R. & Jacoby G. (1989) Extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* **33**, 1131-1136.

Phillips I., Eykyn S., King A., Gransden W.R., Rowe B., Frost J.A. & Gross R.J. (1988) Epidemic multiresistant Escherichia coli infection in West Lambeth Health District. *Lancet* **1**, 1038-1041.

Piatti G., Mannini A., Balistreri M. & Schito A.M. (2008) Virulence factors in urinary Escherichia coli strains: phylogenetic background and quinolone and fluoroquinolone resistance. *J Clin Microbiol* **46**, 480-487.

Pitout J.D. (2009) Recent changes in the epidemiology and management of extended-spectrum beta-lactamase-producing Enterobacteriaceae. *F1000 Med Rep* **1**.

Plos K., Hull S.I., Hull R.A., Levin B.R., Orskov I., Orskov F. & Svanborg-Eden C. (1989) Distribution of the P-associated-pilus (pap) region among Escherichia coli from natural sources: evidence for horizontal gene transfer. *Infect Immun* **57**, 1604-1611.

Popovic T., Bopp C., Olsvik O. & Wachsmuth K. (1993) Epidemiologic application of a standardized ribotype scheme for Vibrio cholerae O1. *Journal of Clinical Microbiology* **31**, 2474-2482.

Power E.G. (1996) RAPD typing in microbiology--a technical review. *J Hosp Infect* **34**, 247-265.

Prats G., Navarro F., Mirelis B., Dalmau D., Margall N., Coll P., Stell A. & Johnson J.R. (2000) Escherichia coli serotype O15:K52:H1 as a uropathogenic clone. *J Clin Microbiol* **38**, 201-209.

Ramos N.L., Saayman M.L., Chapman T.A., Tucker J.R., Smith H.V., Faoagali J., Chin J.C., Brauner A. & Katouli M. (2010) Genetic relatedness and virulence gene profiles of Escherichia coli strains isolated from septicaemic and uroseptic patients. *Eur J Clin Microbiol Infect Dis* **29**, 15-23.

Rangel J.M., Sparling P.H., Crowe C., Griffin P.M. & Swerdlow D.L. (2005) Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerg Infect Dis* **11**, 603-609.

Ratkai C., Peixe L.V., Grosso F., Freitas A.R., Antunes P., Fodor E., Hajdu E. & Nagy E. (2010) Successful application of the DiversiLab repetitive-sequence-based PCR typing system for confirmation of the circulation of a multiresistant Pseudomonas aeruginosa clone in different hospital wards. *Diagn Microbiol Infect Dis* **67**, 202-206.

Reid S.D., Herbelin C.J., Bumbaugh A.C., Selander R.K. & Whittam T.S. (2000) Parallel evolution of virulence in pathogenic Escherichia coli. *Nature* **406**, 64-67.

Reinert R.R., Reinert S., van der Linden M., Cil M.Y., Al-Lahham A. & Appelbaum P. (2005) Antimicrobial susceptibility of Streptococcus pneumoniae in eight European countries from 2001 to 2003. *Antimicrob Agents Chemother* **49**, 2903-2913.

Ribot E.M., Fair M.A., Gautom R., Cameron D.N., Hunter S.B., Swaminathan B. & Barrett T.J. (2006) Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet. *Foodborne Pathog Dis* **3**, 59-67.

Rijavec M., Muller-Premru M., Zakotnik B. & Zgur-Bertok D. (2008) Virulence factors and biofilm production among Escherichia coli strains causing bacteraemia of urinary tract origin. *J Med Microbiol* **57**, 1329-1334.

Rocha E.P., Smith J.M., Hurst L.D., Holden M.T., Cooper J.E., Smith N.H. & Feil E.J. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**, 226-235.

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. *Mol Microbiol* **49**, 55-67.

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 Med* **4**, e329.

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 / Institut Pasteur* **5**, 449-456.

Sabtcheva S., Kaku M., Saga T., Ishii Y. & Kantardjiev T. (2009) High prevalence of the aac(6')-Ib-cr gene and its dissemination among Enterobacteriaceae isolates by CTX-M-15 plasmids in Bulgaria. *Antimicrob Agents Chemother* **53**, 335-336.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A. & Arnheim N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

Saladin M., Cao V.T., Lambert T., Donay J.L., Herrmann J.L., Ould-Hocine Z., Verdet C., Delisle F., Philippon A. & Arlet G. (2002) Diversity of CTX-M beta-lactamases and their promoter regions from Enterobacteriaceae isolated in three Parisian hospitals. *FEMS Microbiol Lett* **209**, 161-168.

Samrook J. & Russell D.W. (2001) *Molecular cloning a laboratory manual*. CSHL press, new york.

Schilling J.D. & Hultgren S.J. (2002) Recent advances into the pathogenesis of recurrent urinary tract infections: the bladder as a reservoir for uropathogenic Escherichia coli. *Int J Antimicrob Agents* **19**, 457-460.

Schmidt H. & Hensel M. (2004) Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* **17**, 14-56.

Selander R.K., Caugant D.A., Ochman H., Musser J.M., Gilmour M.N. & Whittam T.S. (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* **51**, 873-884.

Selander R.K. & Levin B.R. (1980) Genetic diversity and structure in Escherichia coli populations. *Science (New York, N.Y* **210**, 545-547.

Shima K., Yoshii N., Akiba M., Nishimura K., Nakazawa M. & Yamasaki S. (2006) Comparison of PCR-RFLP and PFGE for determining the clonality of enterohemorrhagic Escherichia coli strains. *FEMS Microbiol Lett* **257**, 124-131.

Simpson E.H. (1949) Measurement of diversity [16]. *Nature* **163**, 688.

Skold O. (2001) Resistance to trimethoprim and sulfonamides. *Vet Res* **32**, 261-273.

Smith J.M. (1992) Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.

Smith J.M., Smith N.H., O'Rourke M. & Spratt B.G. (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4384-4388.

Smith Y.C., Rasmussen S.B., Grande K.K., Conran R.M. & O'Brien A.D. (2008) Hemolysin of uropathogenic Escherichia coli evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice. *Infect Immun* **76**, 2978-2990.

Snyder L. & Champness W. (2007a) *"Bacterial genetic analysis" In: Molecular genetics of bacteria*. 3rd edition edn. ASM press pp 1153-161, Washington, DC.

Snyder L. & Champness W. (2007b) *"molecular mechanisms of homologous recombination" In: Molecular genetics of bacteria*. 3rd edn. ASM press pp 430-452, Washington DC.

Snyder L. & Champness W. (2007c) *"Plasmids" In: Molecular genetics of bacteria*. 3rd edn. ASM press pp197-241, Washington, DC.

Soll D.R., Pujol C. & Lockhart S.R. (2007) *laboratory procedures for the epidemiological analysis of mcroorganisms In: manual of clinical microbiology pp129-151*. 9th Edition edn. ASM press, Washington DC.

Soto S.M., Jimenez de Anta M.T. & Vila J. (2006) Quinolones induce partial or total loss of pathogenicity islands in uropathogenic Escherichia coli by SOS-dependent or -independent pathways, respectively. *Antimicrob Agents Chemother* **50**, 649-653.

Spratt B.G. (1999) Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr Opin Microbiol* **2**, 312-316.

Spratt B.G. & Maiden M.C. (1999) Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B Biol Sci* **354**, 701-710.

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. *J Clin Microbiol* **41**, 285-289.

Stamm W.E. (2001) An epidemic of urinary tract infections? *N Engl J Med* **345**, 1055-1057.

Starcic Erjavec M., Rijavec M., Krizan-Hergouth V., Fruth A. & Zgur-Bertok D. (2007) Chloramphenicol- and tetracycline-resistant uropathogenic Escherichia coli (UPEC) exhibit reduced virulence potential. *Int J Antimicrob Agents* **30**, 436-442.

Steinke D.T., Seaton R.A., Phillips G., MacDonald T.M. & Davey P.G. (2001) Prior trimethoprim use and trimethoprim-resistant urinary tract infection: a nested case-control study with multivariate analysis for other risk factors. *J Antimicrob Chemother* **47**, 781-787.

Tarkka E., Ahman H. & Siitonen A. (1994) Ribotyping as an epidemiologic tool for Escherichia coli. *Epidemiology and Infection* **112**, 263-274.

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. *J Clin Microbiol* **43**, 5860-5864.

Tenover F.C., Arbeit R.D. & Goering R.V. (1997) How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol* **18**, 426-439.

Tenover F.C., Arbeit R.D., Goering R.V., Mickelsen P.A., Murray B.E., Persing D.H. & Swaminathan B. (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**, 2233-2239.

Threlfall E.J., Frost J.A., Ward L.R. & Rowe B. (1990) Plasmid profile typing can be used to subdivide phage-type 49 of Salmonella typhimurium in outbreak investigations. *Epidemiol Infect* **104**, 243-251.

Trevors J.T. (1998) Bacterial population genetics. *World Journal of Microbiology and Biotechnology* **14**, 1-5.

Turner K.M., Hanage W.P., Fraser C., Connor T.R. & Spratt B.G. (2007) Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* **7**, 30.

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. *Int J Antimicrob Agents* **35**, 444-450.

Urwin R. & Maiden M.C. (2003) Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* **11**, 479-487.

Valverde A., Canton R., Garcillan-Barcia M.P., Novais A., Galan J.C., Alvarado A., de la Cruz F., Baquero F. & Coque T.M. (2009) Spread of bla(CTX-M-14) is driven mainly by IncK plasmids disseminated among Escherichia coli phylogroups A, B1, and D in Spain. *Antimicrob Agents Chemother* **53**, 5204-5212.

van Belkum A., Struelens M., de Visser A., Verbrugh H. & Tibayrenc M. (2001) Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin Microbiol Rev* **14**, 547-560.

Vidovic S., Germida J.J. & Korber D.R. (2007) Sensitivity of two techniques to detect Escherichia coli O157 in naturally infected bovine fecal samples. *Food Microbiol* **24**, 633-639.

Vila J., Simon K., Ruiz J., Horcajada J.P., Velasco M., Barranco M., Moreno A. & Mensa J. (2002) Are quinolone-resistant uropathogenic Escherichia coli less virulent? *J Infect Dis* **186**, 1039-1042.

Walker R.C. (1999) The fluoroquinolones. *Mayo Clin Proc* **74**, 1030-1037.

Weintraub A. (2007) Enteroaggregative Escherichia coli: epidemiology, virulence and detection. *Journal of medical microbiology* **56**, 4-8.

Whelan S., Lio P. & Goldman N. (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet* **17**, 262-272.

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

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. *Antimicrob Agents Chemother* **53**, 4472-4482.

Woodford N., Fagan E.J. & Ellington M.J. (2006) Multiplex PCR for rapid detection of genes encoding CTX-M extended-spectrum (beta)-lactamases. *J Antimicrob Chemother* **57**, 154-155.

Woodford N., Kaufmann M.E., Karisik E. & Hartley J.W. (2007) Molecular epidemiology of multiresistant Escherichia coli isolates from community-onset urinary tract infections in Cornwall, England. *J Antimicrob Chemother* **59**, 106-109.

Xie J., Foxman B., Zhang L. & Marrs C.F. (2006) Molecular epidemiologic identification of Escherichia coli genes that are potentially involved in movement of the organism from the intestinal tract to the vagina and bladder. *J Clin Microbiol* **44**, 2434-2441.

Yamamoto S. (2007) Molecular epidemiology of uropathogenic Escherichia coli. *J Infect Chemother* **13**, 68-73.

Yu X., Susa M., Weile J., Knabbe C., Schmid R.D. & Bachmann T.T. (2007) Rapid and sensitive detection of fluoroquinolone-resistant Escherichia coli from urine samples using a genotyping DNA microarray. *Int J Med Microbiol*.

Yue L., Jiang H.X., Liao X.P., Liu J.H., Li S.J., Chen X.Y., Chen C.X., Lu D.H. & Liu Y.H. (2008) Prevalence of plasmid-mediated quinolone resistance qnr genes in poultry and swine clinical isolates of Escherichia coli. *Vet Microbiol* **132**, 414-420.

Yumuk Z., Afacan G., Nicolas-Chanoine M.H., Sotto A. & Lavigne J.P. (2008) Turkey: a further country concerned by community-acquired Escherichia coli clone O25-ST131 producing CTX-M-15. *J Antimicrob Chemother* **62**, 284-288.

Zdziarski J., Svanborg C., Wullt B., Hacker J. & Dobrindt U. (2008) Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect Immun* **76**, 695-703.

Zhang L., Foxman B., Manning S.D., Tallman P. & Marrs C.F. (2000) Molecular epidemiologic approaches to urinary tract infection gene discovery in uropathogenic Escherichia coli. *Infect Immun* **68**, 2009-2015.

Zhang W., Qi W., Albert T.J., Motiwala A.S., Alland D., Hyytia-Trees E.K., Ribot E.M., Fields P.I., Whittam T.S. & Swaminathan B. (2006) Probing genomic diversity and evolution of Escherichia coli O157 by single nucleotide polymorphisms. *Genome Research* **16**, 757-767.