ANTIMICROBIAL RESISTANCE AND GENOMIC DIVERSITY OF *ESCHERICHIA COLI* STRAINS IN MALAYSIA

HO WING SZE

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

ANTIMICROBIAL RESISTANCE AND GENOMIC DIVERSITY OF *ESCHERICHIA COLI* STRAINS IN MALAYSIA

HO WING SZE

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: HO WING SZE

Registration/Matric No: SHC100015

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"): ANTIMICROBIAL RESISTANCE AND GENOMIC DIVERSITY OF *ESCHERICHIA COLI* STRAINS IN MALAYSIA

Field of Study: MOLECULAR MICROBIOLOGY

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Escherichia coli is an important nosocomial pathogen and extended-spectrum cephalosporin (ESC) resistance *E. coli* is recognized as the global drug-resistant threat. This study aims to determine the resistance mechanisms of ESC and genetic diversity of Malaysian *E. coli* strains using various phenotypic and genotypic approaches.

One hundred ten non-repeat *E. coli* strains were isolated from the stools of paediatric patients in 2009 from a tertiary hospital in Malaysia. Twenty-one were categorized as extended spectrum beta-lactamase (ESBL)-producers using the double-disk synergy test and ESBL E-test. CTX-M-15 was the pre-dominant CTX-M variant (77%). Class 1 integron was the most common class of integrons found in the *E. coli* strains (76%); however, they lack gene cassette encoding ESBL genes. Using PFGE, three strains displayed DNA degradation (Dnd) phenotype and few CTX-M-15-positive strains with indistinguishable pulsotypes were identified. To better understand the dissemination of CTX-M genes, *E. coli* strains from clinical, zoonotic, food and environment samples isolated between 2002 - 2011 in Malaysia were screened for ESBL-production. Using multi-locus sequence typing (MLST), 6 out of 35 CTX-M positive *E. coli* strains belonged to the sequence type (ST) ST131, an *E. coli* clone notorious for the global dissemination of CTX-M genes. Insertional sequences IS*26* and ISE*cp1* were the most common genetic environments for CTX-M genes in the Malaysian *E. coli* strains.

The degradation phenotype which was observed earlier in this thesis was overcome for 12 Dnd+ *E. coli* strains by adding thiourea into the PFGE run. The *dnd* operon, responsible for the phosphorothioation modification of DNA, was detected in all Dnd^{+} *E. coli* strains by PCR. Further genomic analysis of the genetic environment of *dnd* operons of 52 global *E. coli* genomes revealed a total of 7 types of *dnd*-encoding genomic islands, indicating substantial diversity in these regions. *dnd* operons were more often found in pathogenic *E. coli,* indicating a possible linkage of the *dnd* operons with *E. coli* pathogenicity.

Two highly related *E. coli* strains EC302/04 and EC096/10 with single band difference in their pulsotypes (identified in the first part of this study) were investigated using whole genome sequencing (WGS). EC302/04 and EC096/10 that were isolated from tracheal aspirate and a stool specimen, respectively, shared the same sequence type ST349, and serotype O166:H5. EC302/04 harboured a 140,232 bp IncFII plasmid, pEC302/04, that was absent in EC096/10. Plasmid pEC302/04 is self-transferable and carried resistance genes (bla_{TEM-1}*;* integron-encoded- *sul1*, *cml*, and *aadA*) and two iron-acquisition systems. Phylogenomic analysis with 38 global *E. coli* of various pathotypes revealed that ExPEC and intestinal strains may share similar phylogenetic signals. This may help in identifying commensal-like strains with extraintestinal virulence potential. Phylogenomic analysis also unveiled ExPEC with rare genotypes, indicating the importance of genomic approach in detecting the potential emergence of new ExPEC lineage(s).

Besides providing comprehensive understanding on the resistance mechanisms of the clinically important CTX-M genes, this thesis also demonstrates the usefulness of whole genome sequence approach in investigating the resistance mechanisms and genetic diversity of specific region or whole genome of various types of *E. coli* strains. (500 words)

ABSTRAK

Escherichia coli merupakan patogen nosokomial yang penting dan *E. coli* yang resisten kepada "extended- spectrum cephalosporin (ESC)" diiktiraf sebagai ancaman kepada kesihatan awam di peringkat dunia. Tesis ini bertujuan untuk menentukan mekanisme resisten terhadap ESC serta mencirikan fenotip dan genotip *E. coli*.

110 *E. coli* telah diisolasi daripada najis pesakit pediatrik pada tahun 2009 dari satu hospital pengajian tinggi di Malaysia. Dengan menggunakan "double disk synergy (DDST)" dan "ESBL E-test", dua puluh satu *E. coli* merupakan produser "extended-spectrum beta-lactamase" (ESBL). CTX-M-15 merupakan varian CTX-M yang paling dominan (77%). Kelas I integron merupakan jenis integron yang paling biasa dijumpai di kalangan strain *E. coli* tempatan (76%). Namunbegitu, tiada gen ESBL ditemui dalam integron. Dengan menggunakan PFGE, tiga *E. coli* adalah "untypeable" (Dnd) dan beberapa isolat CTX -M -15- positif mempunyai "pulsotypes" yang amat sama. Untuk memahami penyebaran gen CTX-M secara mendalam, penentuan produser CTX-M telah dilanjutkan ke kalangan *E. coli* daripada pelbagai sumber (klinikal, zoonotik, makanan dan persekitaran) dari tahun 2002-2011. Dengan menggunakan "multi-locus sequence typing" (MLST) dan PCR, 6 isolat dengan O25-ST131 telah dijumpai di kalangan 35 produser CTX-M, satu klon yang terkenal dengan penyebaran gen CTX-M. "Insertional sequence" IS*26* dan IS*Ecp1* merupakan persekitaran genetik yang paling biasa dijumpai untuk gen CTX -M di kalangan *E. coli* Malaysia.

Fenotip Dnd juga telah diatasi dengan penambahan thiourea dalam PFGE untuk 12 Dnd^+ *E. coli*. Dengan meggunakan PCR, operon *dnd* juga telah dikesan di kalangan Dnd^+ *E. coli*. Analisis genomik persekitaran operon *dnd* operon untuk 52 *E. coli* mendedahkan sejumlah 7 jenis organisasi genetic yang amat berlainan, di mana semua adalah berkaitan dengan "genomic island" (GI). Operon *dnd* juga lebih kerap dijumpai di kalangan *E. coli* yang patogenik, menunjukkan kemungkinan hubungan operon *dnd* dengan patogenisiti *E. coli*.

Dari bahagian pertama, dua isolat *E. coli* daripada sampel klinikal yang berbeza didapati berkait rapat dengan meggunakan PFGE. EC302/04 (daripada "tracheal aspirate") dan EC096/10 (daripada sampel najis) berkongsi "sequence type" (ST349) dan "serotype" (O166:H5) yang sama. Walaubagaimanapun, EC302/04 mempunyai satu plasmid IncFII yang bersaiz 140,232 bp (pEC302/04), dimana pEC302/04 tidak dijumpai dalam EC096/10. Plasmid pEC302/04 adalah "self-transferable" dan mempunyai gen resisten *sul1*, *cml* dan *aadA* dalam integron kelas 1. Selain itu, gen resisten $bla_{\text{TEM -1}}$, dan dua sistem "iron-acquisition" juga ditemui dalam pEC302/04. Analisis filogenomik dengan 38 isolat *E. coli* dengan pelbagai "pathotypes" yang diekstrak daripada pangkalan data awam mendedahkan bahawa ExPEC serta isolat "intestinal" berkongsi isyarat filogenetik yang amat sama. Tesis ini juga memberi penekanan terhadap ancaman isolat "intestinal" dalam menyebabkan jangkitan "extraintestinal". Analisis filogenomik juga telah mendedahkan keupayaan *E. coli* dengan genotip yang amat jarang dalam menyebabkan jangkitan "extraintestinal", menunjukkan kepentingan kaedah filogenomik dalam pengesanan kemunculan klon patogen bakteria yang baru.

Tesis ini memberikan pemahaman komprehensif mengenai mekanisme resisten untuk gen CTX -M yang amat penting di Malaysia. Kajian ini juga menunjukkan kegunaan "whole genome sequencing" dalam menyiasat mekanisme gen resisten dan kepelbagaian genetik untukl seluruh genome atau pun ciri-ciri yang spesifik untuk *E. coli*.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Dr. Thong Kwai Lin and Prof. Dr. Yeo Chew Chieng for giving me a great opportunity to pursue my doctorate under their supervision. I'm extremely gratelful for all the patience, guidance and supervision given during the candidature.

I would like to thank my external collaborators, Dr. Ou Hong Yu, Dr. Ooi Peck Toung, Dr. Frederic Bertel and Dr. Gan Han Ming for all the assistance and help given, specifically:

- Dr. Ou Hong Yu for his expertise in the data analysis and interpretation of the genetic environment of *dnd* operons;
- Dr. Ooi Peck Toung for his veterinary expertise in the data interpretation of zoonotic *E. coli* strains;
- Dr. Frederic Bertel for his expertise in phylogenomic analysis of large number of *E. coli* strains;
- Dr. Gan Han Ming for his assiatance in the initial bacterial genome analysis.

A special thanks to my labmates from Laboratory of Biomedical Science and Molecular Microbiology for the immense moral support and friendliness throughout the project. I will definitely treasure all the helpful advices and gifted invaluable friendship.

Last but not least, I would like to thank my parents, grandparents as well as a special friend for all the immeasurable love and abundant understanding throughout my doctoral work.

I dedicate this thesis to all my loved ones as the completion of this research would not be possible without all of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

3’-CS	:	3' conserved segment
5’-CS	:	5' conserved segment
%	:	Percent
>	:	Greater than
~	:	Approximately
≥	:	Same or greater than
APEC	:	Avian pathogenic *E. coli*
ATCC	:	American Type Culture Collection
bp	:	base pairs
CDC	:	Center for Disease Control and Prevention
CLSI	:	Clinical Laboratory Standards Institute
col	:	colicin
conc.	:	concentration
DAEC	:	Diffusely Adhering *E. coli*
ddH2O	:	double distilled water
DEC	:	Diarreagenic *E. coli*
dH2O	:	distilled water
DIP	:	2′2-dipyridyl
DNA	:	Deoxyribonucleic acid
dnd	:	DNA degradation
dndI	:	*dnd*-encoding genomic island
E. coli	:	*Escherichia coli*
e.g.	:	*exempli gratia*
EAEC	:	Enteroaggregative *E. coli*

EDTA	:	Ethylenediaminetetraacetic
EHEC	:	Enterohaemorrahagic *E. coli*
EIEC	:	Enteroinvasive *E. coli*
EMB	:	Eosin methylene blue
environ.	:	environment
EPEC	:	Enteropathogenic *E. coli*
ESBL	:	Extended Spectrum β-Lactamase
ESC	:	Extended-spectrum cephalosporins (ESCs)
et al.	:	*et alibi*
ETEC	:	Enterotoxigenic *E. coli*
exp.	:	expected
ExPEC	:	Extraintestinal pathogenic *E. coli*
FQ	:	Fluoroquinolone
gbk	:	genbank
GI	:	Genomic island
i.e.	:	*id est*
IMR	:	Institute for Medical Research, Malaysia
IPEC	:	Intestinal pathogenic *E. coli*
IS	:	Insertion sequence
kbp	:	kilo base pair
LBA	:	Luria Bertani agar
LBB	:	Luria Bertani broth
MDR	:	Multi-drug resistance
MLST	:	Multilocus sequence typing
mM	:	millinolar
MOH	:	Ministry of Health

MST	:	Microbial source tracking
mth	:	month
NAG	:	National Antibiotic Guideline
Nd	:	nondeterminate
NMEC	:	Newborn meningitis associated *E. coli*
no.	:	number
nsSNP	:	Non-synonymous single nucleotide polymorphism
ORF	:	Open reading frame
PAI	:	Pathogenic island
PBS	:	Phosphate-buffered saline
PCR	:	Polymerase Chain Reaction
PFGE	:	Pulsed-field gel electrophoresis
PFP	:	Pulsed-field profile
PM	:	Phenotypic microarray
PWD	:	postweaning diarrhea
REP	:	Repetitive element sequence-based
Rep-PCR	:	Repetitive Sequence-Based PCR
rpm	:	rotation per minute
SePEC	:	Sepsis-associated *E.coli*
spp.	:	species
sSNP	:	synonymous single nucleotide polymorphism
ST	:	Sequence type
STEC	:	Shigatoxin producing *E. coli*
TA	:	Toxin-antitoxin
TBE	:	Tris-borate-EDTA
tRNA	:	transfer-RNA

UPGMA	:	Unweighted pair group mathematical average
UTI	:	Urinary tract infection
VF	:	Virulence factors
VTEC	:	Verotoxigenic *E. coli*
WHO	:	World Health Organization
yr	:	year
µg	:	microgram
µl	:	microliter
µm	:	micrometer

LIST OF APPENDICES

CHAPTER 1: GENERAL INTRODUCTION

Escherichia coli is a highly versatile bacterium and is the most extensively studied organism worldwide (Blount, 2015). Besides existing as part of the common microbiota in the gastrointestinal tract of many warm-blooded animals (Luo et al., 2011), *E. coli* can also be pathogenic causing a broad spectrum of intra- and extra-intestinal diseases in both humans and animals (Kaper et al., 2004). Treatment using antibiotics is the usual approach used to prevent and treat bacterial infections in clinical and veterinary settings. However, the indiscriminate use of antibiotics has led to the emergence of multidrug resistance bacteria due to selection pressure (WHO, 2014), resulting in ineffective antimicrobial therapy and increased economic burden. In Malaysia, rapid development of extended cepahlosporin resistance in *E. coli* has also been reported (Lim et al., 2009; Sekawi et al., 2008) and its resistance rate has been escalating for the past few years in the local clinical settings (MOH, 2014a, 2012, 2013). This is certainly a cause for concern since extended-spectrum cephalosporins are often used as first-line antimicrobial agents for treating infections caused by Gram-negative bacteria (MOH, 2014; Winokur et al., 2001). Hence it is a routine practice in clinical laboratories to perform routine surveillance of antibiotic resistance and epidemiological investigation especially among pathogenic bacteria for effective infection control.

Although *E. coli* infections affect patients of all ages, certain types of infections are more common in children under 5 years old (CDC, 2011). Hence, in chapter 3, one-hundred and ten non-repeat *E. coli* strains from stool samples of children under the age of 5 years that were obtained between 2009–2010 from a tertiary hospital in Johor Bahru, a city at the southern tip of Peninsular Malaysia, were characterized. The *E. coli* strains were analyzed for their antibiograms, carriage of extended-spectrum beta-lactamase (ESBL) and AmpC genes which may confer non-susceptibility to extended

cephalosporins, and association of beta-lactamase genes with integron. The genetic diversity of the 110 strains was also investigated using pulsed-field gel electrophoresis (PFGE), a gold standard for *E. coli* subtyping.

To have a better understanding on the dissemination of CTX-M genes, the study was further expanded to investigate the prevalence of CTX-M-encoding genes in 189 non-repeat *E. coli* strains from clinical (n=53), zoonotic (n=47), environmental (n=34) and food (n=55) samples that were isolated over a 10-year period (2002-2011) in Malaysia (Chapter 4). The genetic environments that may have contributed to the dissemination of the CTX-M genes in these *E. coli* strains, in particular insertion sequences and complex integrons, were explored in detail using polymerase chain reaction (PCR) and primer walking. On the other hand, the possible clonal dissemination of *E. coli* strains harbouring CTX-M genes was also investigated using multi-locus sequence typing (MLST) and PFGE.

When using PFGE as typing tool, some *E. coli* strains that exhibited DNA degradation phenotype (Dnd^+) yielded smeared DNA, leading to incomplete molecular epidemiological investigations. Thus, this study also aimed to improve the typeability of the Dnd^+ *E. coli* strains and to determine the genetic mechanism conferring DNA degradation. A PCR assay was developed to detect the *dnd* operons responsible for the *dnd* phenotype of Dnd^+ *E. coli* strains. Fifty-two whole genome sequences of *E. coli* harbouring *dnd* operons were then retrieved from the public database, where further genomic analysis of the *dnd* operons together with their respective immediate genetic environments was performed. The association of *dnd* operons of representative *E. coli* with genomic islands designated *dnd*-encoding GIs was also further investigated. The employment of whole genome sequence analysis in this study revealed substantial diversities among the genetic environment of *dnd* operons in *E. coli*.

Two *E. coli* strains, EC302/04 and EC096/10, which were characterized in Chapter 3 displayed nearly identical pulsed-field profiles (with only a single band difference) and were isolated from different patients of the same hospital over a gap of 5 years. Interestingly, EC302/04 was isolated from tracheal aspirate (i.e., extra-intestinal site) whereas EC096/10 was isolated from a stool specimen, where intra- and extra-intestinal sites belonged to two different sites with distinct nutritional niches. Hence, in Chapter 6, detailed characterization of the two closely related *E. coli* strains was carried out using high resolution technique including Next Generation Sequencing (NGS) technology and Phenotypic Microarray (PM) assay. Genomic differences between EC302/04 and EC096/10 were also investigated in detail. Phylogenomic analysis with 38 other *E. coli* genomes sequences that were retrieved from the public database was also performed to provide further insights into the phylogenetic relationships of *E. coli* of various pathotypes.

Overall, the general goal of this thesis is to investigate the antibiotic resistance mechanism and epidemiology of local *E. coli* strains. The objectives of the study are:

a) To investigate the resistance trend, resistance mechanisms and epidemiology of *E. coli* strains from various sources;
b) To improve the typeability of *E. coli* strains with degradation phenotype and investigate the diversity of the *dnd* operon of *E. coli* strains;
c) To characterize two clonally related *E. coli* strains EC096/10 and EC302/04 that were isolated from intra- and extra-intestinal sites using various phenotypic and genotypic approaches.

The findings of this study will provide insights into the current resistance issues and epidemiological data of local *E. coli* strains.

CHAPTER 2: LITERATURE REVIEW

2.1 *Escherichia coli*

Escherichia coli was first discovered by Theodor Escherich from the feces of a child in 1885 (Escherich, 1886). The genus *Escherichia* belongs to a group of bacteria informally known as "coliforms", and is a member of the *Enterobacteriaceae* family. *E. coli* is a Gram-negative, rod-shaped bacterium and facultative anaerobe that commonly inhabit the lower intestine of warm-blooded organisms. *E. coli* can either be motile or immotile, with its locomotion usually driven by peritrichous flagella (Murray, 1999). Today, *E. coli* is one of the best-studied microorganisms and certain strains such as DH10B, JM109, and K12 are often used as a laboratory workhorse (Cronan, 2014). Nevertheless, *E. coli* strains causing diseases at intraintestinal and extraintestinal sites have been reported worldwide and are of great public health concern. *E. coli* consists of three main clinical groups, namely the commensal, extraintestinal pathogenic *E. coli* (ExPEC) and intestinal pathogenic *E. coli* (IPEC), making them a versatile bacterium (Kaper et al., 2004). The intra- and extra-intestinal sites in human hosts are also known to be two distinct habitats with unique nutritional niches where limited supply of iron and nitrogen sources, which are essential for bacteria growth, were found at extraintestinal sites (Hagan et al., 2010; Sandy & Butler, 2010).

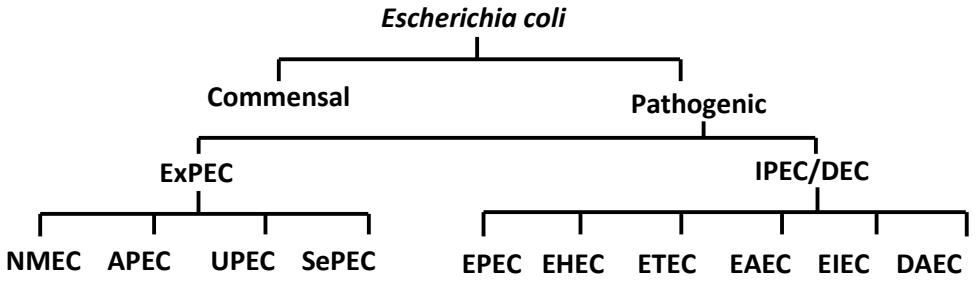


Figure 2.1: Schematic diagram of *E. coli* groups and pathotypes. Pathogenic *E. coli* can be classified as either extraintestinal pathogenic *E. coli* (ExPEC) or intestinal pathogenic *E. coli* (IPEC) (also known as Diarrheagenic *E. coli*, or DEC). The ExPEC can be subdivided into neonatal-meningitis *E. coli* (NMEC), avian pathogenic *E. coli* (APEC), uropathogenic *E. coli* (UPEC) and sepsis-associated pathogenic *E. coli* (SePEC). On the other hand, IPECs can be subdivided into enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC).

2.1.1 Commensal *E. coli*

E. coli typically colonizes the gastrointestinal tract of infants within hours after birth and then establishes mutual benefits with the hosts. *E. coli* exists as harmless symbiont within the host but in rare cases, it may cause infections in immuno-compromised hosts, or when gastrointestinal barriers are violated (Kaper et al., 2004). However, pathogenic *E. coli* harbouring broad range of virulence factors can cause wide spectrum of diseases in both intraintestinal and extraintestinals sites.

2.1.2 Intestinal pathogenic *E. coli*

Diarrheagenic *E. coli* (DEC), which is also known as intestinal pathogenic *E.coli* (IPEC), can cause different types of diarrheal diseases, with six well described pathotypes being recognized up till today (CDC, 2014). Shiga toxin-producing *E. coli* (STEC), which may also be referred to as verocytotoxin-producing *E. coli* (VTEC) or enterohemorrhagic *E. coli* (EHEC) are pathotypes that are most commonly associated with foodborne outbreaks (CDC, 2014). Other diarrheagenic pathotypes are enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli*

(EPEC), enteroaggregative *E. coli* (EAEC) and diffusely adherent *E. coli* (DAEC). Each pathotype is characterized by the possession of specific set of virulence determinants, which is detailed in Table 2.1. The carriage of different virulence factors in *E. coli* determines its pathogenicity and their pathotypes (Bekal et al., 2003; Gómez-Duarte et al., 2009; Nataro & Kaper, 1998; Yu & Thong, 2009). Approximately 2 million foodborne illnesses were documented for IPEC annually in a study which was carried out by Scallan et al. (2011) between years 2000-2008. Of all pathotypes, *E. coli* O157 steals the limelight for causing outbreaks globally, where a total of 390 outbreaks has been reported (as of July 2015), with 65% associated with food while the remaining incidences were due to human contacts with animals, water or human as well (Heiman et al., 2015).

Table 2.1: Clinical and epidemiological features and virulence factors of various *E. coli* pathotypes

Pathotype	Clinical features	Epidemiological features	Virulence factors
Enteropathogenic (EPEC)	Watery diarrhea and vomiting	Infants in developing countries	Bundle-forming pilus, attaching and effacing
Entero-hemorrhagic (EHEC)	Watery diarrhea, hemorrhagic colitis, hemolytic-uremic syndrome	Food-borne, water-borne outbreaks in developed countries	Shiga toxins, attaching and effacing
Enterotoxigenic (ETEC)	Watery diarrhea	Childhood diarrhea in developing countries, traveler's diarrhea	Pili, heat-labile and heat-stable enterotoxins
Enteroaggregative (EAEC)	Diarrhea with mucous	Childhood diarrhea	Pili, cytotoxins
Enteroinvasive (EIEC)	Dysentery, watery diarrhea	Food-borne outbreaks	Cellular invasion, intracellular motility
Diffuse-adhering (DAEC)	Poorly characterized	Older children	Unknown

Adapted from Donnenberg & Whittam (2001)

2.1.3 Extraintestinal pathogenic *E. coli*

Extraintestinal pathogenic *E. coli* (ExPEC) is known to be able to cause extraintestinal infections in almost any organs and represents a major health threat (Russo & Johnson, 2000). ExPEC is able to colonize the intestinal tract of the host asymptomatically in a symbiotic fashion, but has been known to be incapable of causing intestinal infections. Nevertheless, compromised host with weak defense system may be infected with non-virulent *E. coli* strains (Russo & Johnson, 2000). Generally, ExPEC can be determined by the source of isolation (extraintestinal sites) and also the virulence genes they harboured (such as adhesin, iron acquisition systems and protectins) (Köhler & Dobrindt, 2011) (detailed in Table 2.2). However, some *E. coli* from the gut microflora also harboured the above-mentioned virulence genes. Hence, ExPEC are often not easily distinguishable from the commensals despite the difference in isolation sites. Besides, majority of the ExPEC also belonged to phylogenetic group B2, followed by group D at a lesser extent (James R. Johnson & Russo, 2002). In contrast, the IPEC strains often belonged to phylogenetic groups A, B1, D or ungroup lineages (Pupo et al., 1997).

Three main types of clinically-relevant ExPECs have been reported worldwide, namely the UPEC, NMEC and SePEC (Figure 2.1), which are causative agents for urinary tract infections (UTI), neonatal meningitis and sepsis (Johnson & Russo, 2002; Moriel et al., 2010), respectively. On the other hand, APEC, which is an animal-associated *E. coli* pathotype, causes various infections primarily respiratory infections, pericarditis, and septicaemia in poultry (Kaper et al., 2004).

In comparison to the intestinal tract, extraintestinal sites are known to have low abundance of substrates which are essential for bacterial growth such as iron and nitrogen sources. Besides harbouring fitness determinants which can scavenge crucial

nutrients such as iron, vast varieties of virulence factors essential for host mucosal surface colonization, host defense evasion, host invasion, and inflammatory response stimulation are also found in ExPECs (Johnson & Russo, 2002; Table 2.2). Many of these ExPEC-aasociated virulence factors are also encoded on pathogenicity islands (PAIs), which are large, potentially mobile regions of the genome characterized by having a number of mobile genetic elements such as insertion sequences, integrons and transposons (Johnson & Russo, 2002).

Virulence genes associated with ExPEC were reportedly located in both the chromosome and virulence plasmids (Johnson & Nolan, 2009). A bacterial plasmid is an important agent for introducing adaptive traits horizontally to bacterial hosts as well as contributing to bacterial pathogenesis and evolution. Plasmids that carry virulence and resistance genes besides genes essential for their own transmission and maintenance are termed virulence plasmids and resistance plasmids, respectively (Sengupta & Austin, 2011). Large virulence plasmids (>100 kb) belonging to the incompatibility group IncF are widely disseminated in clinically relevant *Enterobacteriaceae* (Villa et al., 2010) including extraintestinal pathogenic *E. coli* (Brolund et al., 2013; Johnson & Nolan, 2009; Johnson et al., 2010; Mellata et al., 2009; Peigne et al., 2009; Woodford et al., 2009). An example of an IncF virulence plasmid that is often associated with ExPEC is the colicin-producing plasmid where colicins are antimicrobial substances produced by certain members of colicin-producing *Enterobacteriaceae* that can kill susceptible strains (Waters & Crosa, 1991). In fact, the successful global dissemination of IncF plasmids may be attributed to its multi-replicon characteristic which enables replication in a broader range of hosts (Villa et al., 2010).

Table 2.2: Virulence factors of ExPEC

Functional category	Virulence factor
Adhesin	Type 1 fimbriae (Fim) P fimbriae (Pap/Prf) S or F1C fimbriae (Sfa/Foc) N-acetyl d-glucosamine-specific fimbriae (Gaf) Heat-resistant agglutinin (Hra) M-agglutinin (Bma) Bifunctional enterobactin receptor/adhesin (Iha) Afimbrial adhesin (Afa) Temperature sensitive hemagglutinin (Tsh)
Invasin	Invasion of brain endothelium (IbeA)
Iron acquisition	Siderophore receptor IreA Aerobactin (Iuc) Yersiniabactin (Ybt) Salmochelin (Iro) Periplasmic iron binding protein (SitA)
Toxins	alpha-Hemolysin (HlyA) Vacuolating toxin (Vat) Enteroaggregative E. coli toxin (AstA) Cytolethal distending toxin IV (CDT 1) Cytotoxic necrotizing factor 1 (CNF-1) Putative hemolysin (HlyF) Colibactin (Clb) Secreted autotransporter toxin (Sat) Serine proteases (Pic)
Protectins	Group II capsule incl. K1 capsule Conjugal transfer surface exclusion protein (TraT) Outer membrane protease T (OmpT) Increased serum survival (Iss) Colicin V (Cva)
Others	d-Serine deaminase (DsdA) Maltose and glucose-specific PTS transporter subunit Pathogenecity island marker (MalX) Flagella Uropathogenic-specific protein (Usp)

Adapted from Köhler & Dobrindt 2011 and Pitout, 2012

2.2 Antimicrobial resistance in *E. coli*

Antimicrobial resistance is now a major clinical problem globally (WHO, 2014). The condition whereby a patient becomes non-responsive to the drugs of choice may not only result in treatment failure, but also lead to the spread of infections. Effective infection control is hence made difficult and antimicrobial resistance has definitely led to increases in the rate of morbidity and mortality along with health care costs (WHO, 2014). *E. coli* is still currently recognized as one of the major causative agents of common health-care associated and community-acquired infections such as UTI, wound infections, bloodstream infections and pneumonia. What makes the current problem more worrisome is the association of *E. coli* with high resistance rates globally (WHO, 2014).

E. coli resistant to third generation cephalosporins has been identified as a public health threat since the beginning of the 21st century (Bradford, 2001; Winokur et al., 2001). Today, the same problem still persists (WHO, 2014). Resistance to cephalosporins is mainly associated with the production of extended-spectrum beta-lactamase (ESBL) (Philippon et al., 2002), with cefotaximase (CTX-M) being the most common ESBL found worldwide (Bonnet, 2004; Cantón & Coque, 2006). *E. coli* strains harbouring $bla_{\text{CTX-M}}$ genes are also often MDR, as CTX-M gene is often located on plasmid carrying other resistance genes that confer resistance to co-trimoxazole, tetracycline, gentamicin, and ciprofloxacin (Pitout & Laupland, 2008).

2.2.1 Extended-spectrum beta-lactamases

Extended-spectrum beta-lactamases (ESBLs) are derived from the non-ESBL-progenitor of β-lactamase, for instance the TEM-1, TEM-2 and SHV-1 β-lactamases, which can only hydrolyze penicillin, ampicillin, and to a lesser degree, carbenicillin and cephalothin, but not extended-spectrum cephalosporins and monobactams (Bush et al.,

1995; Paterson & Bonomo, 2005). With the gain of mutations in the non-ESBL-progenitor of β-lactamase, the ESBL enzymes are made capable of hydrolyzing third-generation cephalosporins (e.g., cefotaxime, ceftriaxone, ceftazidime) and monobactams (e.g., aztreonam) by breaking their β-lactam rings (Figure 2.2), but not carbapenems and cephamycins. ESBLs are also inhibited by beta-lactamase inhibitors such as clavulanic acid (Bradford, 2001).

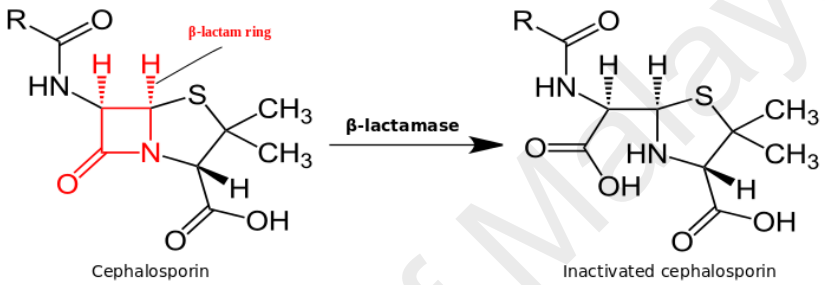


Figure 2.2: Inactivation of the β-lactam ring of cephalosporin by β-lactamase (adapted from http://www.blad.co.in/resistance.php)

2.2.1.1 Types of ESBL

Classical ESBLs can be divided into three major groups namely the TEM-, SHV- and CTX-M-type ESBLs (Paterson & Bonomo, 2005; Pitout & Laupland, 2008).

2.2.1.1.1 TEM

TEM-1 was first reported in year 1965 from *E. coli* in Greece (Datta & Kontomichalou, 1965; Paterson & Bonomo, 2005). TEM-1 is able to hydrolyze ampicllin but with negligible activity against extended-spectrum cephalosporins and is inhibited by clavulanic acid. TEM-1, TEM-2 and TEM-13 have similar hydrolytic profiles and are all non-ESBLs (Jacoby & Medeiros, 1991; Paterson & Bonomo, 2005). In the early 1980s, plasmid-mediated ESBLs such as TEM-3 and TEM-12 that were

derived from the non-ESBL progenitors with amino acid substitutions were found to have enhanced activity against extended-spectrum cephalosporins. Today, the total subtypes of TEM have exceeded 130 types with most of them belonging to ESBLs (Jacoby & Price, 2005; Paterson & Bonomo, 2005). The bla_{TEM} gene in particular TEM-1 is often located on plasmids of *E. coli* and Enterobacteriaceae (Jacoby and Price, 2005). Among hundreds of subtypes of TEM, TEM-10, TEM-12 and TEM-26 are the most common subtypes reported in the North and South America (Paterson et al., 2003).

2.2.1.1.2 SHV

SHV-2, one of the most common SHV (sulfhydryl variable)-type ESBL, differed from the non-ESBL SHV-1 by a single amino acid difference (replacement of glycine by serine at position 238), and can hydrolyze cefotaxime effectively but not so on ceftazidime (Knothe et al., 1983; Paterson & Bonomo, 2005). In fact, most SHV-type ESBLs bear the same amino acid replacement at position 238 (Bradford, 2001). SHV-type ESBLs are widely found in various *Enterobacteriaceae* and outbreaks of SHV-producing bacteria (*Pseudomonas aeruginosa* and *Acinetobacter* spp.) have also been reported in Greece and China (Huang et al., 2004; Paterson & Bonomo, 2005; Poirel et al., 2004). More than 50 subtypes of SHV has been recognized currently, with SHV-5 and SHV-12 being the most common subtypes found in the clinical resistant bacterial isolates in both Europe and America (Paterson et al., 2003; Yuan et al., 1998).\

2.2.1.1.3 CTX-M

Cefotaximase (CTX-M)-producing bacteria can hydrolyze the third-generation cephalosporin, cefotaxime effectively (with MIC values in the resistant range, >64µg/ml) but weakly on ceftazidime (with MIC values in the susceptible range, 2-8 µg/ml). Nevertheless, certain CTX-M subtypes may have high potency in inactivating

ceftazidime, with reported MIC values as high as 256µg/ml (Baraniak et al., 2002; Poirel et al., 2002; Stürenburg et al., 2004). CTX-M can also hydrolyze cefepime efficiently (Tzouvelekis et al., 2000) but with variable efficiency on aztreonam (Paterson & Bonomo, 2005). CTX-M genes were phylogentically classified into 5 main groups based on amino acid sequence similarities, namely the CTX-M-1, CTX-M-2, CTX-M-8, CTX-M-9 and CTX-M-25 (Bonnet, 2004). Today, up to a total of 170 types of CTX-M have been identified (http://www.lahey.org/Studies/other.asp#table1, accessed on 31 May 2015).

Since the first detection of CTX-M in 1986 (Matsumoto et al., 1988), there has been a dramatic increase in the numbers of CTX-M-enzymes. CTX-M has also become the most prevalent ESBLs worldwide in the *Enterobacteriaceae* in the hospitals, community settings and among animal populations (Cantón & Coque, 2006; Eckert et al., 2006; Paterson & Bonomo, 2005). CTX-M-15 producing *E. coli*, which was first identified in India in 1999, is now the most prevalent ESBL variant found worldwide (Cantón et al., 2012; Pitout, 2009) except for Asia (Hawkey, 2008). A recent study by Sidjabat & Paterson (2015) revealed that different CTX-M subtypes have also appeared to predominate in different regions of Asia. According to a recent publication by the Study for Monitoring Antimicrobial Resistance Trends (SMART) (Jean and Hsueh, 2016), CTX-M-15 and CTX-M-14 remained as the predominant ESBL type among intra-abdominal infections-causing *Enterobacteriaceae*. Out of 12 Asia-Pacific countries which participated in the SMART programme (years 2008 – 2014), 6 countries (i.e., Australia, Kazakhstan, Malaysia, New Zealand, Philippines and Singapore) had the highest prevalence for CTX-M-15 among the 1739 studied *E. coli* strains. On the other hand, CTX-M-14 is the predominant ESBL subtype found in *E. coli* from Hong Kong, Korea and Taiwan (Jean and Hsueh, 2016).

CTX-M can be found associated with different genetic platforms including insertional sequences (e.g., IS*Ecp1*, IS*CR1*), integrons, phage-related sequences and conjugative plasmids (Bonnet, 2004; Paterson & Bonomo, 2005; Rossolini et al., 2008). The wide dissemination of CTX-M genes (in particular CTX-M-15 and CTX-M-14) has also been partly attributed to the clonal spread of the *E. coli* O25-ST131 clone (Bradford, 2001; Hawkey & Jones, 2009; Johnson et al., 2010).

2.2.1.1.4 Other ESBLs

There are several non-classical ESBLs that have been reported. OXA-type-β-lactamase has high efficiency in hydrolyzing oxacillin and is predominantly found in *Pseudomonas aeruginosa* (Weldhagen et al., 2003). Many other Gram-negative bacteria including *E. coli* are also common in harbouring OXA (Livermore, 1995). Majority of the OXA-type-β-lactamases are not able to hydrolyze extended-spectrum cephalosporins (ESC) efficiently and are hence not considered as ESBLs. Nonetheless, certain OXA types (e.g., OXA-11, -14, -16, -17, and -19) can hydrolyze cefotaxime and sometimes even ceftazidime and aztreonam (Paterson & Bonomo, 2005).

Other less common ESBLs include PER, BES-1, VEB-1, and CME. PER-1 has been found associated with a *Pseudomonas aeruginosa* outbreak in Italy (Luzzaro et al., 2001), which has been subsequently found in other countries such as France, Italy, Belgium (Claeys et al., 2000; De Champs et al., 2002; Pagani et al., 2004) and Korea (Kwon et al., 2002; Yong et al., 2003). VEB-1, which is plasmid-mediated, share high sequence identities with PER-1 and PER-2 (Poirel et al., 1999) and confers high potency on ceftazidime, cefotaxime, aztreonam and is inhibited by clavulanic acid. Other non-TEM and non-SHV ESBLs such as GES, BES, SFO, and IBC are also found distributed in various geographical regions (Paterson & Bonomo, 2005).

2.2.2 Dissemination of CTX-M genes

The main focus of the mechanism of resistance for ESBL in this study will be on CTX-M since CTX-M is the most common ESBL in *E. coli*. Both mobile genetic elements and vertical transmission play an important role in the dissemination of CTX-M genes (Peirano & Pitout, 2010).

2.2.2.1 The association of CTX-M genes with insertion sequences and phage elements

CTX-M genes have been reported to be associated with several types of mobile genetic elements, mainly the insertion sequences IS*Ecp1* and IS*CR1* and phage related elements, to a lesser extent (Poirel et al., 2012). IS*Ecp1* is an IS element of the IS*1380* family. It is able to recognize a wide range of structurally unrelated DNA sequences that are located adjacent to it as right inverted repeats (IRR) and then mobilize the gene (in this case CTX-M) via one-ended transposition mechanism. Hence, a single copy of IS*Ecp1* located upstream of a CTX-M gene is able to mobilize the CTX-M gene which will then lead to the 5 bp duplication at the target site of the transposed fragment. IS*Ecp1* has an outward-directing promoter sequence which is able to enhance the expression of an adjacent $bla_{\text{CTX-M}}$ gene, where the CTX-M expression is known to be low naturally (Lartigue et al., 2004; Poirel et al., 2012). IS*Ecp1*-like insertion sequence elements have been found to be associated with genes encoding for CTX-M-1, CTX-M-2 and CTX-M-9 (Cantón & Coque, 2006; Poirel et al., 2008).

The CR1 (common region 1, formerly *orf513*) element (Toleman et al., 2006), which is now known as IS*CR1* (an insertion sequence element) is often found to be associated with CTX-M-2 and CTX-M-9 genes. IS*CR1*, which belongs to the IS*91* family, is reportedly able to mobilize neighboring CTX-M genes via a rolling-circle replication

mechanism and is often found to be embedded in complex class I integrons. Integrons are mobile genetic elements that are able to acquire gene cassettes and express them as functional genes including the antibiotic resistance phenotypes (Cambray et al., 2010). Nevertheless, CTX-M genes are usually not present in the integron in the form of a gene cassette but is located downstream of IS*CR1* instead. IS*CR1* is also found to be capable of enhancing the expression of CTX-M-9 gene (Cantón & Coque, 2006; Poirel et al., 2012).

On the other hand, phage-related sequences have only been identified upstream of the CTX-M-10 gene, which was first reported in Spain (Oliver et al., 2005). In fact, similar phage-like elements have been identified upstream of CTX-M-10 genes of various CTX-M-10 producing strains. Nonetheless, the self-transferability of the phage-like structure was not determined (Oliver et al., 2005).

2.2.2.2 CTX-M genes and plasmids

Most $bla_{\text{CTX-M}}$ genes together with their respective immediate genetic environment are usually located in plasmids, facilitating the widespread distribution of $bla_{\text{CTX-M}}$. CTX-M genes have been found to be associated with both narrow host-range (IncFI, IncFII, IncHI2 and IncI) and broad host-range (IncN, IncP-1-a, IncL/M and IncA/C) plasmids (Carattoli, 2009). Among different CTX-M subtypes, CTX-M-9, CTX-M-14, CTX-M-15 and CTX-M-32 were often found associated with epidemic plasmids (Cantón & Coque, 2006). IncFII plasmids are among the most well-known plasmids that are responsible for the widespread distribution of CTX-M-15 genes. Majority of these CTX-M-15-bearing IncFII plasmids are also multireplicon in nature (Villa et al., 2010). On the other hand, the spread of CTX-M-9 in Spain was attributed to the dissemination of narrow host-range IncHI2 plasmids (Gilmour et al., 2004). In fact,

CTX-M-9 as well as CTX-M-32 have been reported to be linked to the epidemic spread of broad host-range IncP-1a and IncN plasmids, respectively (Novais et al., 2006).

2.2.2.3 Clonal dissemination of *E. coli* strains harbouring CTX-M genes

The widespread distribution of CTX-M genes (in particular CTX-M-15) is partly attributed to the single disseminated *E. coli* clone that belongs to sequence type (ST) 131, serotype O25b:H4 and B2 phylogenetic group, namely the ST131-O25b:H4 clone. The ST131-O25b:H4 clone also often harbour IncFII plasmids (Pitout, 2009), and is notorious for being a multidrug-resistance uropathogenic *E. coli* (Cantón et al., 2012; Johnson et al., 2010; Molina-lópez et al., 2010; Woodford et al., 2009). In fact, the *E. coli* O25-ST131 clone is also often co-resistant to fluoroquinolone (Johnson et al., 2010), as the IncFII plasmid often harbour both fluoroquinolone and extended-spectrum cephalosporin resistance genes (Canton and Coque, 2006). Besides CTX-M-15, ST131-O25b:H4 has been identified to harbour CTX-M-9 and CTX-M-14, but at a lower occurrence (Cao et al., 2011; Mora et al., 2010). Nonetheless, *E. coli* of other clones such as ST393 and ST69 (determined using MLST) were also often found associated with CTX-M genes (Riley, 2014).

2.3 Subtyping of *E. coli*

Understanding the genetic diversity of bacterial pathogens is essential for epidemiological investigation, microbial source tracking as well as the early detection of newly emerging clinically important pathogens. Hence, it is important to have a useful tool which can type bacterial isolates quickly and reliably. Various phenotypic and genotypic approaches have been described for *E. coli* subtyping with some commonly used tools being described in detail as below.

2.3.1 Serotyping

Conventional *E. coli* serotyping uses antisera which detects the somatic O- and H-antigens being expressed by *E. coli*, and hence able to distinguish *E. coli* of different serotypes (Orskov & Orskov, 1984). Although useful for short term epidemiologic studies (Ranjbar et al., 2014), serotyping is labour intensive, and time consuming for timely epidemiological investigations (van Belkum et al., 2007; Ranjbar et al., 2014). Nevertheless, molecular serotyping has been made available which can increase the efficiency to determine *E. coli* serotype (Coimbra et al., 2000). One of the ways to perform molecular serotyping for *E. coli* is by amplifying genes encoding flagellin (H-(flagellar) antigen) and O-antigen gene cluster (rfb) (Coimbra et al., 2000; Machado et al., 2000), followed by digesting the amplified region using the previously described primers and restriction enzymes. The serogroup and serotypes can then be determined by matching the restriction patterns to those in the previously described databases (Coimbra et al., 2000; Machado et al., 2000).

Certain *E. coli* serotypes are often associated with particular clinical outcomes. Of all, EHEC with serotype O157:H7 has drawn the limelight as a deadly diarrheagenic pathogen due to its frequent association with outbreaks, attributed to consumption of contaminated food and water (Kaper et al., 2004; Riley, 2014). Another well recognized clone, *E. coli* O104:H4, with combine features of EHEC and EAEC, also made another worldwide headline for causing approximately 3000 diarrheal cases with 17 deaths recorded via contaminated salad vegetables (EFSA, 2011). Besides IPEC, ExPEC also often associated with certain serotypes. *E. coli* with serotypes O4:K12:H5, O6:K2:H1, and O18:K1:H7 are serotypes that are epidemiologically associated with pyelonephritis, bacteremia, and/or neonatal meningitis (Russo & Johnson, 2000).

2.3.2 Multilocus sequence typing

Besides serotypes, *E. coli* clonal lineages are also often determined using multilocus sequence typing (MLST). MLST is a genotyping method with web accessible MLST databases for various bacterial species, including *E. coli* (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli). MLST often uses DNA sequences of selected seven housekeeping genes, which are believed to evolve slowly as they are under stabilizing, and not directional, selective pressure (Wirth et al., 2006). For *E. coli*, the 7 housekeeping genes are *adk* (adenylate kinase), *fumC* (fumerase isozyme C), *gyrB* (DNA gyrase subunit B), *icd* (isocitrate hydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate synthetase), and *recA* (recombinase A).

MLST is often utilized to determine the clonal dissemination of virulent and/or drug resistant *E. coli* clones. Several pandemic clonal lineages of ExPEC based on MLST have been reported to be disseminated globally, including ST131, ST393, ST69, ST95, and ST73 which often associated with urinary tract and bloodstream infections (Riley, 2014). On the other hand, ST1, ST11, ST25, ST270, ST272, ST280, ST582, ST731, ST1200 and ST1283 were found to be exclusively associated with intestinal pathogenic *E. coli* (Köhler & Dobrindt, 2011). Sequence type also often combined with *E. coli* serotype, to name several predominant ExPEC clonal groups including O25:H4-ST131 and O6:H1-ST73, which often linked to extraintestinal infections (Johnson et al., 2010; Riley, 2014).

2.3.3 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) is known for its remarkable discriminatory power and reproducibility that is able to differentiate among closely related strains, including *E. coli* strains that shared the same sequence types determined using MLST

(Brolund et al., 2013; Price et al., 2013; Riley, 2014). In brief, PFGE is a molecular typing method which uses an alternating current (pulsed fields) at a particular switch time which is predetermined to separate large chromosomal DNA fragments. PFGE requires high quality DNA, hence the DNA preparation involves embedding the bacteria in agarose, lysing it in situ, and then digesting the chromosomal DNA with restriction enzyme that cleaves infrequently. The slices of agarose plugs containing the DNA fragments are placed into the wells of an agarose gel, and the restriction fragments are then resolved into different banding patterns. The pulsotypes or pulsed field profiles of the strains are then compared with one another to determine their relatedness by generating a dendrogram (Tenover et al., 1995). Although PFGE is known to have high discriminatory power, it is a time consuming typing method (2-4 days) (van Belkum et al., 2007). Nevertheless, PFGE has been demonstrated to be able to determine whether an outbreak comes from a single source in the case of a suspected outbreak (van Belkum et al., 2007). PFGE has been used to investigate outbreak caused by contaminated food, as well as outbreaks occurred in the hospital and community. Mamlouk et al. (2006) reported an outbreak by drug resistant CTX-M-producing *E. coli* in a Tunisian hospital while Zhou et al. (2002) described a gastroenteritis outbreak in Japan caused by *E. coli* O166:H15. PFGE also successfully confirmed an outbreak of *E. coli* O157:H7 hemorrhagic colitis that is associated with unpasteurized gouda cheese in Canada (Honish et al., 2005). On the other hand, Lim et al. (2009) reported that 47 *E. coli* clinical strains yielded 44 pulsotypes, indicating substantial genetic diversity among the Malaysian *E. coli*.

Nonetheless, PFGE does suffer from problems such as DNA degradation, indicated by smearing in the PFGE lane instead of the appearance of distinct restriction bands. Such phenomenon can be caused by several reasons including premature cell lysis due to undue stress to bacterial cells prior to plug casting, incomplete lysis, insufficient

washing, and nucleolytic derivatives generated during electrophoresis (in the PFGE context, the peracetic acid), all of which have been described by PulseNet USA (http://www.aphl.org/conferences/proceedings/Documents/2009/2009_PulseNet/023Cooper.pdf). The degradation of DNA caused by the first three reasons can be easily overcome by gentle handling of bacterial suspensions, sufficient lysis and washing, respectively. As for the degradation caused by peracetic acid, which will give a characteristic smearing especially on the lower part of the gel, the degradation can be solved by adding free radicals scavengers such as thiourea to the running buffer and/or the agarose gel (Römling & Tümmler, 2000). The added thiourea can scavenge the Tris-peracetic acid generated during gel electrophoresis, and has been reported to improve the typeability of many bacterial isolates including *Pseudomonas aeruginosa*, *Clostridium difficile* and *Vibrio* species (Corkill et al., 2000; Römling & Tümmler, 2000; Takayanagi et al., 2005).

The DNA degradation phenotype (designated *dnd*) was first observed during an electrophoretic separation of DNA of a Gram-positive bacterium, *Streptomyces lividans* (Zhou et al., 1988). A DNA modification system (conferred by a five-gene *dnd* cluster, *dndABCDE*) was then found to be responsible for the *dnd* phenotype by mediating the incorporation of sulphur into the DNA backbone via a process called phosphorothioation (Zhou et al., 2005, 1988). The DNA modification renders the DNA susceptible to cleavage by peracetic acid generated during a PFGE run (Ray et al., 1995; Xie et al., 2012). While the 5-gene cluster (consisting of *dndA* and the *dndBCDE* operon) is found in bacteria across different families (such as *Acidobacteriaceae*, *Clostridiaceae*, and *Streptomycetaceae*), *Enterobacteriaceae* including *E. coli* only harbour a 4-gene *dndBCDE* operon without the *dndA* gene. IscS, a cysteine desulfurase that has a similar function as the *dndA*-encoded protein, was then identified in *E. coli t*o support DNA phosphorothioation (An et al., 2012; Ou et al., 2009).

2.3.4 Phylogenetic grouping

Escherichia coli is composed of four main phylogenetic groups (A, B1, B2, and D). Three candidate markers are used to assigned *E. coli* into one of the phylogroups (A, B1, B2 or D) based on the presence and absence of the three genes (Figure 2.4) (Clermont et al., 2000), which can be carried using PCR. The three genes are: a) *chuA*, a gene required for heme transport in *E. coli*; b) *yjaA*, a gene with unknown function, and c) TSPE4.C2, a putative lipase esterase gene (Clermont et al., 2000; Clermont et al., 2013).

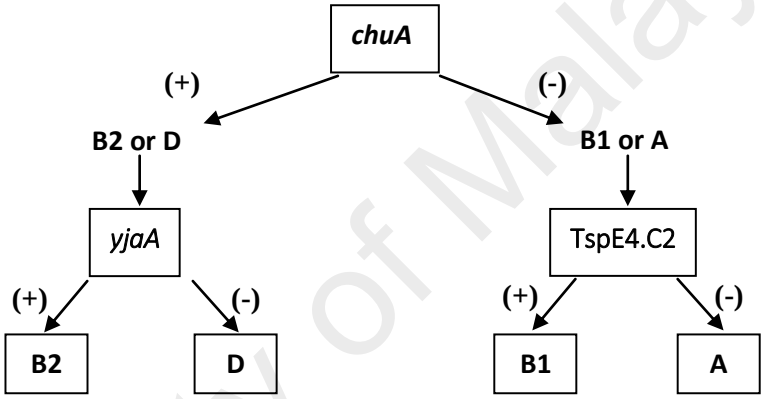


Figure 2.3 Phylogenetic grouping of *E. coli* based on the presence and absence of three genetic markers (Clermont et al., 2000).

Virulent extra-intestinal strains mainly belong to phylogroup B2 and to a lesser extent, phylogroup D while most commensal strains belong to phylogroup A (Clermont et al., 2000; Clermont et al., 2013). In fact, ExPEC are often reported to be derived predominantly from *E. coli* phylogroup B2, followed by phylogroup D since more than a decade ago (Jauréguy et al., 2007; Johnson et al., 1991; Picard et al., 1999; Salipante et al., 2015).

2.3.5 Phenotypic microarray assay

Phenotypic microarray (PM) assay is a high throughput technology that enables the simultaneous evaluation of nearly 2000 phenotypes of a bacterial strain in a single experiment. PM results are useful for correlating genotypes and phenotypes, characterizing cell phenotypes for epidemiological studies, determining gene functions and many more. Generally, each well of the PM plate is designed to test for different phenotypes. When cell respires in the well with substrate that is being inoculated with cell suspension, the tetrazolium dye in the well will be reduced forming a purple-coloured compound (Bochner et al., 2001). The intensity of the colours formed is dependent on how actively the cell respires and the PM results are available in the form of kinetic graphs (Bochner et al., 2001).

PM has been used for various *E. coli* research purposes. Several studies have employed PM for gene knockout or mutation studies for *E. coli* (Bochner et al., 2001; Roesch, 2003; Zhou et al., 2003). Using PM, Alqasim et al. (2014) showed that the pandemic virulent *E. coli* ST131 clone is not metabolically distinct from other ExPEC lineages. The same assay also revealed that by comparing the metabolic profiles of two closely related commensals *E. coli* strain B and K-12, strain B served as a better candidate for production of recombinant proteins due to advantages in certain metabolic reactions (Yoon et al., 2012). PM assay also often couples with whole genome sequence analysis for detail *E. coli* investigations (Blum et al., 2015; Yoon et al., 2012).

2.3.6 *E. coli* and whole genome sequencing (WGS)

E. coli is a highly versatile bacterium, which can exists as commensals, IPEC and ExPEC in humans, animals and environment (Kaper et al., 2004). In general, genome sizes of *E. coli* are approximately 4.5 - 5.9 million bp, with estimated number of ORFs ranging from 4687 to 5526, based on a study which includes 61 *E. coli* genomes from

various pathotypes and sources (Lukjancenko et al., 2010). As the number of *E. coli* genomes increase, the core genome size will decrease, while the pangenome size will increase (Lukjancenko et al., 2010). Nonetheless, core and pangenome sizes vary depending on the study population as *E. coli* from the same pathotypes may share high number of core genes and vice versa (Salipante et al., 2015).

Generally, *E. coli* with the same pathotype, phylogenetic group and/or sequence type are well clustered together in a phylogenomic tree as they shared similar phylogenetic signal (Mcnally et al., 2013; Salipante et al., 2015). With the advent of next generation sequencing, high resolution phylogenomic analysis shed further insights into the evolutionary relationship between or within the same *E. coli* pathotypes. Using phylogenomic analysis, both Salipante et al. (2015) and Hazen et al. (2013) further divided ExPEC and attaching and effacing *E. coli* (AEEC) into different lineages, in which each of the lineages harbour distinct virulence and/or resistance profiles, suggesting the utility of phylogenomic analysis in further distinguishing *E. coli* of the same pathotypes.

Although *E. coli* of the same pathotypes are generally grouped together into a cluster of a phylogenomic tree, the boundaries between different pathotypes have been blurred due to recombination (Mcnally et al., 2013). Recently, Chattaway et al. (2014) described a diarrheagenic *E. coli*, EAEC, in causing extraintestinal infections. Although *E. coli* of ST38 is a successful EAEC clone, it carries several extraintestinal virulence determinants and were often associated with urinary tract infections, indicating the emergence of UPEC/ EAEC hybrid pathotype (Chattaway et al., 2014). Similarly, the hybrid EAEC/EHEC O104:H4 is also notorious for the large hemolytic uremic syndrome outbreak in German, which was caused by contaminated sprouts (EFSA, 2011).

Besides being useful for inferring phylogetic relationships and epidemiological and outbreak investigations, whole genome sequences can be used for many other purposes. WGS is reported by several studies to be feasible for the 4-step routine diagnostics for clinical implementation, namely detection, identification, antibiotic susceptibility profiling and epidemiological typing (Bertelli & Greub, 2013; Fournier et al., 2014; Hasman et al., 2014; Köser et al., 2012). Whole genome sequences of *E. coli* have also been found useful for identification genetic markers (Clermont et al., 2013), and fitness traits such as toxin-antitoxin systems (Norton & Mulvey, 2012; Shao et al., 2011), virulence determinants (Chen et al., 2012) and many more, leading to a step-change in public health microbiology. By comparing the genomes of pathogenic and non-pathogenic *E. coli*, Moriel et al. (2010) identified several vaccine candidates for ExPEC, which may also be effective against IPEC as some antigens are also present in IPEC. In fact, many *E. coli* typing methods that were used to be tedious and time consuming can be carried out using the *in silico* approach, by submitting the whole genome sequences to the relevant webservers. These typing tools includes MLST 1.8 (Larsen et al., 2012) and SerotypeFinder 1.1 (Joensen et al., 2015) which are used to determine the sequence types and serotypes of *E. coli*. Other genomic features such as genomic islands and clustered regularly interspaced short palindromic repeat (CRISPR) elements that are of research interests can also be determined (Bellanger et al., 2014; Dhillon et al., 2015; Grissa et al., 2007; Lange et al., 2013; Soares et al., 2012).

CHAPTER 3: CHARACTERIZATION OF ESBL-PRODUCING *E. COLI* FROM PEDIATRIC WARDS OF A MALAYSIAN HOSPITAL

3.1 INTRODUCTION

Extended-spectrum cephalosporins (ESCs) resistance is recognized as a major drug resistant threat globally (CDC, 2013) as ESC is often used as first-line antimicrobial agents for treating infections caused by Gram-negative bacteria (MOH, 2008; MOH, 2014b; Winokur et al., 2001). As reviewed in Chapter 2, various beta-lactamase genes can be associated with ESC resistance with ESBL and AmpC enzymes (Philippon et al., 2002) representing a major challenge to the public health worldwide. This is because ESBL producing bacteria are often multidrug resistant while AmpC-producing bacteria frequently appear as susceptible to certain ESC *in-vitro* but therapeutic failures have been reported with ESC treatment (Jacoby, 2009). This phenomenon has resulted in increased morbidity and mortality due to treatment failures.

ESBL-encoding genes located on integron-like structures have been reported worldwide at an increasing rate (Bonnet, 2004; Cantón & Coque, 2006). Besides, the pandemic *E. coli* O25-ST131 clone which is notorious for its successful global dissemination of CTX-M-15, has not been reported in Malaysia. There have been very few reports on the prevalence of beta-lactamase genes in *E. coli* isolated from hospitals in Malaysia (Lim et al., 2009; Sekawi et al., 2008; Yusof, 2011) and the Southeast Asian region. Any association between ESBL genes and integrons has not been investigated among Malaysian *E. coli* strains. Although *E. coli* infections affect patients of all ages, certain types of infections are more common in children under 5 years old (CDC, 2011). Thus, this study was undertaken to (1) determine the prevalence of different types of beta-lactamase genes and their respective variants, (2) determine the level of resistance to ESCs conferred by different variants of ESBL and AmpC genes,

(3) investigate the association between integrons and ESBL genes, and (4) investigate the clonality of ESBL- and AmpC-producing *E. coli* strains isolated from children in a Malaysian tertiary hospital.

3.2 METHODOLOGY

3.2.1 Bacterial strains

One-hundred and ten non-repeat *E. coli* strains from stool samples of children under the age of 5 years were obtained in 2009–2010, from a tertiary hospital in Johor Bahru, a city at the southern tip of Peninsular Malaysia. The background information for the *E. coli* strains are detailed in Appendix II a. These strains were previously identified by the clinical microbiology laboratory in the hospital using standard cultural methods and biochemical tests.

3.2.2 PCR confirmation of *E. coli* strains

PCR confirmation targeting *phoA*, which is an *E. coli* housekeeping gene (Kong et al., 1999) was carried out (Appendix III a) using DNA templates prepared using the Wizard Genomic DNA Purification Kit (Promega) (Appendix IV a). Strains that have been validated as *E. coli* were then stored in cryovials of veal infusion broth with 50% glycerol (Invitrogen) at -20°C and -85°C.

3.2.3 Antimicrobial susceptibility testing

Antimicrobial susceptibility of *E. coli* strains was determined using the disk diffusion method according to Clinical and Laboratory Standards Institute (CLSI) guidelines (CLSI, 2015) (Appendix IV b). The antimicrobial agents used were: ampicillin (AMP, 10 μg), tetracycline (TET, 30 μg), trimethoprim/sulfamethoxazole (SXT, 30 μg), nalidixic acid (NAL, 30 μg), streptomycin (STR, 10 μg), cefoperazone

(CFP, 30 μg), kanamycin (KAN, 30 μg), chloramphenicol (CHL, 30 μg), ceftriaxone (CRO, 30 μg), amoxicillin/clavulanic acid (AMC, 20/10 μg), cefotaxime (CTX, 30 μg), ciprofloxacin (CIP, 5 μg), aztreonam (ATM, 30 μg), cefepime (FEP, 30 μg), gentamicin (GEN, 10 μg), ceftazidime (CAZ, 30 μg), amikacin (AMK, 30 μg), meropenem (MEM, 10 μg), and imipenem (IPM, 10 μg). Minimum inhibitory concentration (MICs) was also performed using E-test strips (AB Biodisk) with third generation cephalosporins, namely ceftazidime (CAZ), cefotaxime (CTX), ceftriaxone (CRO), and aztreonam (ATM). *E. coli* ATCC 25922 and ATCC 35218 were used as control strains (CLSI, 2015).

3.2.4 Phenotypic detection of ESBL production

E. coli strains were screened for presumptive ESBL production using four antimicrobial disks containing CAZ (30 mg), CRO (30 mg), CTX (30 mg), and ATM (30 mg) (Appendix IV c i). Presumptive-positive ESBL-producing *E. coli* strains were further tested for the ESBL production using the modified double-disk synergy test (DDST) (Jarlier et al., 1988) (Appendix IV c ii) and E-test ESBL strips (CTX/CTX + clavulanic acid [CT/CTL] and CAZ/CAZ + clavulanic acid [TZ/TZL]) (AB Biodisk) (Appendix IV c iii). These two methods were used together to increase the efficiency of ESBL detection. Strains with positive results for DDST or ESBL E-test were termed ESBL producers. *E. coli* ATCC 25922 and ATCC 35218 were used as control strains (CLSI, 2015).

3.2.5 Genotypic detection of genes encoding for beta-lactamases

Polymerase chain reaction (PCR) was used to detect the presence of seven different types of beta-lactamase genes using primers previously described for the amplification of $bla_{\text{CTX-M}}$ (Pagani et al., 2003); bla_{TEM}, $bla_{\text{OXA-1}}$, and bla_{SHV} (Oliver et al., 2002);

bla$_{VEB}$ (Naas et al., 2001); *bla*$_{DHA}$ (Pai et al., 2004); and *bla*$_{CMY-2}$ (Archambault et al., 2006). CTX-M genes were known to consist of 5 major phylogenetic groups where different CTX-M subgroups predominate in different geographical regions (Bonnet, 2004). Hence, strains that yielded positive amplification for *bla*$_{CTX-M}$ were further subgrouped into CTX-M groups 1, 2, 8/25, and 9 as described by Ensor et al. (2007). PCR detection for the CTX-M-15 gene was carried out using primers described by Conceição et al. (2005). All amplified CTX-M genes (except for CTX-M-15) were purified and submitted to a commercial facility (First BASE Laboratories) for sequencing to determine the specific CTX-M subtype. Representative amplified *bla*$_{CTX-M}$, *bla*$_{TEM-1}$, *bla*$_{OXA-1}$, and *bla*$_{SHV}$ genes with the expected amplicon sizes were also sequenced and validated by sequence alignment. These validated genes were then used as positive controls for subsequent ESBL gene detection using PCR. All primers, expected amplicon sizes and annealing temperatures for the PCR detection of genes encoding for the various beta-lactamases were listed in Appendix III b.

3.2.6 Detection of O25-ST131 clone

E. coli O25-ST131 clone is an international pandemic clone notorious for the dissemination of CTX-M genes, in particular CTX-M-15 (Johnson et al., 2010). Hence, detection of the O25-ST131 *E. coli* clone was carried out by PCR as previously described (Clermont et al., 2009) for CTX-M-15-producing strains. In brief, this rapid PCR targets the *pabB* gene which is specific to the *E. coli* O25-ST131 clone, with inclusion of a positive control named EC1003-1 that has been validated by DNA sequencing and alignment analysis (Appendix III c).

3.2.7 PCR detection of class 1, 2, and 3 integrons

All 110 *E. coli* strains were tested for the presence of class 1, class 2, and class 3 integron-encoded integrases using PCR as described previously (Machado et al., 2005). Amplification of the gene cassette regions was carried out for integrase-positive strains (using the 5'CS/3'CS primers for gene cassette-bearing class 1 integrons and the attI2-F/orfX-R primers for gene cassettes in class 2 integrons) (Machado et al., 2005). All PCR amplicons were purified and submitted to a commercial company for sequencing (First BASE Laboratories). Primers used for detection of integrases and variable gene cassette regions were detailed in Appendix III d.

3.2.8 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) for all 110 *E. coli* strains was performed according to Thong et al. (2007) with minor modifications. The detailed protocols for PFGE and data analysis are found in Appendix IV d i-vi. In brief, the *E. coli* genomic DNA plugs were prepared and then subjected to *Xba*I digestion. PFGE was performed for 24 hours in 0.5X TBE buffer at 14°C in a CHEF Mapper system (Bio-Rad) using pulse times of 2.16 to 54.17 sec.

3.3 RESULTS

3.3.1 Antimicrobial resistance rates among the *E. coli* strains

Eighteen of the 110 (16.4%) *E. coli* strains were susceptible to all the 19 antimicrobials tested and a representative result of antimicrobial susceptibility testing on a Mueller Hinton II plate was included in Figure 3.1. The antimicrobial resistance rates are summarized in the form of graph (Figure 3.2). All the strains were found to be susceptible to carbapenems (i.e. imipenem and meropenem).

Fourty six percent (51/110) of the strains were multidrug resistant (MDR, bacterial strains that are resistant to three or more classes of antimicrobials) (CDC, 2013). Table 3.1 summarizes the resistance profiles of the 51 MDR. Among the MDR strains, high resistance rates were observed for AMP (98%, n = 50/51), TET (75%, n = 38/51), and SXT (73%, n = 37/51) (Appendix V a). Three MDR strains had the highest number of antimicrobial resistance (n = 14) and were resistant to all broad-spectrum cephalopsorins and monobactam tested. Fourteen ESC-resistant and nine ESC-intermediate-resistant *E. coli* strains were also MDR.

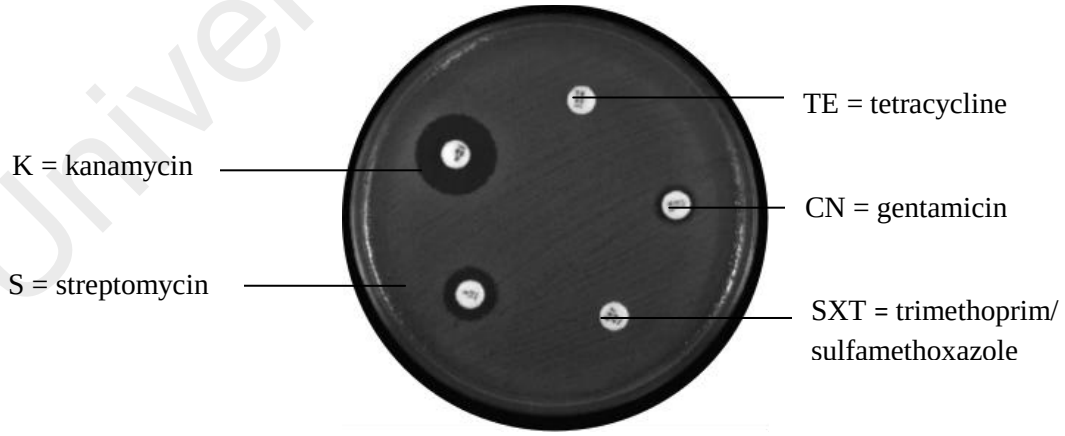


Figure 3.1: Representative picture of antimicrobial susceptibility testing for an *E. coli* isolate on Mueller Hinton II agar plate using Kirby-Bauer disc diffusion method

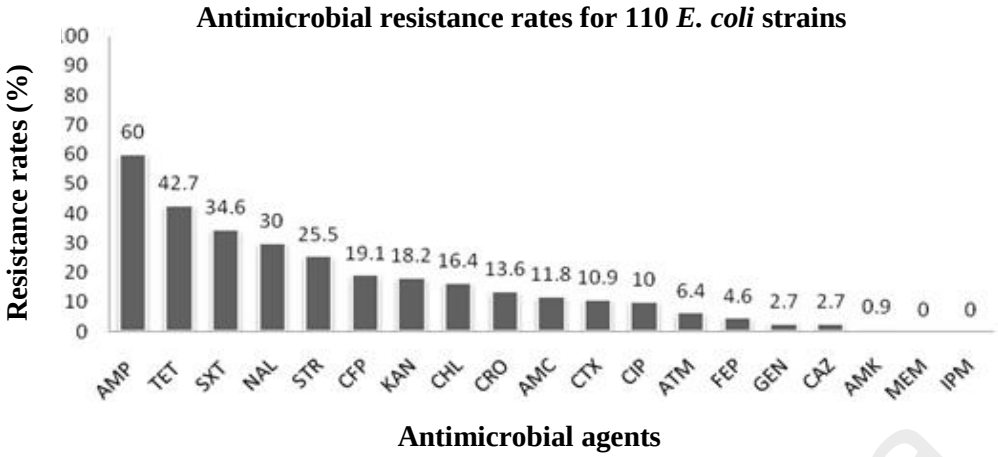


Figure 3.2: Antimicrobial resistance rates for the 110 *E. coli* strains isolated from paedriatic wards of a Malaysian tertiary hospital.

Thirty percent (33/110) of the *E. coli* strains were presumptive ESBL producers based on initial four-antimicrobial disk-diffusion screening. Eighteen were ESBL producers based on DDST (Figure 3.3). Based on the phenotypic confirmatory test (ESBL E-test), 15 strains were ESBL producers while 2 were non-determinate as their MIC values for CTX/CTX + clavulanic acid and CAZ/CAZ + clavulanic acid were above the test range. According to the manufacturer's interpretation, Nd results might suggest the presence of an inhibitor-resistant (such as AmpC enzyme). Thirteen strains were confirmed as ESBL producers by both phenotypic detection tests (DDST and ESBL E-test methods).

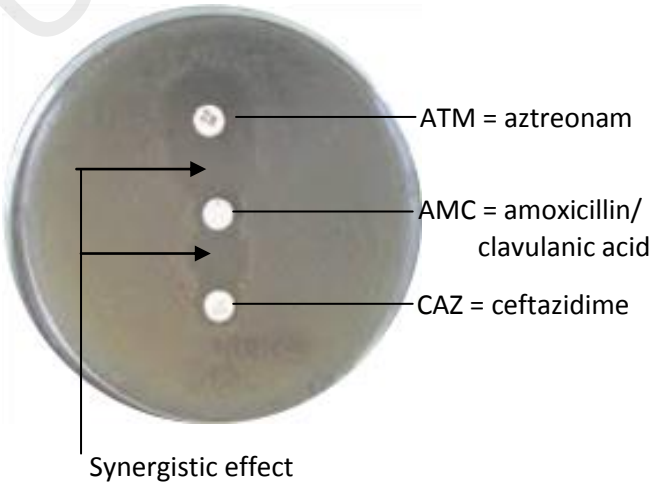


Figure 3.3: Representative picture of double disk synergy testing (DDST) for an *E. coli* isolate on Mueller Hinton II agar plate.

Table 3.1: Antimicrobial resistant phenotypes of 51 MDR *E. coli* strains

No. of antimicrobial resistance	Resistance phenotypes	No. of strains
14	AMP, TET, SXT, NAL, STR, CFP, KAN, CHL, CRO, CTX, CIP, ATM, FEP, CAZ	3
10	AMP, TET, SXT, NAL, STR, CFP, KAN, CHL, AMC, CIP	1
	AMP, TET, SXT, STR, CFP, KAN, CRO, CTX, AMC, ATM	1
8	AMP, TET, SXT ,NAL, STR, KAN, CHL, CIP	1
	AMP, TET, SXT, NAL, STR, CHL, CIP, AMK	1
	AMP, TET, SXT, NAL, CFP, KAN, AMC, CIP	1
	AMP, TET, SXT, NAL, CFP, CHL, CRO, CTX	1
	AMP, TET, SXT, STR, CFP, CHL, CRO, CTX	2
7	AMP, TET, SXT, NAL, STR, KAN, CHL	1
	AMP, TET, SXT, NAL, STR, CHL, CIP	2
	AMP, TET, SXT, NAL, STR, AMC, CIP	1
	AMP, TET, SXT, STR, CFP, KAN, AMC	1
	AMP, STR, CFP, KAN, CRO, CTX, FEP	1
6	AMP, TET, SXT, NAL, STR, CIP	1
	AMP, TET, SXT, NAL, STR, GEN	2
	AMP, CFP, CRO, CTX, ATM, FEP	1
5	AMP, TET, SXT, NAL, STR	1
	AMP, TET, SXT, NAL, KAN	4
	AMP, TET, SXT, KAN, C	3
	AMP, TET, SXT, STR, KAN	1
	AMP, TET, SXT, STR, C	1
	TET, SXT, STR, KAN, CHL	1
	AMP, TET, NAL, CFP, CRO	1
	AMP, CFP, CRO, CTX, ATM	1
	AMP, SXT, NAL, CFP, CRO	1
	AMP, CRO, CTX, AMC, ATM	1
4	AMP, TET, SXT, STR	1
	AMP, TET, NAL, CFP	2
	AMP, TET, STR, AMC	1
	AMP, SXT, NAL, STR	1
	AMP, SXT, STR, CHL	1
3	AMP, TET, SXT	1
	AMP, TET, AMC	1
	AMP, TET, KAN	1
	AMP, SXT, STR	1
	AMP, SXT, NAL	1
	AMP, CFP, AMC	3
	AMP, CRO, AMC	1

3.3.2 Prevalence of different beta-lactamase genes and O25-ST131 clone

Out of the 7 beta-lactamase genes tested (TEM, CTX-M, OXA-1, SHV, VEB, DHA, and CMY-2) using PCR, only 3 were detected among the 110 *E. coli* strains: bla_{TEM} (49.1%, n = 54), $bla_{\text{CTX-M}}$ (11.8%, n = 13), and $bla_{\text{CMY-2}}$ (6.4%, n = 7) (Figures 3.4 a, b, and c). Analysis of the DNA sequence of the bla_{TEM} amplicons revealed that they belonged to subgroup $bla_{\text{TEM-1}}$ and this was the most prevalent beta-lactamase gene detected.

All CTX-M-positive strains were ESBL producers with positive results for both DDST and ESBL E-test methods and the 13 CTX-M genes were further subgrouped into CTX- M-15 (n = 10) and CTX-M-9 (n = 3). Further DNA sequence analyses of the CTX-M-9 genes revealed two subtypes, CTX-M-14 (n = 2) and CTX-M-27 (n = 1). No CTX-M subgroup 2 or 8/25 was detected. All strains with CMY-2 or CTX-M genes were nonsusceptible to third-generation cephalosporins. The O25-ST131 clone was also not found among the 10 CTX-M-15- positive *E. coli* strains.

CTX-M enzymes had higher hydrolytic activity against CTX (MIC, 24–256 µg/ml) than CAZ (MIC, 1–16 µg/ml). Majority of the CTX-M-positive *E. coli* strains (11/13) showed high resistance levels to CTX with MIC values of ≥ 32 µg/ml. The MIC values for CTX-M-15- and CTX-M-14- producing *E. coli* for the four antimicrobial agents were as follows: CTX (16–256 vs. 24–48 µg/ml), CAZ (2–16 vs. 1– 2 µg/ml), CRO (16–256 vs. 48–64 µg/ml), and ATM (3–96 vs. 3–8 µg/ml) (Figure 3.5).

MIC values of CTX, CAZ, CRO, and ATM for CMY-2- producing strains were 1.5–3, 3–12, 2–8, and 3–12 µg/ml, respectively.

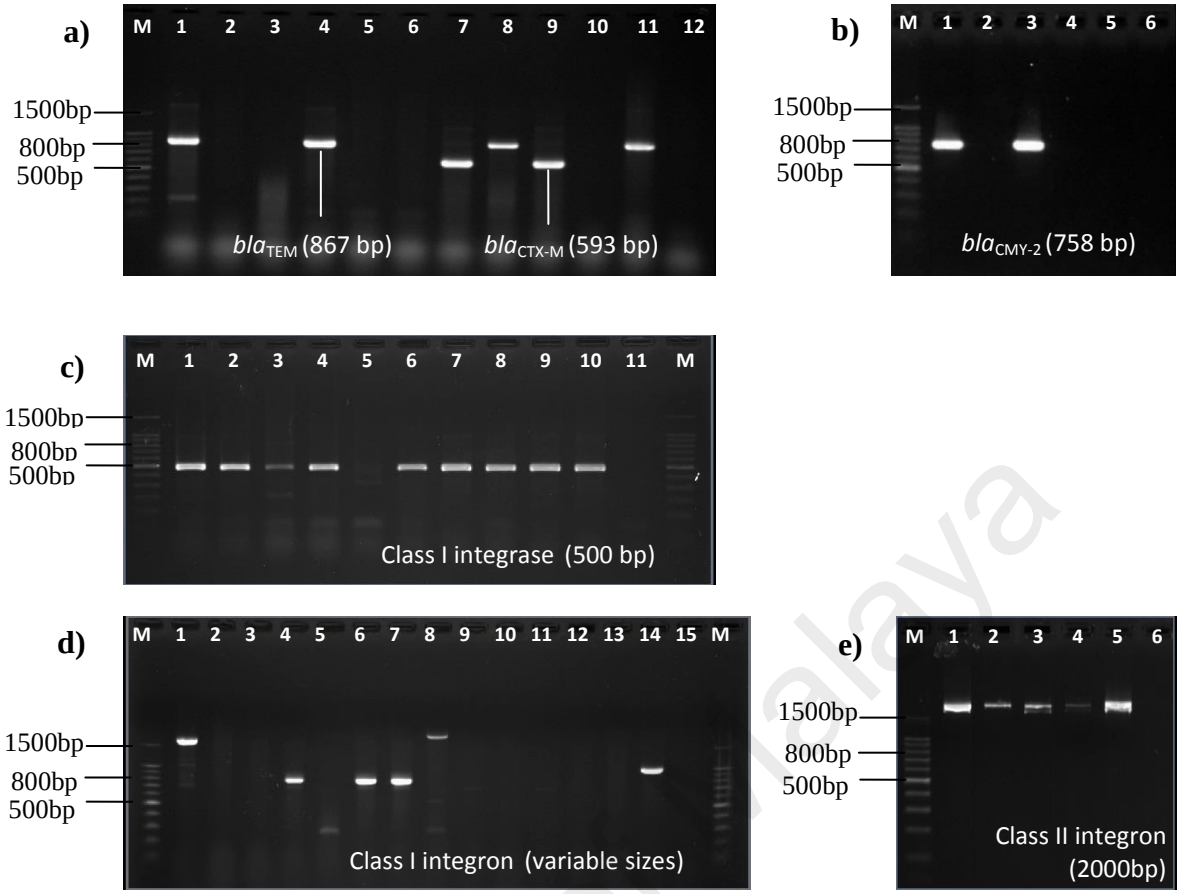


Figure 3.4: Representative agarose gel pictures of PCR amplified products of**: a)** ***bla*$_{\mathbf{TEM}}$ and *bla*$_{\mathbf{CTX\text{-}M}}$**. Lanes 2-11: EC002/09, EC003/09, EC004/09, EC005/09, EC006/09, EC007/09, EC007/09, EC009/09, EC010/09, EC011/09; lane 12: negative control; **b) *bla*$_{\mathbf{CMY\text{-}2}}$.** Lanes 2-5: EC002/09, EC003/09, EC004/09, EC005/09; lane 6: negative control; **c) class 1 integron-encoded integrase**. Lanes 2-10: EC072/10, EC073/10, EC074/10, EC075/10, EC076/10, EC077/10, EC078/10, EC079/10, EC080/10; lane 11: negative control **d) class I integron.** Lanes 2-14: EC088/10, EC087/10, EC086/10, EC083/10, EC081/10, EC080/10, EC079/10, EC078/10, EC077/10, EC076/10, EC074/10, EC073/10, EC072/10; lane 15: negative control **e) class II integron**. Lanes 2-5: EC038/10, EC072/10, EC082/10, EC083/10; lane 6: negative control. Lane M: 100 bp DNA ladder (Promega); Lane 1: positive control.

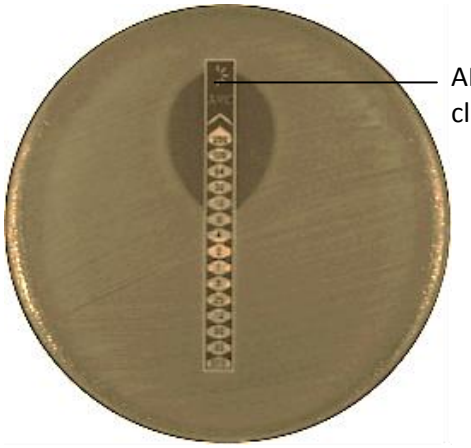


Figure 3.5: Representative picture of minimum inhibitory concentration (MIC) test for an *E. coli* isolate using E-test strip on Mueller Hinton II agar plate.

3.3.3 Prevalence of class 1, 2, and 3 integrons

All 110 *E. coli* strains were screened for class 1, 2, and 3 integron-encoded integrase genes. No class 3 integron-encoded integrase gene was detected. Among the 110 strains, 48 (43.6%) and 4 strains (3.6%) harbored the *intI1* and *intI2* integrase genes, respectively. Of the 48 *intI1*-positive strains, only 21 yielded amplified products with the 5'_CS/3'_CS primers, implying the presence of gene cassettes, while all 4 *intI2* integrase-positive strains harbored gene cassettes as part of class 2 integrons. Six types of gene cassettes were identified (Table 3.2). Three types of *aadA* genes (*aadA1*, *aadA2*, and *aadA5*) were identified, which conferred resistance to STR and spectinomycin. The *aadB* gene (n = 2), which confers resistance to a range of aminoglycosides, namely, GEN, KAN, and tobramycin, was found to be associated with *aadA2* in the class 1 integrons. Four types of *dfr* genes (*dfrA12*, *dfrA17*, *dfrA5*, and *dfrA7*) were detected in class 1 integrons and one type (*dfrA1*) in a class 2 integron. The *dfr* genes encode for the enzyme dihydrofolate reductase that mediates resistance to trimethoprim. The *dfrA* and *aadA* genes can exist alone or in combination with other resistance genes. All *E. coli* strains that harbored *dfr* and *aad* genes were nonsusceptible to SXT and aminoglycosides, respectively. The genotypes of the gene cassettes found in the integrons corresponded to the resistotypes. Only the *sat1* gene was identified in class 2 integron strains, which confers resistance to streptothricin. No beta-lactamase genes

were found within the gene cassettes that were amplified from the integrase-positive *E. coli* strains.

Table 3.2: Contents of class 1 and class 2 integrons gene cassettes detected in 23 integron positive *E. coli* strains

Integrons	Gene cassettes	Amplicon size (bp)	No. of strains
Class 1	*dfrA12- orfF -aadA2*	1900	8
	*dfrA*17-*aadA*5	1600	4
	*aadB- aadA*2	1600	2
	aadA1	1000	1
	dfrA7	800	3
	dfrA5	700	1
Class 2	*dfrA1-sat1-aadA*	2000	2
Class 1 + class 2	*dfrA12- orfF -aadA2* + *dfrA1-sat1-aadA*	1900 2000	2
No. of strains with integrons			23
No. of strains without integrons			87

3.3.4 Genetic relatedness of the *E. coli* strains based on PFGE

PFGE subtyped *Xba*I-digested genomic DNA of 107 strains into 72 PFPs with a Simpson's index diversity value of 0.98, and *F*-value (coefficient of similarity) ranging from 0.45 to 1.0 (Figure 3.6), indicating that the local *E. coli* strains that were obtained from the same hospital are quite diverse. *XbaI* digestion of the genomic DNA resulted in a total of 14 – 26 DNA fragments with sizes ranging from 20 to 800 kb. Three strains

(EC062/10, EC063/10, and EC075/10) were untypeable although their DNA preparation was repeated three times. At 85% similarity, a total of 78 strains were clustered into 29 clonal groups, each comprising of 2 – 4 strains. Most strains within the same clonal group shared high degree of similarities with closely related PFPs and similar genotypes. Another 29 strains were less defined and were not clustered into any apparent clonal group.

The 13 CTX-M-positive strains were subtyped into 11 PFPs. Three out of four strains (EC009/09, EC013/09, and EC019/09) that had the highest detectable MICs for CTX and CRO (256 mg/ml) were grouped together into a clonal group. These three strains also displayed relatively high resistance to ATM (MIC, 32–96 mg/ml) and CAZ (8–16 mg/ml). All three strains were highly related with one band difference and carried both CTX-M-15 and class 1 integron (*dfrA12-orfF-aadA2*). Seven CMY-2-positive strains from different pediatric wards had no apparent linkage as they were found scattered in the dendrogram.

The PFGE profiles of *E. coli* in the present study (isolated between 2009 - 2010) were also compared with profiles of previously isolated *E. coli* (isolated in 2004) in the same hospital (Lim et al., 2009). The genetic relatedness of the *E. coli* strains obtained from the same hospital over a gap of more than 5 years was established from a dendrogram based on their PFGE profiles (Figure 3.7). Overall, these *E. coli* strains were genetically diverse except for two strains, EC096/10 (isolated from a stool sample) and EC302/04 (isolated from tracheal aspirate) that differed by only a single band (PFGE analysis for these two strains was repeated three times) (Figure 3.8).

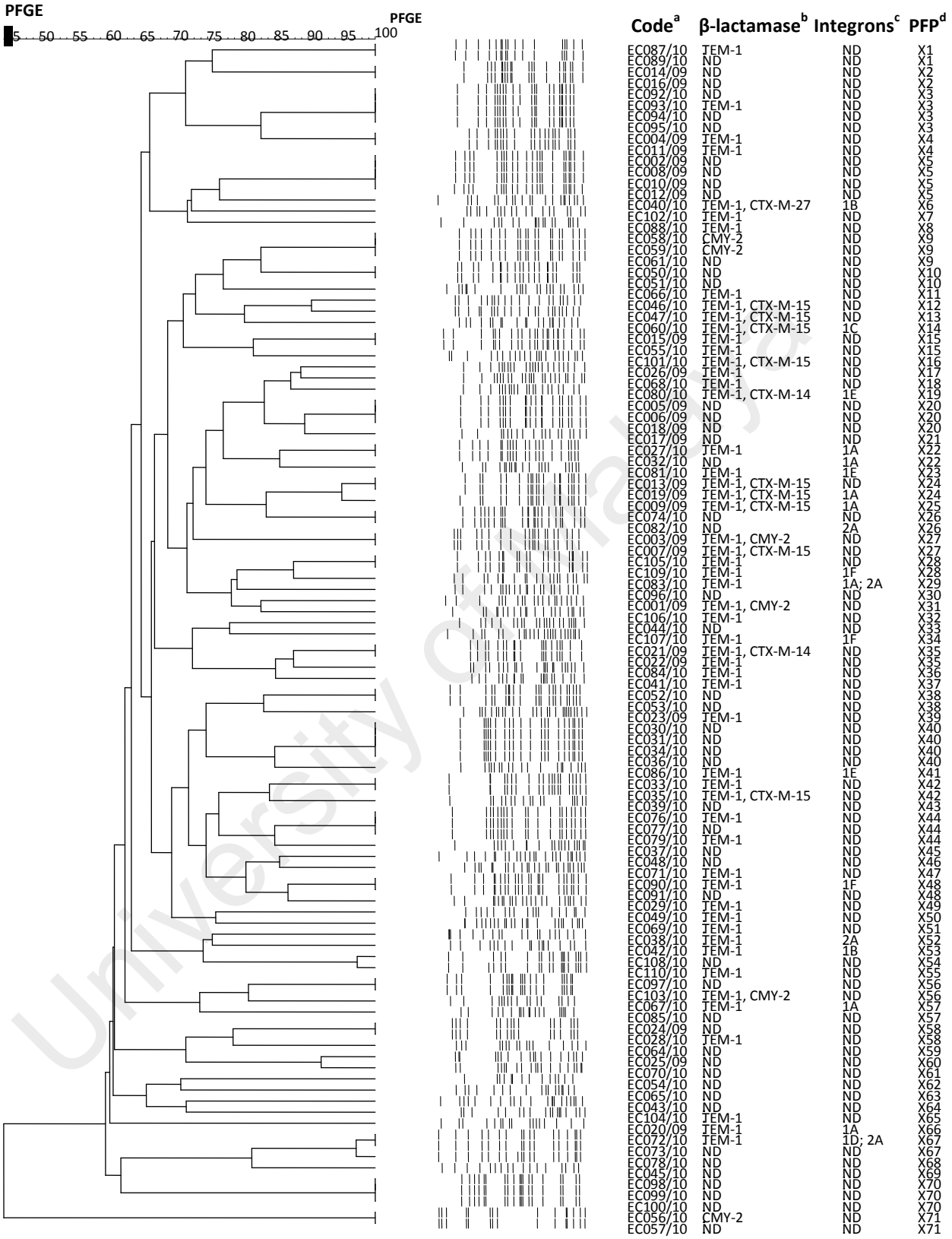


Figure 3.6: Dendrogram generated from PFGE profiles of 107 *E. coli* strains isolated in year 2009 and 2010 from children**.** [a]Strain code; [b]presence of β-lactamase genes; [c]gene cassettes found in class 1 integrons: (1A) *dfrA12-orfF-aadA2*; (1B) *aadB-aadA2*; (1C) *dfrA5*; (1D) *aadA1*; (1E) *dfrA7*; (1F) *dfrA17-aadA5*; gene cassette found in class 2 integron: (2A) *dfrA1-sat-aadA1*; [d]pulsed-field profiles. The scale on top of the dendrogram indicates the percentage of similarity. ND, not detected; UPGMA, unweighted pair group method using arithmetic averages.

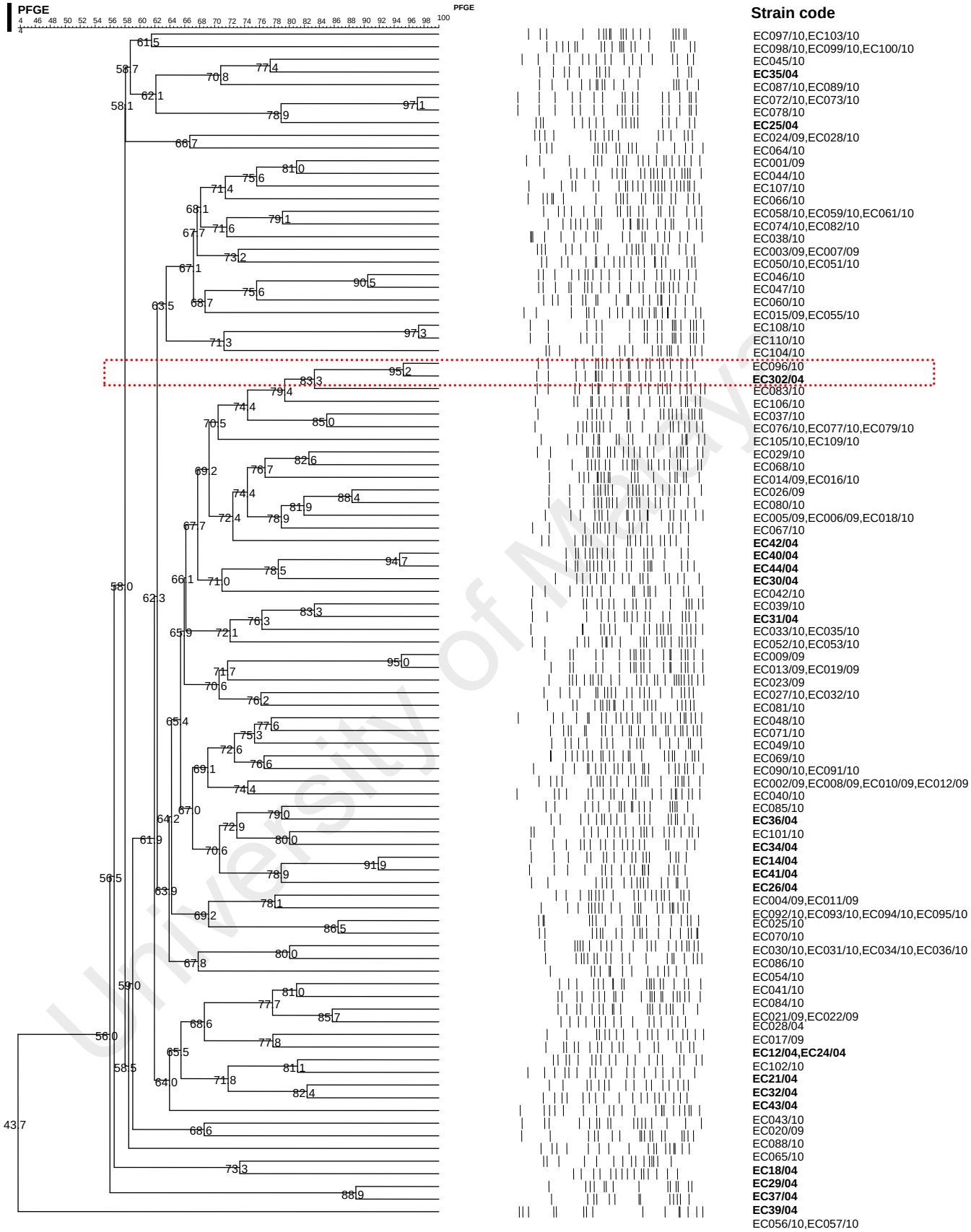


Figure 3.7: Dendrogram generated from PFGE profiles of the 130 *E. coli* strains isolated in years 2004 and 2009/2010 obtained from the same hospital**.** Bold letters of strain codes (/04) indicate that the strains were obtained from the previous study (year 2004) (Lim et al., 2009). Strains EC302/04 and EC096/10 were isolated 5 years apart and showed one band difference in their PFGE profiles (red box).

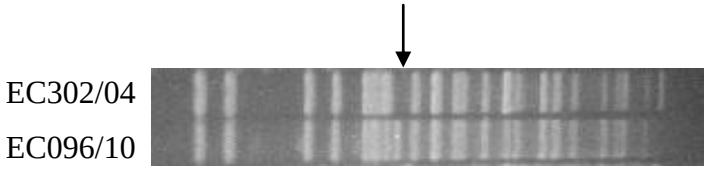


Figure 3.8: Pulsed-field profiles of EC302/04 and EC096/10

3.3.5 Prevalence of integrons in ESBL- and non-ESBL- producing strains

A total of 15 out of the 24 integrons (class 1 and class 2) detected were found in non-ESBL-producing *E. coli* and no significant correlation was found between the ESBL phenotype and the carriage of integrons ($p = 0.23$).

3.4 DISCUSSION

Antimicrobial resistance in *E. coli* from children's stool samples in this study was compared with previous reports from Malaysia (MOH, 2014; Lim et al., 2009) to observe the trend in antimicrobial resistance keeping in mind that the respective study populations were different. All 3 reports shared similar findings whereby *E. coli* strains had low resistance rates to carbapenems (0% - 0.3%) and high resistance rates to ampicillin (60% - 77%). Notably, the resistance rate of cefotaxime reported by the National Surveillance of Antibiotic Resistance 2014 (MOH, 2014) is approximately 100% higher (22.9%) compared to the study from Lim et al. (2009) (11%) as well as this study (10.9%). The hike in resistance rates to the extended-spectrum cephalosporins is of concern as cephalosporin resistant *E. coli* is recognized as one of the drug resistance threat globally (CDC, 2013).

Half (50%, n=12/24) of the non-susceptible ESC (intermediate and resistant) strains were CTX-M positive. CTX-M was the most common type of ESBL observed in both studies (this study: 10.9%; Lim et al. (2009): 17%) with CTX-M-15 as the most common subtype in Malaysia (Lim et al., 2009; Sekawi et al., 2008) and India (Shakil et al., 2010). Although CTX-M-14 is the most prevalent variant of CTX-M in other Asian countries such as Korea, China and Taiwan (Song et al., 2009; Yan et al., 2006; Yu et al., 2007), it is relatively low in Malaysia with rates of 17% (n=2/12) in this study and 31%, (n=4/13) in a previous study (Sekawi et al., 2008). Perhaps, the prevalence of a particular genotype is associated with different geographical regions. To the best of our knowledge, this is the first report of the presence of a CTX-M-27-producing *E. coli* strain in Malaysia. The low prevalence of CTX-M-27 concurred with the reports from Korea (1.3%) (Song et al., 2009) and Canada (0.6%) (Pitout et al., 2007) and Saudi Arabia (3.2%) (Al-Agamy et al., 2014).

E. coli strains that carried CTX-M genes exhibited higher resistance to cefotaxime than ceftazidime (MICs of 8 - 32 fold dilutions), consistent with the classical phenotype of CTX-M enzymes. Nevertheless, a broad range of resistance levels to cefotaxime (MICs of 24 - 256 μg/ml) for the *E. coli* strains that carried the CTX-M-encoding genes was observed, similar to a previous report by (Edelstein et al., 2003). CTX-M-15 has stronger hydrolytic activities against ceftazidime (MIC: 2 - 16 μg/ml) compared to its parental enzyme, CTX-M- 14 (MIC: 1 - 2 μg/ml). *E. coli* strains containing CTX-M-27 also exhibited higher resistance levels to ceftazidime compared to CTX-M-14 (MIC = 8 versus 1 - 2 μg/ml) which is in agreement with a previous report from Bonnet et al. (2003).

Besides CTX-M-encoding genes, AmpC genes also play an essential role in conferring non-susceptibility to broad spectrum cephalosporins (Jacoby, 2009). In this

study, CMY-2 was the only AmpC gene detected in 6 of the 110 *E. coli* strains. Five out of 6 CMY-2-positive strains were non-susceptible to ESC while another CMY-2-positive strain had reduced susceptibility. All CMY-2-positive strains were resistant to amoxicillin/ clavulanic acid, a known trait of AmpC-type β-lactamase (weakly inhibited by β-lactamase inhibitors) (Bauernfeind et al., 1996), and consistent with an AmpC-type β-lactamase phenotype. The CMY-2 producing strains also had a higher resistance level (MICs) to ceftazidime than for cefotaxime, which is in agreement with the findings of Philippon et al. (2002). Plasmid-encoded AmpC-type β-lactamase-producing strains are often not resistant to third generation cephalosporins because they are found in low copy numbers (Hanson, 2003). It would thus be of interest to find out if the AmpC-encoding genes in these *E. coli* strains are plasmid- or chromosomally-encoded.

The remaining 7 strains that were non-susceptible to ESC did not contain any ESBL or AmpC genes that were tested in this study, indicating that resistance to cephalosporins may be conferred by other resistance mechanisms (Pfeifer et al., 2010), such as OXA of other subtypes, as well as rare ESBL types (bla_{PSE} and bla_{PER}) (Bradford, 2001) in which their detection by PCR detection was not carried out in this study.

The association between ESBL genes and integrons has been reported worldwide (Machado et al., 2005; Machado et al., 2007; Vinué et al., 2009), but no such studies have yet been carried out for Malaysian *E. coli* strains. In this study, class 1 integrons were the most common integrons in the Malaysian *E. coli* strains, but no gene cassettes that contained ESBL genes were detected. Sequence analyses of the amplicons revealed a predominance of gene cassettes that conferred resistance to streptomycin and spectinomycin (*aadA* genes) and trimethoprim (*dfr* genes), similar to an earlier report by White et al. (2001). A total of 15 out of 24 integrons (class 1 and class 2) were found in

non-ESBL-producing *E. coli*. Integrons were also not significantly found in either ESBL or non-ESBL producers (P>0.05). These findings inferred that integrons are unlikely to have contributed to the dissemination of the ESBL genes in the Malaysian *E. coli* strains. Previous studies by Machado et al., (2005; 2007) and Vinué et al. (2009) also indicated that integrons did not contribute much to the dissemination of ESBL genes unless these genes were located within integrons (such as the CTX-M-9 gene that was linked to the In60 integron). The transmissibility of ESBL genes could thus be linked to other mechanisms such as insertion elements (Machado et al., 2005; Machado et al., 2007) and plasmids (Saladin et al., 2002).

E. coli strains that contained CTX-M genes were genetically diverse. This observation concurred with other studies (Niumsup et al., 2008; Vinué et al., 2009) which reported low clonal relationships among the strains containing CTX-M genes. This indicates that the dispersion of the *E. coli* strains with CTX-M genes was not likely due to the dissemination of a particular resistant clone harboring the gene. However, the 3 strains that contained CTX-M-15 (EC013/09, EC019/09 and EC009/09) were highly related. These strains shared similar PFPs and resistotypes and might have derived from the same clone. It was found that the PFPs of EC096/10 that was isolated from a stool sample in this study, and EC302/04 that was isolated from tracheal aspirate from an earlier study (Lim et al., 2009) differed by only a single band. Similar observations whereby epidemiologically unrelated *E. coli* strains had closely related PFGE patterns have been reported by Mamlouk et al. (2006). Further analysis of these strains showed that they were of different resistotypes. No integron was detected in both strains while $bla_{\text{TEM-1}}$ was detected only in EC302/04. According to van Belkum et al. (2007), strains with indistinguishable typing patterns do not necessarily belong to the same clone because different strains may yield indistinguishable patterns using the same typing method. Nevertheless, the high discriminatory power of PFGE strongly implies that

EC096/10 and EC302/04 were derived from the same clone that had persisted over five years in the same hospital environment with gain or loss of resistance due to the acquisition or deletion of respective genes, which are likely to be located on mobile genetic elements. This was further explored when the whole genome sequences for EC096/10 and EC302/04 were determined and analyzed (Chapter 6).

3.6 CONCLUSION

In summary, the 110 *E. coli* strains isolated from pediatric wards of a tertiary hospital in the state of Johor, Malaysia showed high resistance rates to ampicillin and total susceptibility to carbapenems. The MDR rates were lower compared to other previously reported Malaysian strains. Class 1 integrons remained the major class of integrons in the Malaysian *E. coli* clinical strains. Nonetheless, no gene cassette encoding ESBL genes was found and no significant association could be made between ESBL production and the carriage of integrons. CTX-M-15 was the most common ESBL found in this study although its low prevalence was reported in other parts of Asia, indicating that the prevalence of genotypes is perhaps associated with different geographical regions. Possible clonal expansion was observed for a few CTX-M-15-positive strains but no O25-ST131 clone was detected, indicating that *E. coli* of other clones may have played a role in the clonal dissemination of CTX-M-15-producing *E. coli* in Malaysia. Besides dissemination via clonal expansion, CTX-M genes have also been found associated with various mobile genetic elements. Nonetheless, both the O25-ST131 clone and integrons did not seem to play a role in the spread of CTX-M-15-producing *E. coli* in Malaysia. Hence, further studies focusing on the cephalosporin resistance mechanisms of *E. coli* strains from various sources and time frames are necessary to further understand the dissemination of such resistance determinants in the local *E. coli* strains.

CHAPTER 4: GENETIC DIVERSITY OF ESBL-PRODUCING *E. COLI* STRAINS FROM MALAYSIA AND THE GENETIC ENVIRONMENT SURROUNDING THE GENES ENCODING THE CTX-M TYPE ESBL

4.1 INTRODUCTION

The previous chapter (Chapter 3) had revealed that CTX-M-15 was the most common ESBL subtype identified in the Malaysian pediatric *E. coli* strains but no integron-encoded gene cassette encoding ESBL genes was found. A few CTX-M-producing strains with indistinguishable pulsotypes were also identified in Chapter 3 but the renowned CTX-M-15-harboring *E. coli* O25-ST131 strain was not detected. Nonetheless, mobile genetic elements such as insertion sequences and complex integrons often associated with the CTX-M genes amongst the *E. coli* population (Lartigue et al., 2004; Literacka et al., 2009; Poirel et al., 2008). In addition, *E. coli* of different STs harbouring CTX-M genes can also predominate different geographical regions (Riley, 2014).

To better understand the prevalence and genetic mechanisms associated with the dissemination of CTX-M genes, the strains included in this chapter has been expanded encompassing local strains from various sources, time frames and geographical locations in Malaysia. This chapter aimed to determine the (1) prevalence of CTX-M genes and their respective variants, (2) sequence types (STs) for fluoroquinolone-resistant and/or CTX-M-bearing *E. coli*, (3) genetic environment for the CTX-M genes and (4) genetic diversity of CTX-M-producing strains using PFGE. The presence of particular *E. coli* clone (determined using MLST and PFGE) in Malaysia may provide further understanding on the contribution of clonal spread to the dissemination of CTX-M genes.

4.2 METHODOLOGY

4.2.1 Bacterial strains

A total of 189 non-repeat *E. coli* isolates collected from various samples over a 10-year period (2002-2011) from different parts of Malaysia were studied. A total of 53 clinical isolates (tracheal aspirates, n=2; stool, n=28; blood, n=4; urine, n=7; pus, n=1; swabs, n=6; body fluids, n=2; tracheal secretory, n=2; tissue, n=1) were obtained from 6 public Malaysian hospitals (Kota Bharu Hospital, Ipoh Hosptial, Sultanah Aminah Hospital, Kuala Lumpur Hospital, Queen Elizabeth Hospital and University Malaya Medical Centre) located in six states in Malaysia. Forty-seven isolates were from zoonotic samples that were recovered from swabs (nasal, rectal and tongue) of pigs from swine farms located in Perak and Penang (Ho et al., 2013). Another 55 *E. coli* isolates were isolated from food samples that were obtained from different sources (seafood, n=23; dairy products, n=18; grilled meat, n=11; vegetables, n= 2 and poultry, n=1) and various locations in Selangor. On the other hand, a total of 34 environmental samples were obtained from aquatic environment in Kelantan and Selangor. Of the 53 clinical isolates, 33 of them were previously determined to be ESBL-producers (Chapter 3; Lim et al., 2009). The characterization of the remaining *E. coli* isolates has not been reported before. All *E. coli* strains were confirmed by positive PCR-amplification of the housekeeping gene *phoA* of *E. coli* (Kong et al., 1999) (Appendix III a).

4.2.2 Antibiotic susceptibility test to screen for putative ST131 clone

Antibiotic susceptibility of *E. coli* strains was determined using disk diffusion method according to the CLSI guidelines (CLSI, 2015). Antimicrobial agents used were: aztreonam (ATM, 30 μg); ceftriaxone (CRO, 30 μg); cefotaxime (CTX, 30 μg); ceftazidime (CAZ, 30 μg); nalidixic acid (NAL, 30 μg), ciprofloxacin (CIP, 5 μg), imipenem (IPM, 10 μg) and meropenem (MEM, 10 μg) (Appendix VI b). Only selected

antimicrobials were used to screen for ST131 clone as this clone is often resistant to extended-spectrum cephalosporins and fluoroquinolones (Johnson et al., 2010; Zhao & Hu, 2013). Carbapenems were also included due to its clinical importance as the "last resort" antimicrobial agent (Johnson et al., 2010).

4.2.3 Phenotypic and genotypic detection of ESBL

Presumptive ESBL-producers were determined by using four antimicrobial disks containing CAZ (30 mg), CRO (30 mg), CTX (30 mg), and ATM (30 mg) (Appendix VI c i). ESBL production of presumptive ESBL-producers was determined using modified double disk synergy test (DDST) described by Jarlier et al., (1988) (Appendix VI c ii) and E-test ESBL strips (CTX/CTX + clavulanic acid [CT/CTL] and CAZ/CAZ + clavulanic acid [TZ/TZL]) (AB Biodisk) (Appendix VI c iii). Two methods were used together to increase the efficiency of ESBL detection. Strains that showed positive results for either DDST or E-test ESBL strips were categorized as ESBL-producers.

The presence of CTX-M genes was confirmed via PCR and DNA sequencing analysis (Pagani et al., 2003). For CTX-M positive strains, 4 sets of primers were used to subgroup the CTX-M genes into groups 1, 2, 8/25, and 9 (Ensor et al., 2007). CTX-M-15 was detected using the primers previously described by Conceição et al. (2005) (Appendix III b).

4.2.4 Minimum inhibitory concentrations

The minimum inhibitory concentrations (MICs) for selected antimicrobial agents (cefotaxime, ceftazidime, ceftriaxone and aztreonam) were determined using E-test strips (AB Biodisk, Solna, Sweden) (Appendix IV b). *E. coli* ATCC 25922 and ATCC35218 strains were used as controls (CLSI, 2015).

4.2.5 Phylogenetic grouping

E. coli strains are composed mainly of 4 phylogenetic groups (A, B1, B2 and D) with extraintestinal pathogenic *E. coli* often belonging to phylogroups B2 and D (Clermont et al., 2000). The phylogroup of the fluoroquinolone and/or cephalosporin non-susceptible (intermediate and resistant) and/or CTX-M positive *E. coli* strains was determined using a triplex PCR assay as previously described (Clermont et al., 2000) (Appendix III e).

4.2.6 Detection of the O25-ST131 clone and MLST

Rapid detection of the O25-ST131 *E. coli* clone was carried out (Clermont et al., 2009) for fluoroquinolone (FQ) and/or cephalosporin non-susceptible and/or CTX-M-producing strains (Appendix III c). For CTX-M positive strains that were non-O25-ST131, multilocus sequence typing (MLST) was carried out using 7 housekeeping genes (Wirth et al., 2006) to determine their sequence types (Appendix III f; IV e i). Representatives of ST131 clones determined by the rapid PCR detection method were also further validated by MLST. Allelic profile and sequence type (ST) were determined using the scheme provided by the MLST website (Appendix IV e i).

4.2.7 Pulsed-Field Gel Electrophoresis (PFGE)

PFGE analysis for the CTX-M-positive strains was performed as previously described with minor modifications (Thong et al., 2007). The protocols for PFGE and data analysis were described in detail in Appendix IV, d, i-vi. Simpson's index was calculated using a webtool (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool).

4.2.8 Genetic environment of CTX-M genes

The genetic environments of the CTX-M genes in CTX-M-positive *E. coli* strains were investigated using PCR and validated by DNA sequencing. The presence of IS*26*, IS*Ecp1*, and IS*903* was determined by PCR using primers targeting the intragenic regions of the respective genes (Literacka et al., 2009; Poirel et al., 2003; Woodford et al., 2004). The linkage between IS*26* and ISE*cp1* with CTX-M genes was determined using primers previously described (Eckert et al., 2006; Pagani et al., 2003; Woodford et al., 2004). The possible inclusion of CTX-M genes in a complex class 1 integron was also studied by determining the linkage between 5'CS, *qacE1*, *orf513*, CTX-M, IS*903*, *orf477*, *mucA* and *orf3* (Dhanji et al., 2011; Eckert et al., 2006; Pagani et al., 2003; Pallecchi et al., 2004) (Appendix III; g, i-iii).

4.3 RESULTS

4.3.1 Antimicrobial resistance rates and prevalence of ESBL-producers

The highest antimicrobial resistance rates of the 189 strains were observed for CIP (37.6%) followed by NAL (35.45%), CTX (23.3%), CRO (22.2%), ATM (21.2%) and CAZ (19.6%). None of them were resistant to the carbapenems (i.e., meropenem and imipenem) tested (Table 4.1). *E. coli* strains from environmental samples had the lowest resistance rate towards the antimicrobials tested followed by zoonotic samples while the clinical strains have the highest resistance rates to all the antimicrobials tested. However, it should be noted that 62% (n=33/53) of the clinical strains have been categorized as ESBL-producers in previous studies (Chapter 3, Lim et al., 2009). On the other hand, a total of 60.9% (n=115/189) of *E. coli* strains were non-susceptible to fluoroquinolones and/or third generation cephalosporins, of which a total of 46 were ESBL-producers, with 36 from clinical, 3 from environmental, 2 from food and the remaining 5 from zoonotic samples (Table 4.1).

Table 4.1: Prevalence of antimicrobial resistance and ESBL-producers among *E. coli* strains isolated from various sources

Source	Total no. of strains	Number of resistant strains (%)												Strains non-susceptible to FQ and 3rd generation cephalosporins	Number of ESBL-producers						CTX-M positive strains	
		NAL		CIP		CRO		ATM		CAZ		CTX			DDST		E-test ESBL		DDST or E-test ESBL			
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	No.	%	No.	%	No.	%	No.	%
Clinical	53	35	52.24	23	32.4	34	81	35	87.5	29	78.4	35	79.6	43	36	83.7	34	87.2	36	78.3	37	100
Environ.	34	2	2.99	4	5.6	3	7.1	3	7.5	3	8.1	3	6.8	6	1	0	3	7.7	3	6.5	0	0
Food	55	16	23.88	12	16.9	0	0	0	0	0	0	0	0	22	2	4.7	1	2.6	2	4.3	0	0
Zoonotic	47	14	20.9	32	45.1	5	11.9	2	5	5	13.5	6	13.6	44	5	11.6	1	2.6	5	10.9	0	0
TOTAL	189	67	35.45	71	37.6	42	22.2	40	21.2	37	19.6	44	23.3	115	43	22.8	39	20.6	46	24.3	37	19.6

*Numerals shaded in grey were mentioned in section 4.4.1.

4.3.2 Molecular detection of the *E. coli* O25b-ST131 clone and phylogenetic grouping

A total of 115 *E. coli* strains that were non-susceptible to fluoroquinolones and/or third generation cephalosporins were subjected to PCR detection of the O25b-ST131 clone. A total of 6 clinical strains were positive for O25b-ST131 using the primers previously described by Clermont et al. (2009) (Figure 4.1 a) while all non-clinical strains did not belong to ST131 clone.

Of the 115 strains, majority (67%) belonged to phylogenetic group A. Notably, 100% (n=22/22) of *E. coli* strains from food samples and 91% (n=40/44) of the zoonotic strains that were non-susceptible to fluoroquinolones and/or third generation cephalosporins belonged to phylogenetic group A (Table 4.2; Figure 4.1 b).

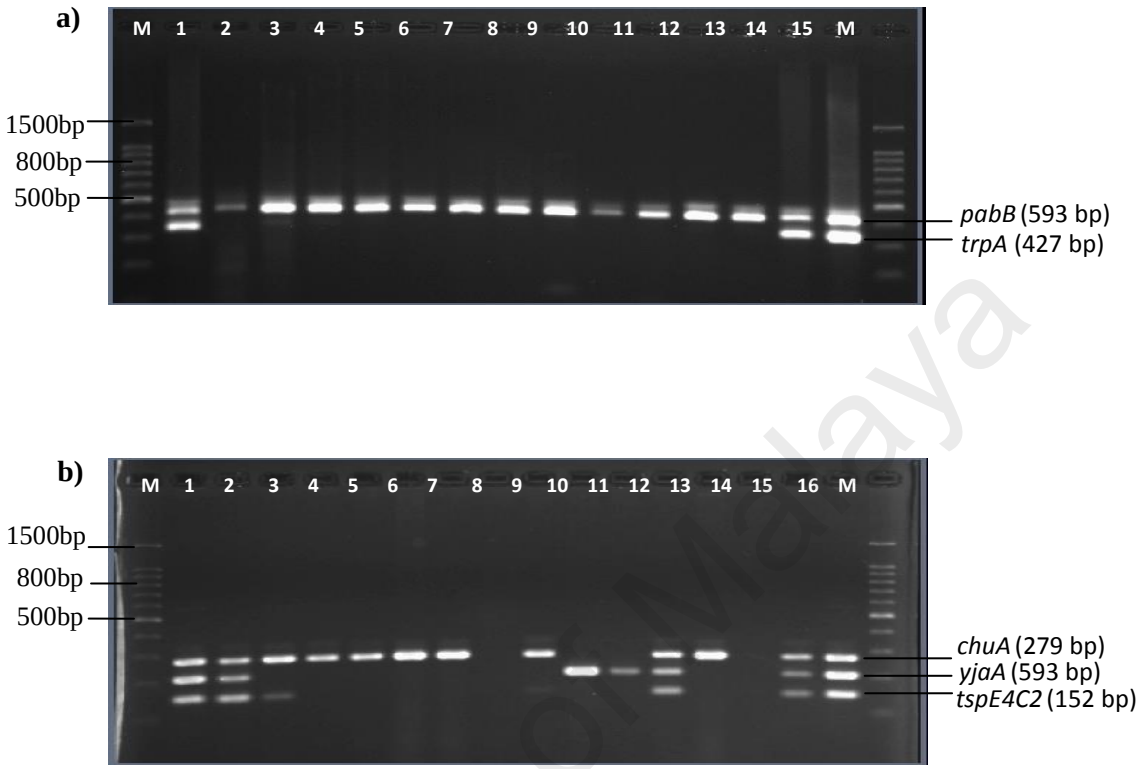


Figure 4.1: Representative agarose gel pictures of: a) PCR detection of *E. coli* O25:ST131 clone. . Lanes 2-15: EC058/10, EC059/10, EC068/10, EC075/10, EC077/10, EC080/10, EC082/10, EC087/10, EC090/10, EC101/10, EC106/10, EC108/10, EC109/10, EC021/10. **b) PCR-based *E. coli* phylogrouping.** Lanes 2-16: EC059/10, EC060/10, EC061/10, EC062/10, EC063/10, EC064/10, EC065/10, EC066/10, EC067/10, EC068/10, EC069/10, EC070/10, EC071/10, EC072/10, EC073/10. Lane M: 100 bp DNA ladder (Promega); Lane 1: positive control.

Table 4.2: Prevalence of phylogroups A, B1, B2 and D in *E. coli* strains that were isolated from various sources and non-susceptible to fluoroquinolones and third generation cephalosporins

Source	Total no. of strains	No. of strains non-susceptible to FQ and 3rd generation	Phylogroup			
			A	B1	B2	D
		No.	No.	No.	No.	No.
Clinical	53	43	12	6	11	14
Environ.	34	6	3	1	2	0
Food	55	22	22	0	0	0
Zoonotic	47	44	40	1	0	3
TOTAL	189	115	77 (66.96%)	8 (6.96%)	13 (11.30%)	17 (14.78%)

4.3.3 CTX-M-15 is the most prevalent CTX-M subtype with the highest MIC values for third generation cephalosporins and aztreonam

Out of 46 ESBL-producing *E. coli*, 35 were positive for CTX-M genes with CTX-M-15 as the predominant subtype (n=28, 80.0%), followed by CTX-M-14 (n=5, 14.3%), CTX-M-27 (n=1, 2.9%) and CTX-M-9 (n=1, 2.9%). All 35 CTX-M-positive *E. coli* were from clinical samples.

E. coli strains harboring each type of CTX-M enzyme showed the following MIC values for the respective third generation cephalosporin and aztreonam: CTX-M-15 (CRO: 16-256 μg/ml, CTX: 16-256 μg/ml, CAZ: 2-48 μg/ml, ATM: 3-256 μg/ml); CTX-M-14 (CRO: 32-256 μg/ml, CTX: 16-256 μg/ml, CAZ: 0.5-4 μg/ml, ATM: 3-8 μg/ml); CTX-M-27 (CRO: 256 μg/ml , CTX: 128 μg/ml, CAZ: 8 μg/ml, ATM: 8 μg/ml) and CTX-M-9 (CRO: 24 μg/ml, CTX: 48 μg/ml, CAZ: 1.5 μg/ml, ATM: 3 μg/ml). Overall, *E. coli* strains harboring the CTX-M enzymes showed higher MIC values for cefotaxime compared to ceftazidime.

4.3.4 Predominant sequence types among CTX-M-producing *E. coli* strains

A total of six O25-ST131 clones were detected in this study using a triplex PCR detection method (Clermont et al., 2009). The sequence type of the O25-ST131 clones was further validated by full MLST (Figure 4.2). Other predominant ST types that were observed among the CTX-M-producing *E. coli* strains include ST1171 (n=4), ST354 (n=3), ST349 (n=2) and ST69 (n=2) (Figure 4.3).

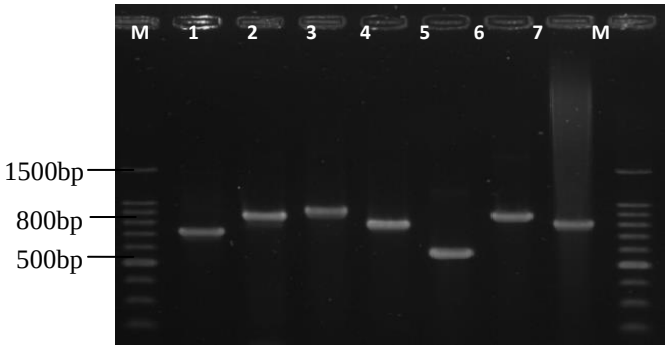
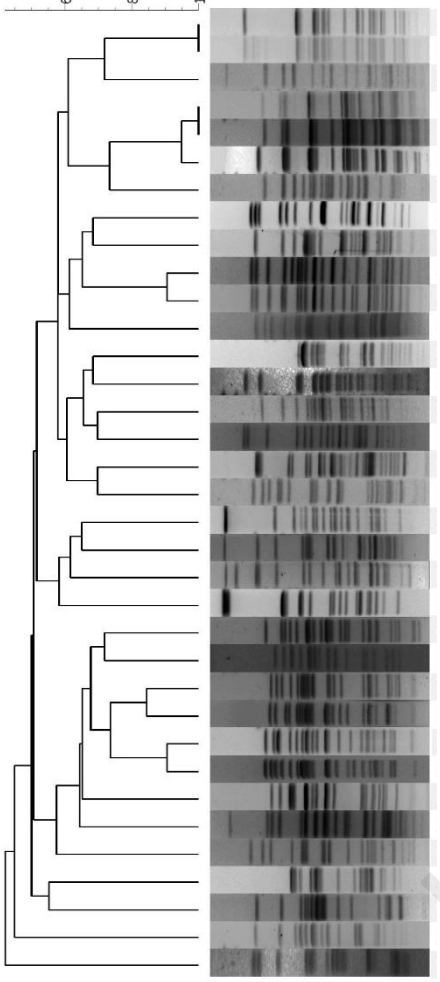


Figure 4.2: Representative agarose gel picture of multilocus sequence typing of an *E. coli* isolate. Lane M: 100 bp ladder (Promega). Lanes 1 - 7: *recA* (780 bp), *icd* (878 bp), *mdh* (932 bp), *fumC* (806 bp), *adk* (583 bp), *gyrB* (911 bp), *purA* (816 bp).

4.3.5 PFGE revealed high genetic diversity among CTX-M-positive strains

A total of 33 pulsotypes were observed for 35 CTX-M-producing strains, with a Simpson's index diversity value of 0.997, and F-value (coefficient of similarity) ranging from 0.42 to 1.0 (Figure 4.3), indicating the high genetic diversity of the CTX-M-producing *E. coli* strains.



Strain code	Year of isolation	CTX-M subtype	Phylo-group	ST[1]	MIC value (S/I/R*) CRO	CTX	CAZ	ATM	Genetic environment of CTX-M genes
EC033/10	2010	CTX-M-15	D	69	16 (R)	16(R)	2 (S)	3 (S)	unknown-CTX-M-15-orf477
EC035/10	2010	CTX-M-15	D	69	16 (R)	16 (R)	2 (S)	3 (S)	unknown-CTX-M-15-orf477
EC/1003-8	2010	CTX-M-15	D	354	256 (R)	256 (R)	6 (S)	32 (R)	unknown-CTX-M-15-orf477
EC013/09	2009	CTX-M-15	A	1771	256 (R)	256 (R)	12 (I)	48 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC019/09	2009	CTX-M-15	A	1771	256 (R)	256 (R)	16 (R)	96 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC009/09	2009	CTX-M-15	A	1771	256 (R)	256 (R)	32 (R)	32 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1004-10	2010	CTX-M-15	B1	156	256 (R)	256 (R)	12 (I)	64 (R)	unknown-CTX-M-15-orf477
EC007/09	2009	CTX-M-15	A	1771	256 (R)	256 (R)	12 (I)	64 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1003-4	2010	CTX-M-15	B1	602	64 (R)	48 (R)	3 (S)	16 (R)	ISEcp1-CTX-M-15-orf477
EC046/10	2010	CTX-M-15	B2	372	48 (R)	48 (R)	3 (S)	8 (I)	unknown-CTX-M-15-orf477
EC047/10	2010	CTX-M-15	B2	4582	64 (R)	48 (R)	4 (S)	12 (I)	ISEcp1-CTX-M-15-orf477
EC060/10	2010	CTX-M-15	D	349	48 (R)	256 (R)	4 (S)	12 (I)	unknown-CTX-M-15-orf477
EC/1003-10	2010	CTX-M-15	A	617	256 (R)	256 (R)	16 (R)	64 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC300KT	2004	CTX-M-9	B2	95	24 (R)	48 (R)	1.5 (S)	3 (S)	qacE-orf513-CTX-M
EC080/10	2010	CTX-M-14	B1	4583	48 (R)	24 (R)	1 (S)	3 (S)	unknown-CTX-M-14-unknown
EC101/10	2010	CTX-M-15	A	226	16 (R)	16 (R)	8 (I)	3 (S)	unknown-CTX-M-15-orf477
EC/1003-2	2010	CTX-M-15	D	349	32 (R)	128 (R)	16 (R)	6 (S)	unknown-CTX-M-15-orf477
EC/1003-9	2010	CTX-M-15	D	57	256 (R)	256 (R)	24 (R)	128 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1003-5	2010	CTX-M-15	D	354	256 (R)	256 (R)	24 (R)	94 (R)	unknown-CTX-M-15-orf477
EC/1004-6	2010	CTX-M-15	A	2197	256 (R)	256 (R)	16 (R)	32 (R)	unknown-CTX-M-15-orf477
EC/1004-2	2010	CTX-M-15	D	405	256 (R)	256 (R)	24 (R)	48 (R)	unknown-CTX-M-15-orf477
EC/1003-7	2010	CTX-M-15	D	354	256 (R)	256 (R)	8 (I)	48 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1004-7	2010	CTX-M-15	B2	131	256 (R)	256 (R)	12 (I)	32 (R)	unknown-CTX-M-15-orf477
EC261KT	2004	CTX-M-14	B2	131	32 (R)	16 (R)	0.5 (S)	3 (S)	ISEcp1-CTX-M-14-unknown
EC/1004-5	2010	CTX-M-14	B2	2279	256 (R)	256 (R)	3 (S)	3 (S)	ISEcp1-CTX-M-14-IS903
EC/1004-9	2010	CTX-M-14	B2	131	32 (R)	256 (R)	1.5 (S)	6 (I)	ISEcp1-CTX-M-14-IS903
EC/1003-1	2010	CTX-M-15	B2	131	64 (R)	128 (R)	16 (R)	16 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1004-1	2010	CTX-M-15	B2	131	256 (R)	256 (R)	48 (R)	256 (R)	IS26- ΔISECp1-CTX-M-15-orf477
EC021/09	2009	CTX-M-14	B2	131	64 (R)	48 (R)	2 (S)	8 (I)	ISEcp1-CTX-M-14-IS903
EC040/10	2010	CTX-M-27	B2	12	256 (R)	128 (R)	8 (I)	8 (I)	int-orf513-CTX-M-27-orf3
EC/1004-4	2010	CTX-M-15	A	46	256 (R)	256 (R)	12 (I)	32 (R)	IS26- ΔISECp1-CTX-M-15-orf477
EC/1003-6	2010	CTX-M-15	B1	533	256 (R)	256 (R)	16 (R)	48 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1004-3	2010	CTX-M-15	B1	224	256 (R)	256 (R)	12 (I)	32 (R)	IS26-ΔTn3-ISEcp1-CTX-M-15-orf477
EC/1003-3	2010	CTX-M-15	D	117	32 (R)	256 (R)	16 (R)	3 (S)	ISEcp1-CTX-M-15-orf477
EC187KT	2004	CTX-M-15	A	5890	256 (R)	256 (R)	32 (R)	64 (R)	unknown-CTX-M-15-unknown

Figure 4.3: Dendrogram generated from PFGE profiles of the CTX-M-positive *E. coli* strains. *S = susceptible; I = intermediate; R = resistance. ST[1] means "sequence type".

4.3.6 Regions surrounding CTX-M-encoded genes

Besides the dissemination of CTX-M-encoded genes via clonal expansion, there are also other mechanisms that are directly involved in the mobilization of CTX-M-encoded genes. These genetic elements include insertion sequences, complex *sul1*-type class 1 integrons bearing IS*CR1* or phage-related elements (D'Andrea et al., 2013). A total of 10 types of genetic environments were found among the 35 *E. coli* strains that were positive for $bla_{\text{CTX-M}}$, with 5 types being observed for strains carrying the CTX-M-15 genes, 3 types for the CTX-M-14 genes, while 1 each for CTX-M-9 and CTX-M-27 genes. Intriguingly, *orf477*, encoding a hypothetical protein (Alves et al., 2014), was found downstream of 27 out of 28 CTX-M-15 genes but was not associated with CTX-M-9, CTX-M-14 and CTX-M-27-encoded genes. In fact, *orf477* has been found downstream of wide variety of CTX-M subtypes such as CTX-M-1, CTX-M-15, CTX-M-3 and CTX-M-32 although its function is currently unknown (Eckert et al., 2006; Vinué et al., 2008; Zhao & Hu, 2013). Meanwhile, IS*26* and ISE*cp1* were commonly found upstream of $bla_{\text{CTX-M-15}}$ (Figure 4.3). Nevertheless, the $bla_{\text{CTX-M-15}}$ upstream regions for 14 *E. coli* strains could not be determined as all PCR targeting various reported upstream regions failed to yield any positive band.

Out of five CTX-M-14-producing strains, four harbours ISE*cp1* as the upstream of $bla_{\text{CTX-M-14}}$ genes. On the other hand, IS*903* was found downstream of $bla_{\text{CTX-M-14}}$ in three *E. coli* strains whereas in the remaining two *E. coli* strains, the downstream region for $bla_{\text{CTX-M-14}}$ could not be determined. These two types of genetic environments for the $bla_{\text{CTX-M-14}}$ gene were also commonly found in many other studies (Eckert et al., 2006; Kim et al., 2011; Lartigue et al., 2004; Zhao & Hu, 2013). The one and only CTX-M-27 gene identified in this study was found embedded in a complex class 1 integron bearing IS*CR1*. IS*CR1* was also identified as the upstream region of the $bla_{\text{CTX-M-9}}$ gene in this study. It should be noted that IS*CR1* was formerly recognized as *orf513* (Toleman et al., 2006). Despite the varieties of genetic elements associated with the genes encoding the various CTX-M-subtypes, the genetic

environment of several *bla*$_{CTX-M}$ genes could not be determined, suggesting possible novel genetic structures and arrangements flanking these CTX-M-encoded genes.

4.4 DISCUSSION

Antimicrobial resistance is now a major public health concern worldwide (WHO, 2014) including Malaysia (MOH, 2014), due to its association with high mortality and morbidity. Appropriate surveillance programs are needed to address the antibiotic resistance issue by determining the scope, magnitude and trends of the resistance problem. Third generation cephalosporins are often used as the first line treatment for *E. coli* infections and increased resistance to third generation cephalosporins has emerged as a public health threat worldwide since the 1990s (Dancer, 2001). In Malaysia, the resistance rates to third generation cephalosporins as well as fluoroquinolones are also on the rise. Reports from the National Surveillance of Antibiotic Resistance of the Ministry of Health, Malaysia (MOH, 2014a, 2011, 2012, 2013) which covered at least 36 hospitals (36 hospitals in 2011 and 2012; 37 hospitals in year 2013; 39 hospitals in 2014) showed elevated resistance to cefotaxime (years 2011-2012-2013-2014: 15.8%- 20.2%- 22.9%- 22.9%), ceftazidime (11.7%- 14.8%- 17.1%- 16.9%) as well as ciprofloxacin (21.2%- 23%- 23.4%- 23.2%). On the other hand, although carbapenem resistance poses another great threat to the public health sector globally, the resistance rate of *E. coli* to carbapenem (antibiotics of last resort) in 36-37 local clinical settings across Malaysia were still low [year 2011-2012-2013-2014: imipenem (0.2%- 0.2%- 0.3%- 0.3%); meropenem (0.3%- 0.3%- 0.3%- 0.2%)] (MOH, 2011, 2012, 2013). Hence, increased resistance to third generation cephalosporins in *E. coli* isolates remained more of an issue rather than carbapenem resistance in the Malaysian health care sector.

High resistance to third generation cephalosporins and fluoroquinolones was observed for *E. coli* from clinical samples and zoonotic samples, respectively. On the other hand, *E. coli*

strains from environmental and food sources have relatively low resistance rates to all the antibiotics tested. The results are not unexpected as antibiotics are often used in food animals for nontherapeutic purposes such as growth promotion (Marshall & Levy, 2011). In addition, many other factors may also contribute to the difference in ESBL resistance rates in *E. coli* strains isolated from different sources, such as the isolation time point, geographical regions and types of samples. A study from the neighbouring country Thailand by Booyasiri and coworkers (2014) revealed a high prevalence of ESBL-producing *E. coli* from healthy workers from both farms and food factories, i.e., 72% and 75.5%, respectively. As for food animals, ESBL-producing *E. coli* is most prevalent in pigs (i.e. 69.5%) while broiler chickens has relatively lower prevalence rate of 38.4%. On the other hand, environmental samples such as canal water also had different prevalence of ESBL-producing *E. coli* in different regions within the country, with 80% in Bangkok and 0% in central province. The findings by Booyasiri et al. (2014) also concurred with our findings that *E. coli* strains from different isolation sources have different prevalence of ESBL-producing *E. coli*.

In this study, 91% of *E. coli* isolated from zoonotic samples and 100% from food samples that were non-susceptible to third generation cephalosporins and fluoroquinolones belonged to phylogroup A. It has been reported that *E. coli* strains of phylogroup A consisted mainly of commensals (Carlos et al., 2010), indicating that the *E. coli* strains from zoonotic and food origins in this study were possibly of commensal origin. On the other hand, phylogroups B2 and D, which consisted mainly of extraintestinal pathogenic *E. coli* strains (Carlos et al., 2010), were found mainly in clinical samples (Table 4.2). Phylogroups were differentiallly distributed among the *E. coli* obtained from different sources, a finding which concurred with that of Carlos et al., (2010). Nevertheless, it should be noted that *E. coli* strains from clinical and zoonotic samples were obtained from different sites of the hosts.

This is also the first study reporting on the emergence of the *E. coli* O25:H4-ST131 clone in Malaysia, indicating that O25:H4-ST131 has disseminated to this part of the world. By screening *E. coli* strains that were resistant to third generation cephalosporins and/or fluoroquinolones, six O25:H4-ST131 strains harbouring CTX-M genes were found, with one being detected as early as in the year 2004. O25-ST131 is an *E. coli* clone notorious for its role in the worldwide expansion and dissemination of CTX-M-type ESBL genes (particularly subtype CTX-M-15). In addition, the O25-ST131 clone is also often resistant to fluoroquinolones (Leflon-Guibout et al., 2008), a trait that was observed in all the 6 O25-ST131 strains in this study. *E. coli* O25-ST131 strains are thus resistant to the two most commonly used therapeutic antibiotics (i.e., the extended-spectrum cephalosporins and fluoroquinolones) (WHO, 2014), thereby limiting their therapeutic options. On the other hand, although some studies have reported that the O25-ST131 clones can be fluoroquinolone resistant but susceptible to extended-spectrum cephalosporins (Johnson et al., 2009; Platell et al., 2011), none of the fluoroquinolone resistant and extended-spectrum cephalosporins susceptible strains in this study were found to be O25-ST131. Besides, all six O25-ST131 clones in this study belonged to phylogroup B2, a phylogroup which commonly consists of extraintestinal pathogenic strains (Clermont et al., 2000). In fact, all six were isolated from clinical specimens, with five out of the six O25-ST131 clones obtained from extraintestinal sites (blood, n=2; urine, n=2; tracheal secretion=1) while the remaining isolate was obtained from a stool sample.

CTX-M is the most common ESBL type found in *E. coli* with CTX-M-15 being the most prevalent subtype found in major parts of the world but were reportedly scarce in Asia except for India and Pakistan (Hawkey, 2008). However, our studies (Chapter 3 and this chapter) showed otherwise. Reviewing the current status of ESBLs in neighbouring countries with shared borders with Malaysia (including land and maritime borders), CTX-M-15 is indeed the predominant CTX-M subtype, which concurred with the findings in this study. Severin et al.

(2010, 2012) revealed that CTX-M-15 is the most prevalent ESBL subtype in both *E. coli* and *K. pneumoniae* strains in Indonesia. In Thailand, a study that included two major university hospitals revealed that CTX-M-14 is the most prevalent CTX-M subtype (43.6%) followed by CTX-M-15 (37.2%) (Kiratisin et al., 2008). On the other hand, another study from Thailand reported CTX-M-15 as the predominant CTX-M subtype (Chanawong et al., 2007). Relatively scarce published data are available from Singapore where Tan et al., (2010) reported that CTX-M was the most prevalent ESBL-subtype in one of the hospitals. However, the CTX-M subtypes were not determined (Tan et al., 2010). CTX-M-15 has also been reported to be the most prevalent CTX-M subtype in both Vietnam and the Philippines (Kanamori et al., 2011; Trang et al., 2013) while there is a total lack of any such published data from Brunei. In line with these reports, the Study for Monitoring Antimicrobial Resistance Trends (SMART) group also recently showed that CTX-M-15 is indeed the most prevalent CTX-M-subtype in Malaysia, Philippines, Singapore, Thailand, and Vietnam (Sheng et al., 2013).

Besides the emergence of the O25-ST131 clone, non-ST131 CTX-M positive *E. coli* strains also played an important role in the wide dissemination of CTX-M genes. We encountered several other clones in this study such as ST349 (n=2), ST69 (n=2), ST354 (n=3) and ST1171 (n=4). Surprisingly, two CTX-M-15-positive strains were found to belong to ST349, a rare genotype that has been previously reported to be associated with non-ESBL producers (Ho et al., 2012). ST354, another *E. coli* clonal complex notorious for the dissemination of CTX-M genes in many countries especially in Korea (Kim et al., 2011) was also found in Malaysia, where all three ST354 clones harboured $bla_{\text{CTX-M-15}}$. Unexpectedly, another prevalent sequence type that harboured $bla_{\text{CTX-M-15}}$ in this study, is the ST1171, which has not been reported previously. Three out of 4 of these strains were closely related (with 1 band difference observed among their respective pulsotypes) (Figure 4.3). Only one each of ST10, ST117 and ST405 were identified among the 35 CTX-M-positive strains, although they have been reported to be the dominant clonal groups in other countries (Coque et al., 2008;

Nicolas-Chanoine et al., 2013). The difference in prevalence and distribution of CTX-M genes that are associated with clones of different sequence types indicates that particular clones harbouring $bla_{\text{CTX-M}}$ are possibly associated with different geographical regions, which had also been suggested by Peirano & Pitout (2010).

Interestingly, 5 types of different genetic environments were found for the 6 *E. coli* ST131 clones that harboured the CTX-M genes (Figure 4.3). Two types of genetic environments were observed for the CTX-M-14 genes in the ST131 clone, namely the IS*Ecp1*-CTX-M-14-unknown and IS*Ecp1*-CTX-M-14-IS*903*. Although ISE*cp1* and IS*903* are often located upstream and downstream of CTX-M-14 genes, respectively, CTX-M-14 had been shown to be mobilized by IS*Ecp1* via one-ended tranposition process while IS*903* did not play a role in the mobilization process (Poirel et al., 2005). As for the CTX-M-15 genes, three types of genetic environments were identified (i.e., IS*26*-ΔTn*3*-IS*Ecp1*-CTX-M-15-*orf477*, IS*26*-ΔIS*Ecp1*-CTX-M-15-*orf477*, unknown-CTX-M-15-*orf477*). The Tn*3* transposon, which is disrupted by the IS*Ecp1*-CTX-M-15 element, is one of the most common genetic environments for CTX-M-15 genes and has been observed in previous studies (Price et al., 2013; Smet et al., 2010). The IS*Ecp1* of the IS*Ecp1*-CTX-M-15 element that is truncated by IS*26* and which has been previously described (Dhanji et al., 2011; Eckert et al., 2006), was also found in the ST131 clone of this study. Despite having the same sequence type (i.e., ST131), all 6 strains displayed distinct pulsed-field-profiles. Notably, strain EC/1004-5 which belonged to ST2279 shared a similar pulsed-field profile with the ST131 strain EC/1004-9, with only a three band differences being observed between the two profiles. In fact, ST2279 is a single locus variant of ST131, with both sequence types belonging to the ST131 complex. Interestingly, although all 7 strains of the ST131 complex were clustered together in the dendrogram, 6 out of 7 had very different PFPs (with similarities of > 65%), indicating that PFGE has a higher discriminatory power in distinguishing *E. coli* strains, in comparison with MLST. This is not surprising as MLST includes a limited number of housekeeping genes (7

loci) while pulsed-field profiles is the gross macrorestriction profiles of the whole bacterial genome, hence its higher discriminatory power.

4.5 CONCLUSION

This study served as the first study reporting on the presence of *E. coli* O25-ST131 clone in Malaysia. CTX-M-15 is found to be the most prevalent CTX-M subtype in the Malaysian *E. coli* strains, as well as in the Southeast Asian region. Besides the clonal dissemination of CTX-M genes, mobile genetic elements also played an essential role in the mobilization of CTX-M genes. Analysis of the genetic environment of CTX-M genes revealed a variety of mobile elements, with insertion sequences IS*Ecp1*, IS*26* and IS*903* as the main genetic elements associated with CTX-M genes. The different genetic environments for strains belonging to the same sequence types also indicates the substantial diversity of the genetic environment surrounding $bla_{\text{CTX-M}}$, reflecting the plasticity of the region which harbors the CTX-M-encoding genes. The CTX-M-producing strains that were found to be clonal using MLST had diverse pulsed-field profiles, indicating the higher discriminatory power of PFGE. A high resolution method, such as the whole genome sequencing approach which may be used to investigate the genetic environment of CTX-M genes and phylogenetic relationship of *E. coli* strains simultaneously, is definitely an attractive approach for future epidemiological studies. In fact, whole genome sequencing of bacterial strains served as the ultimate approach for microbial diagnosis, phylogenetic analysis and genetic diversity of various bacterial strains due to its falling cost and turnaround time (Köser et al., 2012). In the next two chapters of this thesis, whole genome sequence analysis was used to study the genetic diversity of various *E. coli* strains, be it specific features of interest of the *E. coli* genomes or high resolution phylogenomic analysis.

CHAPTER 5: DNA DEGRADATION PHENOTYPE AND THE DIVERSITY OF THE *DND* OPERON OF *ESCHERICHIA COLI* STRAINS

5.1 INTRODUCTION

PFGE, the method of choice for epidemiological studies for *E. coli,* yielded incomplete molecular epidemiological data in Chapter 3 as several *E. coli* strains displayed DNA degradation phenotype (Dnd^+), resulting in untypeable strains. The *dnd* cluster, *dndABCDE* was then found to be responsible for the degradation phenotype of a *S. lividans* by mediating the incorporation of sulphur into the DNA backbone via a process called phosphorothioation (Zhou et al., 2005, 1988) (reviewed in section 2.3.3). Nonetheless, instead of a 5 genes *dndABCDE* cluster, *E. coli* harboured a 4 gene *dnd* operon (*dndBCDE*).

The *dnd* operons are often located in chromosomal islands, where highly diverse genetic contexts were observed across different bacterial species (Ou et al., 2009). However, there is a dearth of research on the diversity of the *dnd*-encoding genomic islands from the same bacterial species. To date, only one *dnd*-encoding genomic island has been described for the enteric pathogen *Salmonella enterica* (Ou et al., 2009). A more distantly related bacterium, *Mycobacterium abscessus,* was reported to have conserved regions flanking their *dnd*-encoding GIs but the description of the *dnd*-encoding GIs itself was not available (Howard et al., 2013). The lack of data regarding the genetic environment of the *dnd* operons makes comprehensive comparative analysis between the same and different bacterial species difficult. Although there are ample bacterial genome sequences in the public domain, comparative studies on the *dnd* clusters of *E. coli* and other bacteria are still lacking. Hence, this chapter described the effort to overcome DNA degradation and development of a PCR assay to detect the *dnd* operon in 12 Dnd^+ *E. coli* strains. The association between the presence of the *dnd* operon and the Dnd phenotype was also determined. This study further determines the genetic environment of the *dnd* operons in *E. coli*. The outcome of the study may provide further

insights into the DNA degradation phenotype and the diversification of the *dnd* operons in *E. coli*.

5.2 METHODOLOGY

5.2.1 Pulsed-field gel electrophoresis

A total of 12 *E. coli* strains that were reported to be untypeable from previous studies (Ho et al., 2013; Chapter 3) were further analyzed in this study. Seven of the *dnd*+ strains were verotoxigenic-*E. coli* (VTEC) isolated from different piglets in the same swine farm. The other five were clinical samples obtained from two medical centers (A and B) and from different sources (stool, blood, urine, and swab). The DNA degradation observed in these strains was confirmed by repeating the PFGE three times. PFGE was carried out on *Xba*I-digested genomic DNA of both dnd^+ and dnd^- strains with and without the addition of 50 μM thiourea (Sigma Aldrich, USA) into the 0.5 × TBE buffer and agarose gels (Römling & Tümmler, 2000) (Appendix IV; d; i-v, vii). *Xba*I-digested *Salmonella* enterica serovar Braenderup H9812 was used as the DNA size marker.

5.2.2 PCR assay detecting *dndB, dndC, dndD* and *dndE* genes

A PCR assay targeting the internal sequences of the *dndB*, *dndC*, *dndD* and *dndE* genes were developed based on the *dnd* operons of 3 dnd^+ *E. coli* genome sequences (B7A, 55989 and SE11 with accession numbers of CP005998.1, NC_011748 and NC_011415, respectively) using Primer3 (http://www.ncbi.nlm.nih.gov/tools/ primer-blast/) (Ye et al., 2012). PCR detection of the *iscS* gene was also carried out using primers described previously (An et al., 2012). PCR detection of *dndB, dndC, dndD, dndE* and *iscS* genes was carried out for 12 dnd^+ (which displayed the degradation phenotype) and 48 dnd^- (without degradation phenotype) strains (Appendix III; h). Five each of *Shigella sonnei, Salmonella* serovar Typhimurium and

Salmonella serovar Enteriditis strains that do not display the DNA degradation phenotype were also included as negative controls. The presence of all 4 *dnd* genes was indicative of the presence of the *dnd* operon. Strain that did not yield any of the expected amplicons was considered as negative for the *dnd* genotype. PCR amplified *dndB, dndC, dndD,* and *dndE* genes were sequenced for validation at a commercial facility (First BASE Laboratories).

5.2.3 Association of the DNA degradation phenotypes and genotypes

Fisher's exact test was used to determine the association of the DNA degradation phenotype and genotype using R (version 2.12.2, Vienna, Austria, [www.R-project.org]) (R Development Core Team, 2013). A significance level of $p < 0.05$ was considered as statistically significant.

5.2.4 Comparative analysis of the *dnd* operon with 1 kb of their respective immediate vicinity

The three complete *dnd* operons of *dnd*$^+$ *E. coli* strains (B7A, 55989 and SE11) available from dndDB (*dnd* database, http://db-mml.sjtu.edu.cn/dndDB) were used as query sequences because the *dnd* operons as well as their association with genomic islands (GI) have been reported (Ou et al., 2009). BLASTN against GenBank databases of nr/ nt and WGS was carried out with Enterobacteriaceae set as 'organism of selection' (Altschul et al., 1990). The sequences of the *dnd* operons of all positive hits were then extracted. The immediate vicinity of each of the *dnd* operon was also included by extracting sequences 1000 bp upstream and downstream from the *dnd* operon. A distance of 1000 bp was selected because PCR can then be employed to examine the immediate upstream and downstream regions of the *dnd* operons of the strains in this study. Furthermore, the 1000 bp flanking regions should be sufficient to determine the neighboring gene(s) of the *dnd* operon since an ORF of *E. coli* generally has an average length of approximately 1000 bp (Brown, 2002). All positive hits were considered as

dnd operon positive although the *dnd* phenotypes for most hits have not been reported before. Genetic relationships of the *dnd* operons together with their immediate genetic environment were determined by constructing phylogenetic trees using the maximum-likelihood (ML) method with MEGA 5. The *dnd* operons of *E. coli* genomes were further subgrouped based on their respective immediate upstream and downstream regions as different genetic environments may indicate different evolutionary origins.

5.2.5 Identification, designation and analysis of *dnd*-encoding genomic islands in *E. coli*

One *E. coli* genome was selected from each of the sub-groups (based on their respective immediate genetic environment) for genomic island identification. In cases where only draft genomes were available for that particular subgroup(s), the genome was selected based on the longest assembled contig where the *dnd* operon is located. Genomic islands harbouring the *dnd* operon were determined using the VRprofile server (http://bioinfo-mml.sjtu.edu.cn/VRprofile/) and mGenomeSubtractor tool (Shao et al., 2010). The common features of a genomic island such as organism-atypical G + C contents, integration into the 3′-end of tRNA genes, and integrase-encoding sequences were also analyzed for the putative laterally acquired *dnd*-encoding GIs. A standardized naming system for *dnd*-encoding GIs was also proposed in this study to avoid confusion. Using "dndI_IEC 55989" as an example, "dndI" means "*dnd*-encoding genomic island"; "_I" means "first *dnd*-encoding GI identified in the respective strain" while "EC 55989" depicts the "strain code". The genetic maps, GC contents of representative genomic regions and *dnd*-encoding GIs were illustrated using Easyfig (Sullivan et al., 2011). BLASTN comparison between the *dnd*-encoding GIs were performed with default settings using Easyfig (Sullivan et al., 2011). Search for homologues against the GenBank databases of nr/nt and WGS using BLASTN (http://blast.ncbi.nlm.nih.gov/blast) were carried out for ORFs of representative *dnd*-encoding genomic islands, to determine their match to *dnd*-negative or *dnd*-positive bacterial genomes.

5.2.6 PCR determination of immediate genetic environment of *dnd* operons for 12 *dnd*$^+$ *E. coli* strains

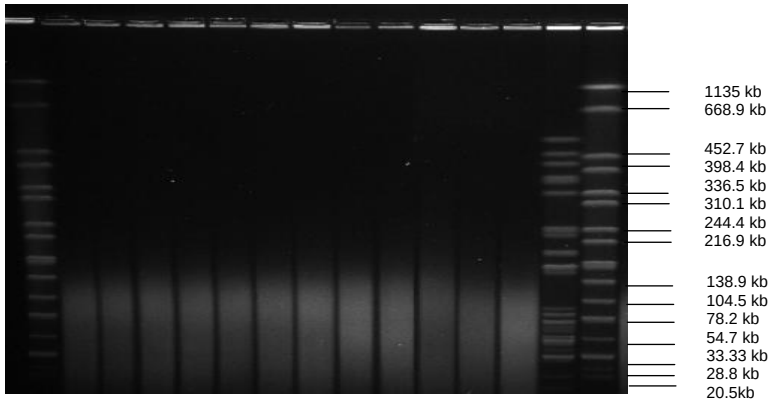
Primers targeting the immediate upstream and downstream regions of the *dnd* operon were also designed based on the 52 sequences extracted from the various *E. coli* genomes. A total of 13 primer pairs were designed based on the 7 types of genetic environments found for *E. coli* (Appendix III; j). The primers were then used for PCR amplification to determine the immediate genetic environment for the 12 *dnd*$^+$ *E. coli* strains isolated in this study.

5.3 RESULTS

5.3.1 Addition of thiourea improved the typeability of *dnd*$^+$ strains by 100%

Without the addition of thiourea to the PFGE run, all 12 *dnd*$^+$ *E. coli* strains yielded degraded DNA despite repeated attempts (3 times) (Figure 5.1 (a)). However, when thiourea was added to the running buffer, there was a 100% improvement in typeability for all the *E. coli* strains (Figure 5.1(b)). Seven zoonotic VTEC strains (lanes S1-S7) and two clinical strains (lanes A1-A2) shared two indistinguishable pulsotypes (i.e., clonal) (Figure 5.1(b)) while another 3 (lanes A3, B1, B2) were genetically different.

a) M S1 S2 S3 S4 S5 S6 S7 A1 A2 A3 B1 B2 C M



b) M S1 S2 S3 S4 S5 S6 S7 A1 A2 A3 B1 B2 C M

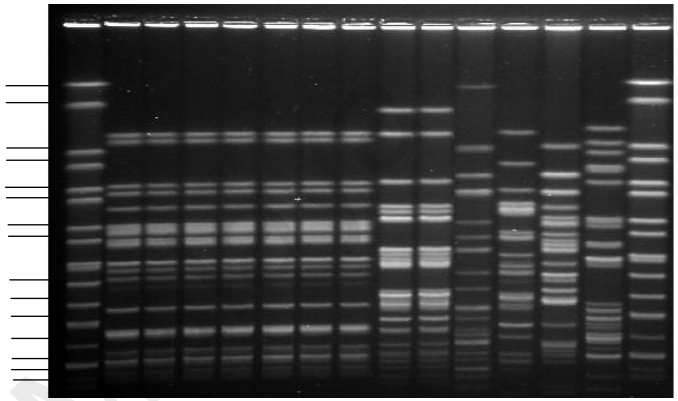


Figure 5.1: Effects of thiourea improved the typeability of *dnd*$^{+}$ strains by PFGE. Gels obtained following a PFGE run: (a) without the addition of thiourea; (b) with the addition of 50 μM thiourea to the agarose gel and running buffer. Lane M: *Xba*I digested H9812 as DNA marker; lanes S1-S7, VTEC isolates from pigs; lanes A1-A3, clinical isolates obtained from medical center A; lanes B1-B2, clinical isolates obtained from medical center B; C, control *E. coli* isolate which is typeable with and without the addition of thiourea (a and b). Gel images were captured using GelDoc (BioRad, Hercules, CA) digital gel documentation system.

5.3.2 Sequence analysis of *dndB-E* genes and association of the DNA degradation phenotypes and genotypes

PCR detection of the *dndC* gene using primers described previously (Wang et al., 2011) revealed that out of 12 *dnd*$^+$ *E. coli* strains, only 7 zoonotic VTEC (Figure 5.1, lanes S1-S7) and 2 clinical (Figure 5.1, lanes B1-B2) *dnd*$^+$ strains were positive for the *dndC* gene. Another 3 *dnd*$^+$ strains (A1-A3) were negative for the *dndC* gene and these strains were obtained from the same medical center and source. The failure to amplify the *dndC* gene from 3 out 12 *dnd*$^+$ strains using previously described primers (Wang et al., 2011) prompted us to develop a PCR assay targeting the internal regions of *dndB*, *dndC*, *dndD*, and *dndE* based on the *dndBCDE* DNA sequences of *E. coli* available in dndDB (Ou et al., 2009). All 12 *dnd*$^+$ strains yielded the expected PCR amplicons for all 4 genes (*dndB, dndC, dndD*, and *dndE*) using the designed primers listed in Appendix III; h (Figure 5.2). All *dnd*$^-$ strains (including *E. coli*, *Shigella sonnei*, *Salmonella* Typhimurium and *Salmonella* Enteriditis) that did not display the degradation phenotype were negative for the *dnd* genes. Significant correlation was thus found between the presence of the *dnd* gene cluster and the *dnd* phenotype ($p < 0.05$).

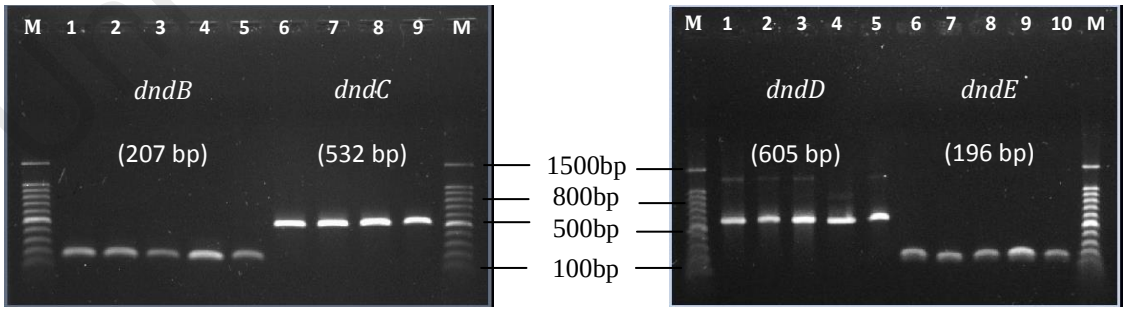


Figure 5.2: Representative pictures of agarose gel of PCR amplified products of *dndB*, *dndC*, *dndD* and *dndE*. Lane M: 100 bp DNA ladder (Promega). Lanes 1, 6: A1; Lanes 2, 7: A2; Lanes 3, 8: A3; Lanes 4, 9: B1; Lane 5, 10: B2.

At least two sets of the amplified *dndB-dndE* genes from different strains with a similar pulsotype were sequenced. It was found that the *dndB-dndE* genes from different strains with the same pulsotype shared identical (100%) nucleotide sequences. Multiple sequence alignment (MSA) of the nucleotide sequences of the *dndB-dndE* genes with *E. coli* SE11, 55989 and B7A were also carried out. Using *E. coli* 55989 as reference, the pairwise alignments showed that *E. coli* S1-S7, B1 and B2 shared *dndB-E* sequences that are highly similar to *E. coli* 55989 (BLASTN identities of *dndB-C* = 100%; *dndD* = 99.8% and *dndE* = 98.5%). Strains A1 and A2 harboured identical *dnd* genes but were different from *E. coli* 55989 (*dndB* = 89.9%; *dndC* = 85.2%; *dndD* = 98.2; *dndE* = 95.9%). The *dndB-E* sequences of strain A3 were the most distantly related to those of *E. coli* 55989 (*dndB* = 87.9%; *dndC* = 89.6%; *dndD* = 80.6%; *dndE* = 84%) but were highly similar to *E. coli* B7A (identical *dndB-D* sequences with 2 SNPs with respect to *dndE*) (Appendix V; b). Hence, the inability to amplify the *dndC* genes using primers described previously (Wang et al., 2011) for strains A1-A3 was likely due to the diversity in the sequences of the *dnd* operons in various *E. coli* strains. Further *in silico* analysis also revealed that mismatches at the primer priming sites (n = 2 – 10) were found for *E. coli* 55989, SE11 and B7A. Nonetheless, the primers described by Wang et al., (2011) were for the expression studies of the *dnd* operon from *Salmonella* enterica (although closely related but not identical to *E. coli*). Furthermore, to the best of our knowledge, there are no other appropriate primers available for the detection of the *dndBCDE* gene cluster in *E. coli*. It should also be noted that the DNA sequences of the *dndB-E* genes included in the MSA are partial CDS. The DNA sequences of *dndB-dndE* genes from each pulsotype have been deposited in GenBank under accession nos. KJ702391-KJ702410. PCR detection of *iscS* revealed that all *E. coli* (both dnd^+ and dnd^- strains), *Salmonella* spp. and *Shigella* strains were positive for the *iscS* gene (Figure 5.3). In fact, *iscS* is also found in *Citrobacter* spp., *Klesiella* spp., and

Morganella spp. using BLASTN indicating that this cysteine desulfurase gene may be ubiquitous in different enterobacterial strains, thus supporting its suggested importance in sustaining fundamental life processes (Fontecave & Ollagnier-de-Choudens, 2008). However, this also indicates that the presence of the *iscS* gene is not a good genotypic marker for the *dnd*$^{+}$ phenotype. Representative *iscS* sequence obtained from this study have been deposited in GenBank under accession no KC839813.

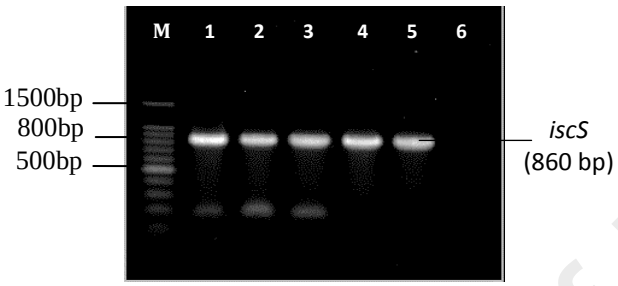


Figure 5.3: Representative picture of agarose gel of PCR amplified *iscS* gene. Lane M: 100 bp DNA ladder (Promega). Lanes 1-6: A1, EC001/09, *Shigella sonnei* TC8/00, *Salmonella* serovar Typhimurium STM 032, *Salmonella* serovar Enteriditis SE400/05, negative control.

5.3.3 *dnd* operons from different bacterial genera were grouped into distinct clusters

Using the three *dnd* operons (*E. coli* SE11, 55989 and B7A) as queries, a total of 106 positive hits (with BLASTN e-value = 0, identity ≥ 85% and coverage ≥98%) were obtained from both databases of non-redundant nucleotide collection (nr/nt) and whole genome shotgun contigs (WGS). All three queries also yielded the same number of positive hits including genomes from other species from *Enterobacteriaceae* besides *E. coli*, including *Serrati*a spp., *Erwinia* spp. and *Citrobacter* spp. Only 101 out of 106 *dnd* operons were included in the comparative sequence analysis because the remaining *dnd* operons were located in the gaps of their respective draft genomes and were thus, incomplete. The accession numbers of all 106 genomes with positive hits were summarized in Appendix V; c, which also included the *E. coli* strain background information.

A phylogenetic tree was constructed using the multiple sequence alignment of the nucleotide sequences of the 101 *dnd* operons (Figure 5.4). All *dnd* sequences from *Salmonella* spp. (blue) and *Erwinia amylovora* (green) were grouped into two distinct separate clusters. On the other hand, the *dnd* sequences from the genomes of other *Enterobacteriaceae* such as *Citrobacter* spp., *Cronobacter* spp., *Enterobacter* spp. and *Serratia* spp. were generally not clustered into any groups (Figure 5.4). As for *E. coli* (red), 45 out of 53 of them were grouped together into one main cluster. The remaining 8 were placed at different nodes. Another phylogenetic tree was constructed using the nucleotide sequences of 100 *dnd* operons (*E. coli* MS45-1 was excluded because the immediate vicinity of its *dnd* operon was located in the gaps, hence the total number of *E. coli* genomes in the analysis was reduced to 52) together with their respective immediate vicinity (1 kb) (Figure 5.5) to determine whether the immediate genetic environment will cause any changes to the phylogenetic tree. Surprisingly, both trees (Figures 5.4 and 5.5) yielded highly similar results with only slight differences. For example, *Cronobacter turicensis* z3032 is more closely related to *Enterobacter mori* LMG 25706 (both forms a sister group) than to *Dickeya dadantii* 3937 by comparing their *dnd* operons only (Figure 5.4). On the other hand, *Cronobacter turicensis* z3032 is more closely related to *Salmonella enterica* strains when the *dnd* operons and their respective immediate vicinity were included in the analysis (Figure 5.5). Overall, both trees yielded similar results.

Based on the sequence similarities of the immediate upstream and downstream regions (1 kb) of the 52 *dnd* operons of *E. coli*, a total of 7 subgroups were further identified. The 7 subgroups were designated a, b, cI, cII, d, e and f and were included in Figure 5.4. It was observed that genomes harbouring highly conserved *dnd* operons also shared similar immediate genetic environments and vice versa. Of the 7 subgroups,

groups cI and cII shared similar immediate downstream but different upstream regions while all other groups were very different with respect to their immediate genetic environment. Interestingly, 4 of the 52 *dnd* operons from *E. coli* genomes belonged to two *E. coli* draft genomes (KOEGE 118 and KTE 9). MSA analysis revealed that the two copies of the *dnd* operon and their associated genetic environment in the same *E. coli* genome are different and are distantly located from each other in the phylogenetic tree generated (groups cI and e) (Figures 5.4 and 5.5).

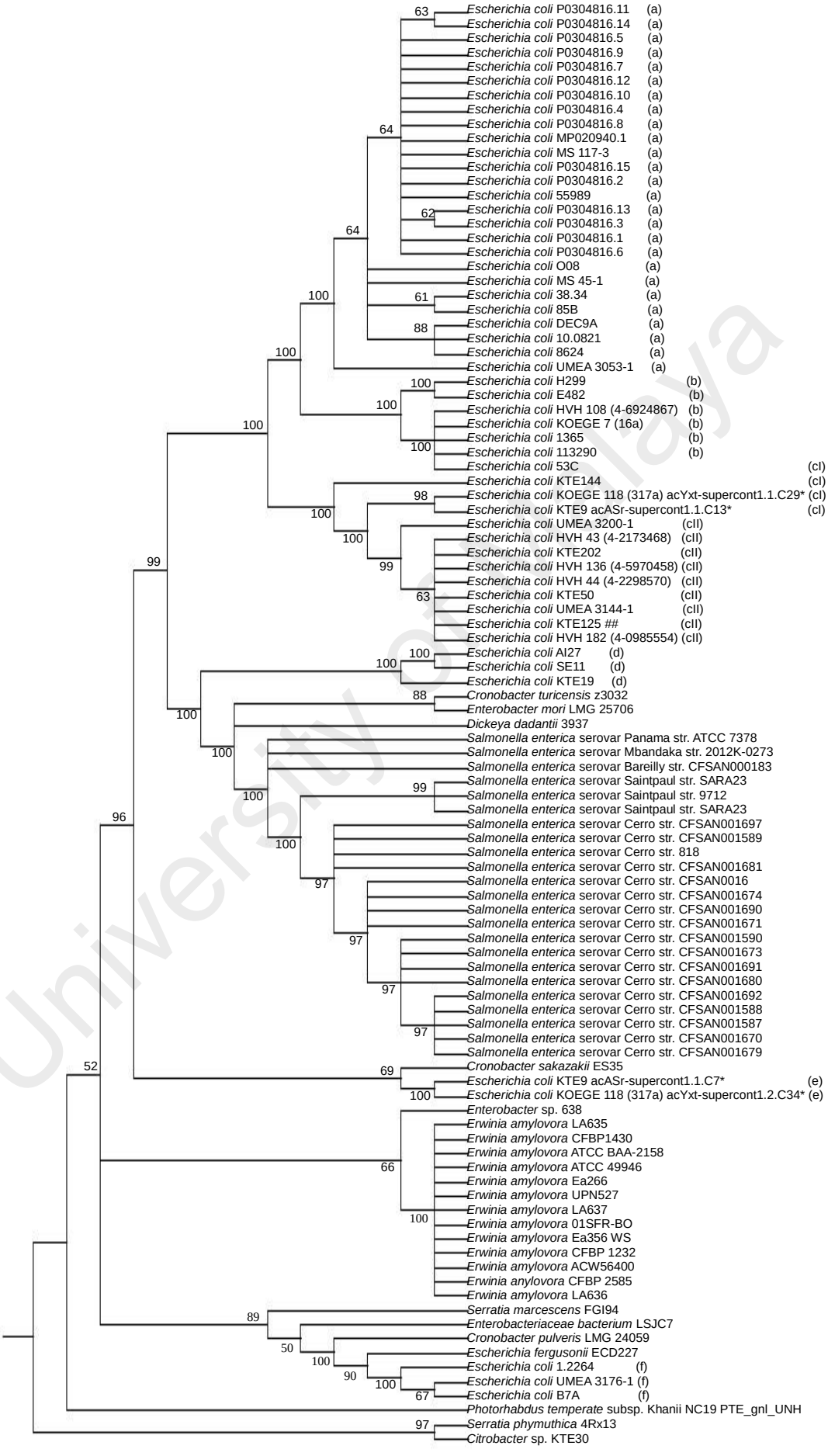


Figure 5.4: Phylogenetic tree of *dnd* operons from *Enterobacteriaceae*. Asterisk (*) indicates that the genome contains two copies of *dnd* operons. Different alphabets (a, b, cI, cII, d, e, f) depict the subgroups based on the immediate genetic environment. Colored genome strain codes facilitate visualization. Maximum likelihood (ML) method was used to construct the phylogenetic tree using MEGA5. Bootstrap confidence values greater than 50% are shown at branches. Nodes with less than 50% bootstrap value were collapsed.

Figure 5.5: Phylogenetic tree of *dnd* operons together with their respective immediate genetic environment (1kb) from *Enterobacteriaceae***.** Asterisks (*) indicate that the genome contains two copies of *dnd* operons. Different alphabets (a, b, cI, cII, d, e, f) depict the subgroups based on the immediate genetic environment. Maximum likelihood (ML) method was used to construct the phylogenetic tree using MEGA5. Bootstrap values have been calculated using 1000 replicates. Bootstrap confidence values greater than 50% are shown at branches. Nodes with less than 50% bootstrap value are collapsed.

5.3.4 Flanking regions of 12 *dnd*$^+$ *E. coli* strains resemble three different types of genetic environments

PCR detection of the immediate flanking regions of the *dnd* operon for the 12 *dnd*$^+$ *E. coli* strains obtained in this study revealed three types of immediate genetic environment. Strains S1-S7, B1 and B2 harboured *dnd* operons with associated genetic environment similar to group a; strains A1 and A2 to group cII; and strain A3 to group f. All three types of immediate genetic environments identified in the *dnd*$^+$ *E. coli* strains using PCR matched to those determined by *in silico* analysis, indicating that there were likely only 7 types of different immediate genetic environment identified to date for the *E. coli* *dnd* operon.

5.3.5 Diverse *dnd*-encoding genomic islands were observed for various representative *E. coli* genomes

A genome from each of the seven subgroups determined by the sequence similarity of their respective immediate genetic environment was selected for further genomic analysis. Only 1 genome was selected from each subgroup because it was impossible to determine all complete GIs carrying the *dnd* operon as most of the genomes in the public database were partially sequenced. All 7 *dnd* operons from the selected *E. coli* genomes were found to be associated with GIs. The 7 *dnd*-encoding GIs were then designated based on the naming system described in the Methodology section. Thus, the *dnd*-encoding GI located in *E. coli* 55989 (accession no. CU928145.2) was designated

"dndI_IEC 55989"; KOEGE 7 (16a) acYxM- supercont1.4.C20 (accession no. AWAA010 00020.1) as "dndI_IEC KOEGE 7"; KOEGE 118 (317a) acYxt-supercont1.1.C29 (accession no. AWAR01000029.1) as "dndI_IEC KOEGE 118"; UMEA 3200-1 acYwY- supercont1.1.C4 (accession no. AWCH0100 0004.1) as "dndI_IEC UMEA 3200-1"; SE11 (accession no. AP009240.1) as "dndI_IEC SE11", KTE9 acASr-supercont1.1.C7 (accession no. ANVJ010 00007.1) as "dndI_IIEC KTE 9" and UMEA 3176-1 acYyH- supercont1.8.C36 (accession no. AWCA01000036.1) as "dndI_IEC UMEA 3176-1". The genetic maps of the 7 types of *dnd*-encoding GIs are illustrated in Figure 5.6 Based on the genetic maps of the 7 *dnd*-encoding GIs, it was found that the genetic contents of the GIs harbouring the *dnd* operons were variable (Figure 5.6) although all the GIs described were obtained from *E. coli*. This indicates the potential contribution of GIs to the evolution and diversification of closely related bacteria. All 7 *dnd*-encoding GIs were also associated with tRNA and integrase genes (Figure 5.6). Further exploration of the genetic environment revealed the presence of three conserved genes (*dptF*, *dptG*, *dptH*), found upstream of the *dnd* operons, where both gene clusters were often separated by 1 to 3 hypothetical proteins (Figures 5.6, 5.7, 5.8 and 5.9). The three conserved genes were found in the vicinity of the *dnd* operon of all *E. coli* genomes included in the MSA analysis using BLASTN with the exception of *E. coli* strains C496_10 and O08 where only draft genomes were available. Thus, we could not be certain if these genes were absent or they were lost in the gaps of the incomplete genome sequences.

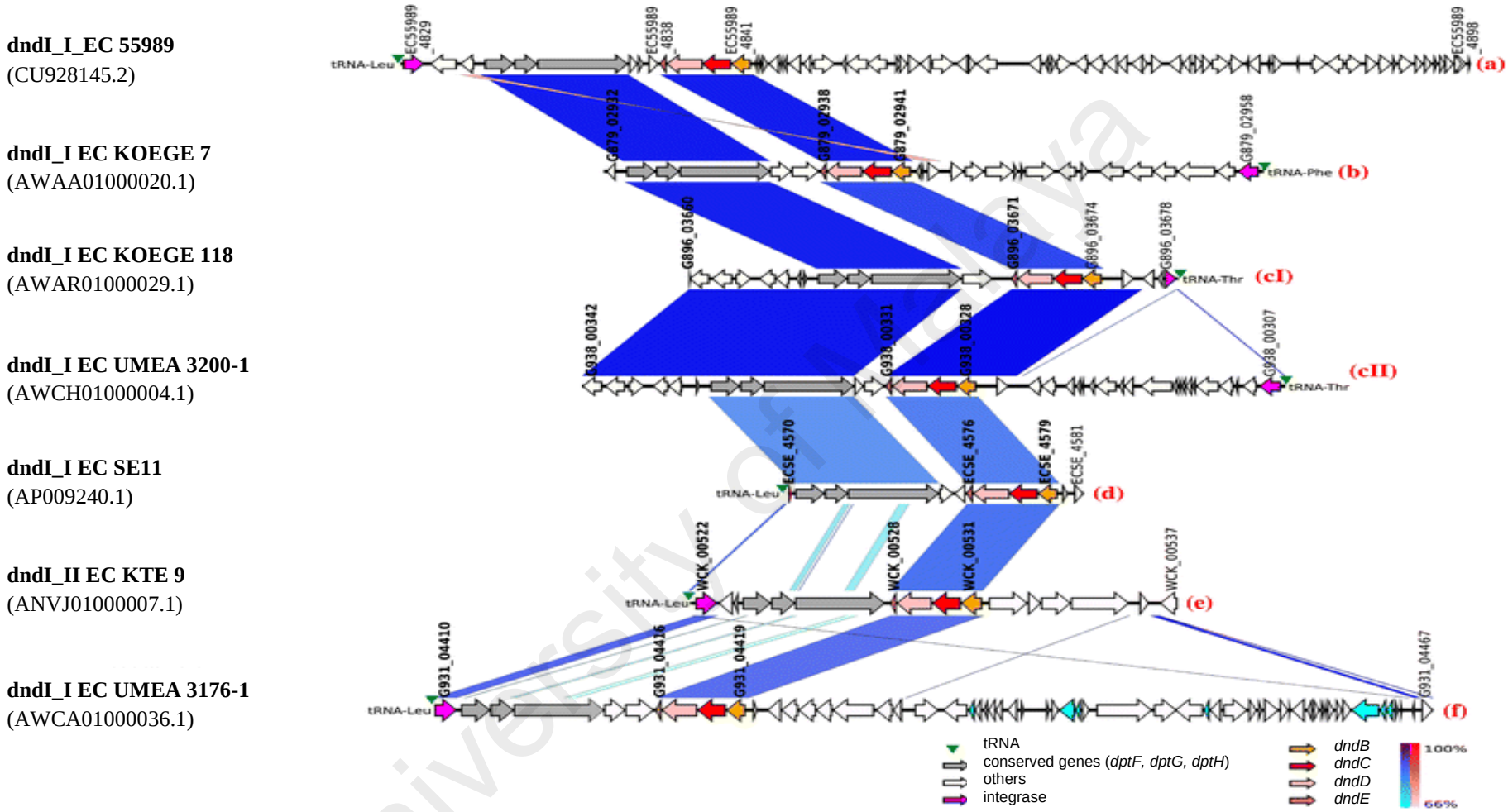


Figure 5.6: Genetic map of 7 different *dnd*-encoding genomic islands of *E. coli*. Same-strand DNA similarity is shaded blue (dark blue-sky blue) while reverse similarity is shaded red (red-light pink). Coding sequences are displayed as arrows. Major features are displayed and are colored to facilitate visualization. Order of genomic islands in the map was arranged according to the phylogenetic tree constructed using the *dnd* operons. Red alphabet in brackets indicates the subgroup determined based on the immediate genetic environment of the respective *dnd* operons. Integrase gene is the first element of the *dnd*-encoding genomic island (GI).

E. coli KOEGE 118 and KTE 9 harboured 2 copies of the *dnd* operons each, and all were associated with GIs as well. The *dnd*-encoding GIs in KOEGE 118 were designated dndI_IEC KOEGE 118 (acYxt-supercont1.1.C29; accession no, AWAR01000029.1) and dndI_IIEC KOEGE 118 (acYxt-supercont1.2.C34; AWAR01000034.1) whereas in KTE9, they were designated dndI_IEC KTE 9 (acASr- supercont1.1.C13; ANVJ01000013.1) and dndI_IIEC KTE 9 (acASr-supercont1.1.C7; ANVJ01000007.1) with their genetic maps depicted in Figure 5.7. Interestingly, both dndI_IEC KOEGE 118 and dndI_IEC KTE9 are identical whereas dndI-IIEC KTE9 is identical to dndI-IIEC KOEGE 118 (covering the *dnd* operon and its immediate vicinity) but dndI-IIEC KOEGE 118 had an additional 74 kb segment containing a type VI secretion system (T6SS) that was absent from dndI-IIEC KTE9 (Figure 5.7). Thus, in terms of the *dnd* operon and its immediate genetic vicinity, both dndI_IIEC KOEGE 118 and dndI_IIEC KTE 9 were grouped within the same group e (Figures 5.4 and 5.5). It should be noted that not all genomes from the same subgroup shared identical GIs as they are often very large in size, and variation may have occurred within the GIs caused by other mobile genetic elements (Bellanger et al., 2014).

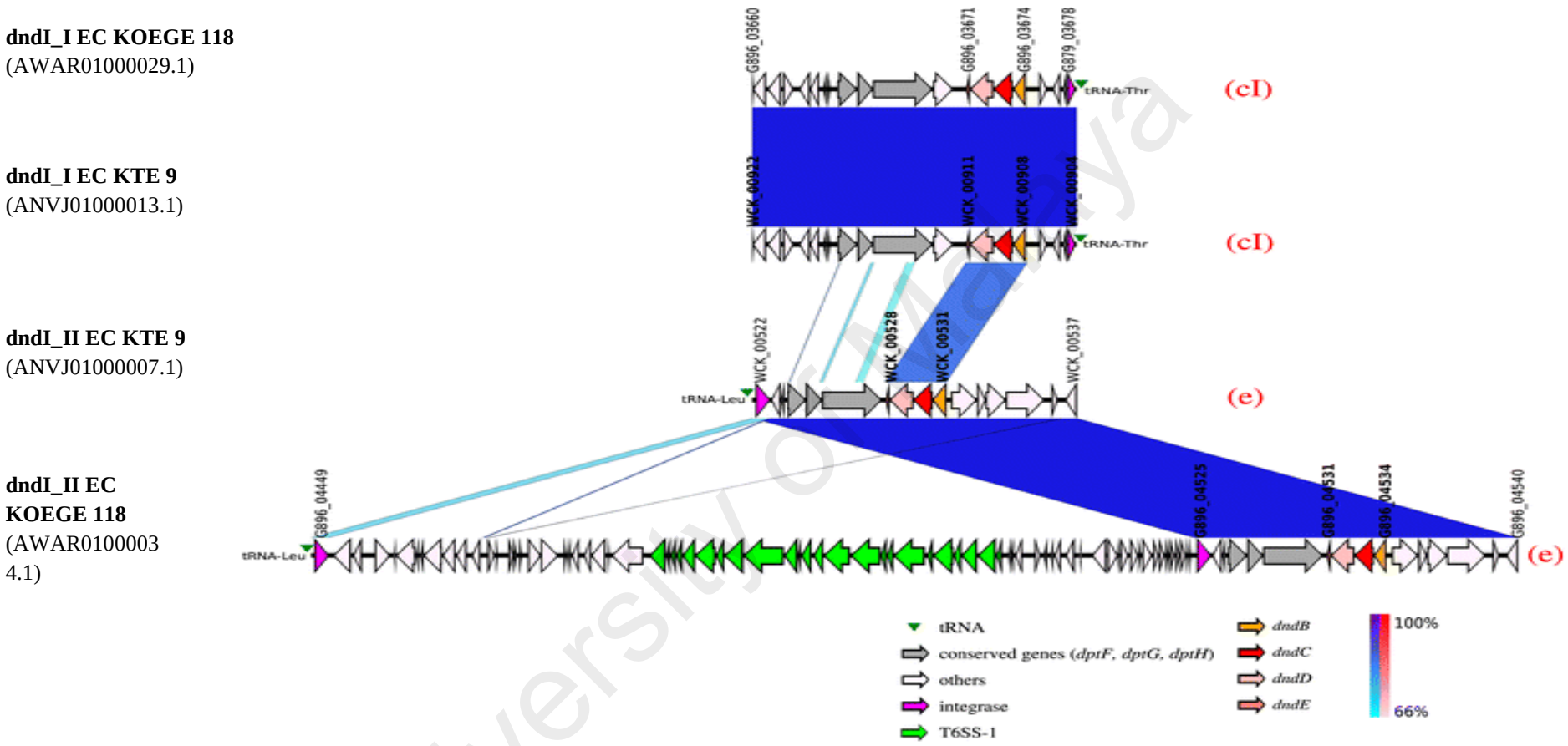


Figure 5.7: Genetic map of the *dnd*-encoding genomic islands of *E. coli* KTE9 and KOEGE118, which harboured 2 copies of *dnd* operons. The integrase gene is the first element of the *dnd*-encoding genomic island (GI) although the genetic map is illustrated according to the orientation of the *dnd* operons to facilitate visualization. Same-strand DNA similarity is shaded blue (dark blue-sky blue) while reverse similarity is shaded red (red-light pink). Coding sequences are displayed as arrows. Major features are displayed and are colored to facilitate visualization. Red alphabet in brackets indicates the subgroup determined based on the immediate genetic environment of the respective *dnd* operons. An additional DNA segment of 74 kb was observed for dndI_IIEC KOEGE 118 compared to dndI_IIEC KTE 9. dndI_IEC KOEGE 118 is identical to dndI_IEC KTE 9.

5.3.6 *dnd*-encoding genomic islands displayed mosaic structures

Due to the great diversity of the islands harbouring *dnd* operons, three representative islands from *E. coli* strains B7A, 55989 and KOEGE 118 belonging to different groups (Figure 5.4) were selected for further investigation to determine the possible evolutionary forces driving the great variation between the *dnd*-encoding GIs. Based on the GC content (Figure 5.8), the *dnd*-encoding GIs displayed mosaic structures and can be separated into several modules. Further examination of the presence of ORFs from the *dnd*-encoding GIs in other *E. coli* genomes further supported their mosaic characteristics. The *dnd* modification systems, *dpt* restriction systems, together with the hypothetical proteins in between the two systems were collectively found associated with *dnd*-positive genomes only (ORFs underlined by green-coloured bars, Figure 5.8). On the other hand, majority of the ORFs outside the *dpt*- and *dnd*-clusters are present in both *dnd*-positive and *dnd*-negative *E. coli* genomes (brown-coloured underlined bars, Figure 5.8), with the exception of an integrase gene (*E. coli* B7A) and 6 ORFs located immediately upstream of the *dnd* operon in KOEGE118, which are found associated with *dnd*-positive genomes only. Intriguingly, some hypothetical proteins are strain-specific, which can be observed in *E. coli* strain 55989 (yellow-coloured bar) (Figure 5.8).

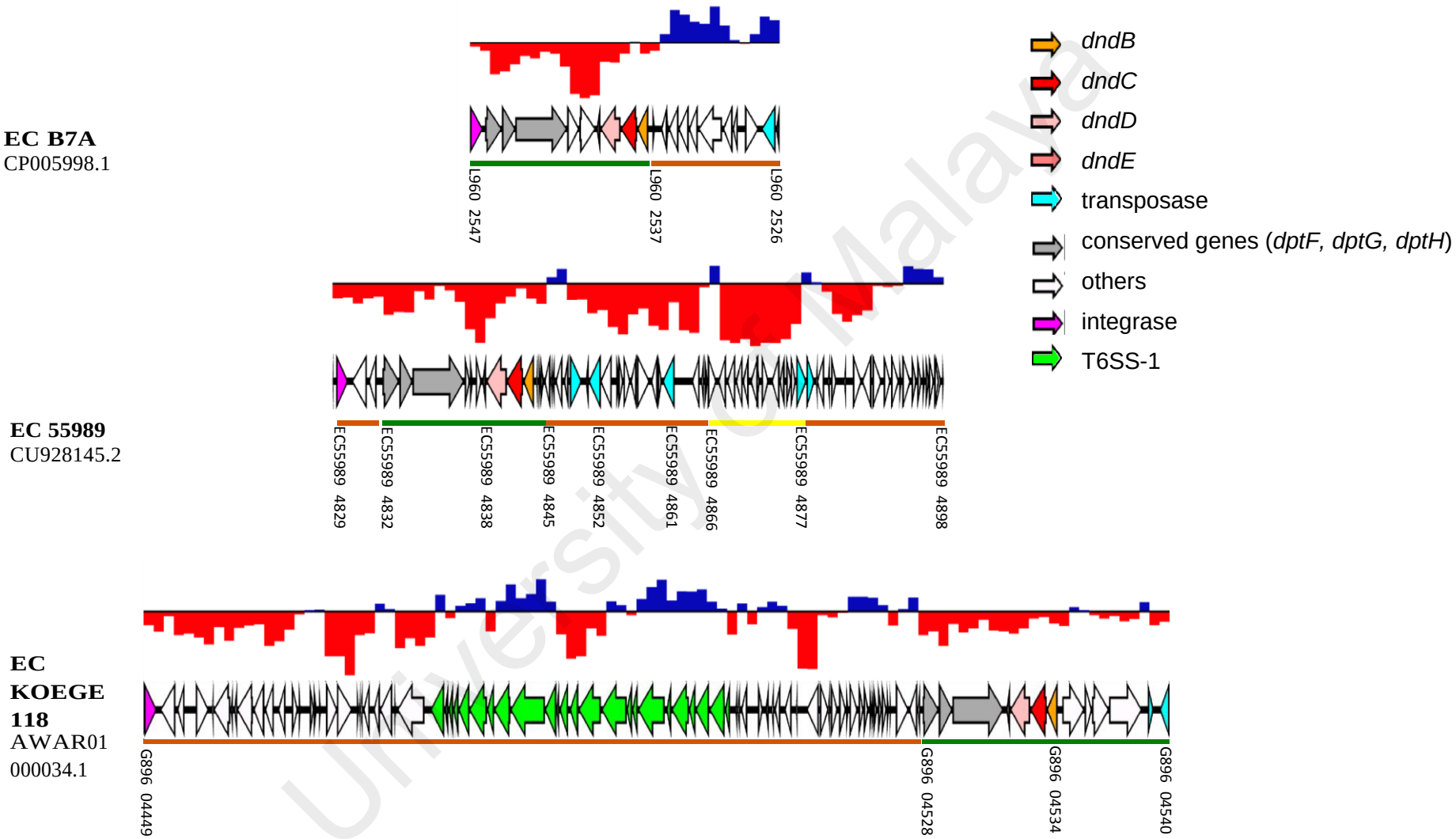


Figure 5.8: GC content of three representative *dnd*-encoding genomic islands. GC content of *dnd*-encoding GIs are shown above their respective genetic maps, with blue and red region showing GC content above and below 50%, respectively. The *dnd*-encoding GIs consist of several modules, depicted by the coloured bars below their respective genetic maps. ORFs above the brown coloured bars were found in both *dnd*-positive and *dnd*-negative *E. coli* genomes. ORFs that are underlined by green bars are strictly associated with *dnd*-positive genomes only while ORFs underlined by the yellow bar are strain-specific.

5.3.7 Different genetic contents were identified at the *leu* tRNA insertion sites of *dnd*-negative and *dnd*-positive *E. coli* genomes

All *dnd*-encoding GIs determined in our study were associated with a tRNA-encoding gene, an integration hotspot in bacterial genomes (Ou et al., 2006). This prompted us to investigate the genetic contents at the same tRNA gene site (*leu* tRNA) in *dnd*-negative *E. coli* genomes. The tRNA insertion sites of four genomic regions (1 *dnd*-negative and 3 *dnd*-positive) from *E. coli* strains CFT073, UMEA 3176-1, B7A and 55989 were compared (Figure 5.9). Upstream regions of the GIs that were inserted into the *leu* tRNA gene sites for all 4 *E. coli* genomes were highly conserved, indicating similar upstream backbone (Figure 5.9). While a 15.6 kb non-*dnd*-encoding GI (consisting mainly of hypothetical proteins) was found inserted at the *leu* tRNA gene site of *E. coli* CFT073, similar insertion sites in *E. coli* strains B7A, UMEA 3176-1 and 55989 were occupied with *dnd*-encoding islands of variable sizes (53.3 kb – 59.5 kb). The downstream regions of the genomic islands for *dnd*-negative (*E. coli* CFT073) and *dnd*-positive (*E. coli* UMEA 3176-1 and B7A) genomes were also conserved (Figure 5.9), indicating that the *leu* tRNA gene site in *E. coli* is likely a variable region with other types of GIs found in *dnd*-negative strains. Despite the highly conserved upstream and downstream flanking regions for *dnd*-encoding GIs for *E. coli* strain UMEA 3176-1 and B7A, slight variations within the highly similar GIs were still observed with transposases found in the variable regions (Figure 5.9).

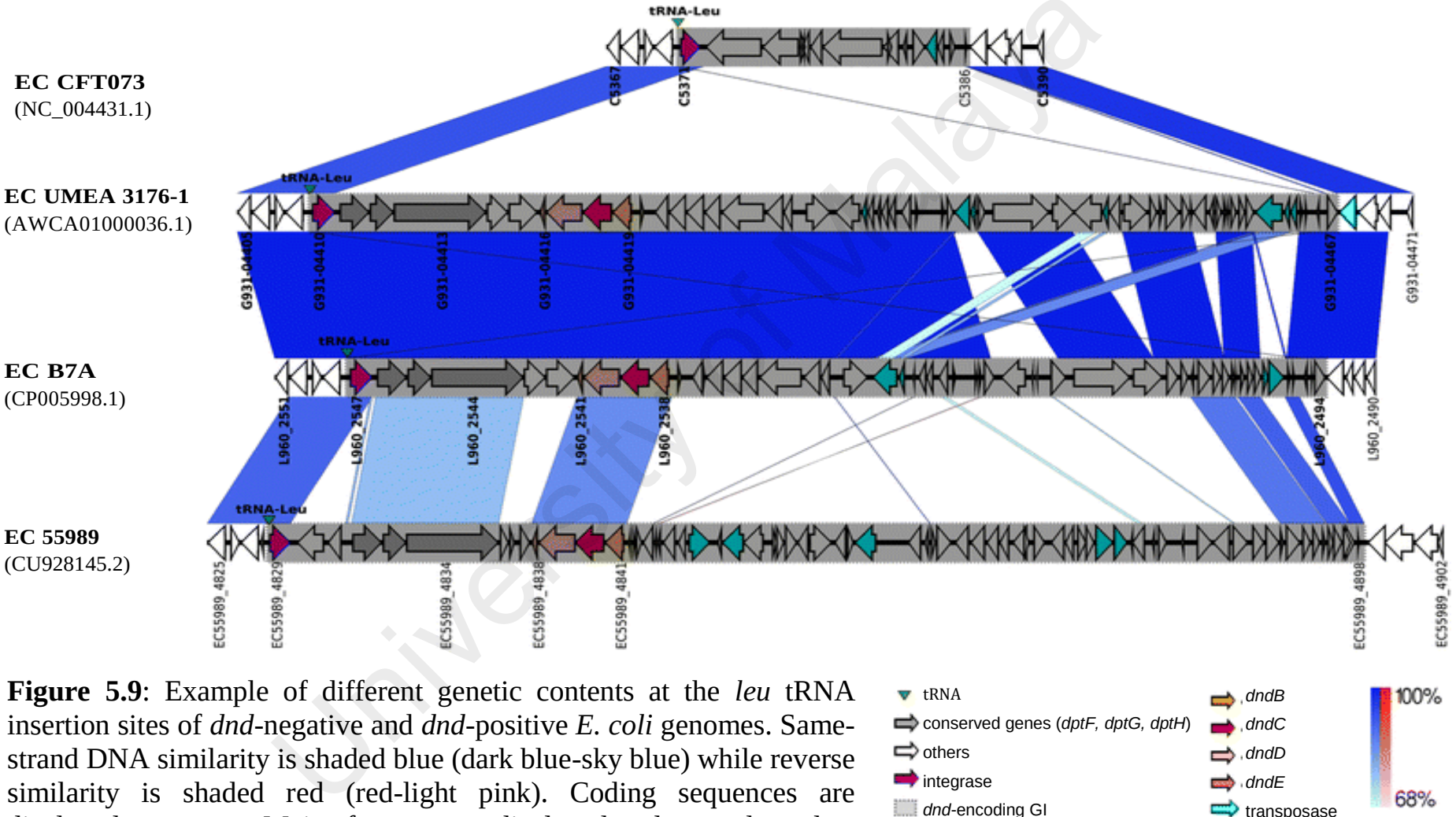


Figure 5.9: Example of different genetic contents at the *leu* tRNA insertion sites of *dnd*-negative and *dnd*-positive *E. coli* genomes. Same-strand DNA similarity is shaded blue (dark blue-sky blue) while reverse similarity is shaded red (red-light pink). Coding sequences are displayed as arrows. Major features are displayed and are coloured to facilitate visualization.

5.4 DISCUSSION

Although PFGE is the method of choice for subtyping *E. coli*, certain Dnd$^+$ *E. coli* strains are untypeable and this yield incomplete molecular epidemiological data. The addition of thiourea to the PFGE running buffer has resolved the untypeable problem for all the 12 Dnd$^+$ *E. coli* strains. In this study, 9 out of the 12 Dnd$^+$ strains were represented by two distinctly different pulsotypes [S1-S7 (n = 7); A1-A2 (n = 2)] (Figure 5.1), suggesting that the Dnd$^+$ phenotype might be an inherent clonal trait for strains with the two respective pulsotypes. No Dnd$^-$ strains were found to share similar pulsotypes with the Dnd$^+$ strains when we compared the pulsotypes with those of previous chapters (Chapters 3 and 4). The 9 Dnd$^+$ strains were multidrug resistant with 2 extended-spectrum beta-lactamase (ESBL) producers whereas 7 of the 9 strains were pathogenic VTEC and were clonal (Chapter 3; Ho et al., 2013). The resistance and pathogenic states of these Dnd$^+$ strains may therefore pose possible threats to public health, hence improvement in their typeability is important to generate valuable epidemiological data. The pathogenic state of the Dnd$^+$ strains also prompted us to further explore the strain background of all 52 *E. coli* genomes that carried the *dnd* operons. Surprisingly, majority of them (n = 44 of 52; 85%) were pathogenic in nature with 26 being diarrheagenic *E. coli* and 18 extra-intestinal pathogenic *E. coli*. To the best of our knowledge, no study has reported this finding before and the link between virulence and *dnd* operons has yet to be explored. Undoubtedly, bias might be present as pathogenic *E. coli* strains have higher level of public interest, and hence the availability of their genome data may be plenty when compared to non-pathogenic *E. coli*. Nevertheless, this finding may provide a caveat for future studies on the possible interesting relationship between pathogenicity and the presence of the *dnd* operon in *E. coli*.

The *dnd* clusters of different bacterial species have reportedly a diverse nature (Ou et al., 2009), not only with regard to their respective sequence similarities, but also the genetic context of the *dnd* clusters. Some bacterial species harboured a 4-gene *dndBCDE* operon whereas others carried a five-gene *dnd* cluster (*dndA* and *dndBCDE* operon). In fact, *dndA* is not found in the *dnd* clusters of all *Enterobacteriaceae* reported in dndDB (including *E. coli*), indicating that the gene contents in the *dnd* clusters for closely related bacteria are similar. This is also the reason why *dndA* gene was not included as a target gene for prediction of degradation phenotype of *E. coli* in our study. Although DndA might be absent in some *dnd* clusters, studies have revealed that IscA can serve as a functional homologue of DndA and may be found in more than one copy in the genome (He et al., 2007; Wang et al., 2011). As previously reported (Ou et al., 2009), the *dnd* cluster of *E. coli* belonged to the 4-gene *dndBCDE* operon. All 12 Dnd+ *E. coli* strains in this study were positive for the *dndB-dndE* genes and the presence of *dndB- dndE* genes was significantly associated with the DNA degradation phenotypes ($p < 0.5$). This indicates that the presence of these genes can be a potential genetic marker to predict the phenotype of DNA degradation. Howard et al. (2013) developed an assay to predict the DNA degradation phenotype of *Mycobacterium abscessus* based on the fact that the *dnd* gene cluster was associated with a genomic island of the *M. abscessus* complex and the genetic environment of the *dnd*-encoding GI was highly conserved. One pair of primers targeted the *dndC* gene while another pair confirmed the presence or absence of the *dnd* operon. If the *dnd* operon is present, the amplicon will be too large to be amplified. On the other hand, if the *dnd*-encoding GI is absent, a short bridge amplicon spanning the GI insertion site will be generated for Dnd$^-$ *M. abscessus* strains (Howard et al., 2013). However, this approach was not applicable in this study due to the variable genetic environment of the *E. coli dnd* operons based on MSA analysis. Hence for *E. coli*, the detection of all 4 *dndB-dndE* genes would be a better

alternative to confirm the presence of the *dnd* operon.

The dissemination of *dnd* operons was reportedly facilitated by mobile genetic elements especially GIs (He et al., 2007; Ou et al., 2009). Hence, representative *E. coli* genomes harbouring *dnd* operons were subjected to genomic island determination. Our data showed that all the *dnd* operons in the selected genomes were in fact located in GIs. Out of 31 *dnd* operons described in dndDB, 29 were also reported to be located in chromosomal islands, one was plasmid-encoded while another one is located in a plasmid-derived chromosomal segment. These results showed the potential role of genomic islands in facilitating the horizontal gene transfer of *dnd* operons. Genomic islands have also been known to be the mechanism of diversification contributing to bacterial evolution (Juhas et al., 2009). However, it should be noted that majority of the genomes of *E. coli* in this study were draft genomes. Thus, we were unable to ascertain if the *dnd*- encoding GIs identified were chromosomally located or plasmid-encoded. Further analysis would be necessary to verify their location.

The genetic environment of the *dnd* operon in *E. coli* is very diverse (Figures 5.5 – 5.9). On the other hand, the *dnd* operon together along with its immediate vicinity in *Erwinia amylovora* and *Salmonella enterica* appeared to be conserved within the same bacterial species (Figures 5.4 and 5.5), indicating that their genetic environment could be similar. Analysis of representative *E. coli* genomes has shown that their respective *dnd* operons were associated with genomic islands that possess diverse genetic context. Different genetic fragments of the island that corresponded to regions of different types of *E. coli* genomes (*dnd*-positive or *dnd*-negative) indicate that the genetic elements were likely acquired from different origins. Genetic variation can also be observed among highly conserved *dnd*-encoding GIs due to insertion or deletion of transposases along with other hypothetical proteins (Figure 5.9). A large T6SS cluster was also found

inserted into the upstream of another *dnd*-encoding island (Figure 5.7). Furthermore, marked difference in GC content was also observed across the *dnd*-encoding GIs, supporting the mosaic composition of these *dnd*-containing GIs. The mosaic structure of the *dnd*-encoding GIs (Figure 5.8), presence of phage-like integrases and transposable elements flanking the GIs as well as across the *dnd*-encoding GIs (Figure 5.9), suggest multiple integration or recombination events, which has also been reported by He et al. (2007) for the *dnd*-encoding GI of *Streptomyces lividans* 66. Insertion and deletion of genetic elements as well as recombination appeared to play an essential role in the constant evolution of these GIs.

Besides the diverse genetic context of the *dnd*-encoding GIs of *E. coli*, the tRNA genes associated with these islands can also be variable. The 3′-end of tRNA genes frequently served as the insertion sites for foreign DNA fragments (Ou et al., 2006). All 9 *dnd*-encoding GIs described in this study were associated with tRNA genes, but only limited to three types (*leu*, *thr* and *phe* tRNA genes) (Figures 5.6 and 5.7), which have also been reported to be among the most frequently targeted tRNA genes for insertion in *E. coli* (Germon et al., 2007; Ou et al., 2006). One of the possible reasons that particular tRNA genes often served as an integration hotspot is that the integrases of GIs may have specificity for certain tRNA genes (Boyd et al., 2009). For *E. coli* KOEGE 118 and KTE 9 which harboured two *dnd*-encoding GIs within the same genome, their respective *dnd*-encoding GIs were also associated with different tRNA genes (*leu* and *thr* tRNA genes) (Figure 5.7). However, it should be noted that not all GIs are associated with tRNA genes (Langille et al., 2012).

This is the first report describing two *dnd* operons found in the same bacterial genome. However, the benefits to its host, if any, are unknown. Several advantages to bacteria with DNA phosphorothioation modification have been proposed. One of them

is that the modified host DNA is able to act as an antioxidant as phosphorothioation gave the modified DNA chemical reduction property. Thus, the modified DNA is resistant to growth inhibition and oxidative double-stranded DNA breakage caused by H_2O_2 (Xie et al., 2012). H_2O_2 is generated during aerobic metabolism and studies have revealed that aerobic *E. coli* can generate high levels of H_2O_2 that are toxic enough to cause damage to their own DNA. On the other hand, exogenous H_2O_2 can also be found in the environment as competing organisms can release H_2O_2 as toxin to suppress the growth of competitors (Park et al., 2005). Bacteria can be killed by high levels of H_2O_2 generated endogenously and exogenously. Hence it is essential to protect the host bacteria against damage caused by peroxide and DNA that has been modified by phosphorothioation was hypothesized as one of the mechanisms that can offer protection to the host (Park et al., 2005). Nevertheless, the biological implication of having two copies of the *dnd* operons within a genome is currently unknown.

All the *dnd*-encoding GIs identified in this study also harboured *dptFGH*, a set of three conserved *dnd*-linked orthologues which are often found near the *dnd* operon. *E. coli* KTE9 and KOEGE118 which harboured 2 *dnd*-encoding GIs in the same genome also carried the three conserved genes within the islands. The conserved *dnd*-linked orthologues have been shown to function as restriction systems in *Salmonella* enterica serovar Cerro87 (Xu et al., 2010) and recently, in *E. coli* B7A (Cao et al., 2014). The restriction system is able to restrict foreign DNA (such as phages and plasmids) that lacked the specific phosphorothioation modification when the host DNA is protected by the modification, and hence prevents the invasion of heterologous DNA (Xu et al., 2010). Phosphorothioate DNA modifications incorporated by Dnd proteins (DndABCDE) that work together with the DptFGH-restriction system function as a restriction-modification (R-M) system, which is evident in *E. coli* B7A (Cao et al., 2014). However, the same study also ruled out other known R-M mechanisms, namely

the types I, II, III and IV R-M systems (Cao et al., 2014), implying that the Dpt-Dnd system is likely a new R-M-type mechanism. These two systems which are located in the same genomic island provide selective advantages to the bacterial host and hence may be stably maintained (Howard et al., 2013; Wang et al., 2011). Nevertheless, the three conserved *dnd*-linked orthologues were not identified in all *dnd*-encoding GIs described in previous studies (He et al., 2007; Ou et al., 2009) but only in several diverse bacterial strains such as *E. coli*, *Hahelia chejuensis*, *Oceanobacter* sp., and *Bacillus cereus*. This serves to underline not only the diversity of the genetic environment surrounding the various *dnd* operons but also their possible roles in their respective hosts. Perhaps they are like toxin-antitoxin systems which have multiple biological functions depending on their location in the genome and their respective hosts (Chan et al., 2012) which have adopted and adapted these genetic modules for cellular function. Intriguingly, previously described *dnd*-encoding GI of *Geobacter* uraniireducens Rf4 was also found to carry putative toxin-antitoxin loci (Shao et al., 2011). The presence of the *dnd* operons in such varied genomic islands in *E. coli* and their biological implication will hopefully be revealed in the near future.

5.5 CONCLUSION

Oftentimes, PFGE has to be rerun in cases where DNA degradation is observed for untypeable strains, to confirm the non-typeability of Dnd^+ strains, which is time consuming. The designed PCR assay coupled with the use of thiourea in PFGE may improve the detection and typeability of Dnd^+ *E. coli* strains simultaneously. Our study also showed that genomic islands not only play a potential pivotal role in facilitating the horizontal gene transfer of *dnd* operons, but also essential in generating diversity within the genomics of closely related bacteria exhibiting degradation phenotypes. Since the *dnd* operon is often associated with mobile genetic elements, the operon may be integrated into genetically diverse strains followed by dissemination. DNA that has been modified by phosphorothioation was also hypothesized as one of the mechanisms that can offer protection to the host against damage caused by peroxide. Therefore the Dnd^+ phenotype can be observed in both genetically closely related and diverse strains, and may be stably maintained in the Dnd^+ strains due to its possibility in giving a selective advantage to the host. The finding that *dnd* operons were more often found in pathogenic *E. coli* should also be further investigated for their possible linkage.

CHAPTER 6: WHOLE GENOME SEQUENCE ANALYSIS OF TWO CLONALLY-RELATED *ESCHERCHIA COLI* STRAINS EC096/10 AND EC302/04 ISOLATED FROM INTESTINAL AND EXTRAINTESTINAL SITES

6.1 INTRODUCTION

Two clinical *E. coli* isolates (EC096/10 and EC302/04) that were characterized in Chapter 3 were found to be closely related using PFGE as typing tool, where only 1 band difference was observed for their respective pulsotypes. EC096/10 was isolated from a stool specimen (intestinal site) whereas EC302/04 was isolated from a tracheal aspirate (extraintestinal site) over a gap of 5 years from the same tertiary hospital in Malaysia. The encounter of two clonal *E. coli* strains obtained from different sites in the human host(s) is indeed interesting as the intestinal and extraintestinal environment in human hosts were known to be different especially with regards to the iron and nitrogen concentration (Hagan et al., 2010). To better understand the biology of these two closely related *E. coli* strains EC096/10 and EC302/04, whole genome sequencing was applied. Comparative genomic analyses, *in silico* phylogenetic grouping, MLST and virulence profiling of additional 38 *E. coli* genomes of different pathotypes that were extracted from the public database were also determined. Besides highlighting the genetic complexity of ExPEC strains, our study also uncovered a potential sporadic *E. coli* clone that was often associated with extraintestinal infections, where their serotypes, resistotypes and plasmid replicon types were further investigated and compared.

6.2 METHODOLOGY

6.2.1 Bacterial strains

Two *E. coli* strains EC302/04 and EC096/10 were obtained from the same tertiary hospital in the state of Johor, Malaysia, over a gap of 5 years. *E. coli* EC302/04 is a multidrug-resistant strain that was isolated from the tracheal aspirate of an ICU patient in year 2004 (Lim et al., 2009) while EC096/10 was isolated from the stool sample of a pediatric patient in year 2010 (Ho et al., 2012). Both patients were admitted to the same hospital with an unknown epidemiological link. Using PFGE, These two strains were found to be highly related with only a single band difference (Chapter 3).

6.2.2 Whole genome sequence analysis

6.2.2.1 Genome sequencing, assembly, annotation and analysis of EC302/04 and EC096/10

Whole genome sequencing of EC302/04 and EC096/10 was performed using the Illumina HiSeq 2000 (100-bp read length) with an insert size of 300 bp. The reads were trimmed and assembled *de novo* using CLC Genomics workbench 5.0 (CLC Bio, Denmark). The genome annotation of EC302/04 and EC096/10 was performed by the RAST annotation server (Aziz et al., 2008). Number of tRNAs and rRNAs were identified using tRNAscan-SE and RNAmmer, respectively (Lagesen et al., 2007; Lowe & Eddy, 1997). Virulence genes were identified by mapping the annotated sequences against the Virulence Factor Database (VFDB) (Chen et al., 2012) using BLASTn. Pathogenecity islands (PAI) were determined by local BLAST (Altschul et al., 1990) against the PAI database retrieved from PAIDB (Yoon et al., 2015). BLAST hits with e-value ≤1e–10, query coverage ≥80% and nucleotide identity ≥80% were considered as positive hits. All results generated using webservers were further verified using BLASTn. The prediction of a type III secretion effector was carried out at the SIEVE

server (Samudrala et al., 2009). The single nucleotide polymorphisms (SNPs) from core regions (synonymous and non-synonymous mutations) of EC302/04 and EC096/10 were also determined using PGAP (Zhao et al., 2012).

6.2.3 Comparative genomic analysis of EC96/10 and EC302/04 with 38 *E. coli* genomes obtained from NCBI database

6.2.3.1 Phylogenomic analysis of *E. coli* strains from various sources

Whole genome sequences of 24 pathogenic *E. coli* (including IPEC and ExPEC) and three intestinal *E. coli* strains sharing the highest whole genome nucleotide sequence homology (E-value =0; Query coverage= 52 – 98%; nucleotide identity: 98%) with EC096/10 and EC302/04 determined using BLASTn (Altschul et al., 1997) were extracted from the NCBI database. An 'intestinal strain' is defined as a strain isolated from the intestinal tract of a patient but does not belong to any IPEC pathotype. Another 11 commensals that were commonly used for *E. coli* comparative genomics analysis were also included to represent the sequence diversity of *E. coli* (Bertels et al., 2014; Lukjancenko et al., 2010; Touchon et al., 2009). In total, an additional 38 whole genome sequences of different *E. coli* pathotypes were included in the phylogenomic analysis. The phylogenomic tree of 40 *E. coli* genomes (including EC302/04 and EC096/10) was constructed by mapping against six reference genomes with REALPHY (Reference sequence Alignment-based Phylogeny builder) (Bertels et al., 2014) (Table 6.1); an approach which is able to construct an accurate tree including genomes from substantially divergent bacterial strains. The six reference genomes are genomes that were evenly distributed across the tree. The merged alignments were then used to generate a phylogenetic tree with 100 bootstrap replicates using PHYML (Guindon et al., 2010).

Table 6.1: Genomes of 40 *E. coli* used for comparative genomic analysis

***E. coli* genome**	**Accession number**	***E. coli* genome**	**Accession number**
536	CP000247	HS	CP000802
12009	AP010958	IAI 1	CU928160
55989*	CU928145	IAI 39*	CU928164
042	FN554766	IHE3034	CP001969
2009EL-2050	CP003297	KTE147	ANWN00000000.1
2009EL-2071	CP003301	MC6003	ATNV00000000.1
5-366-08_S1_C3	JONE00000000.1	MG1655*	U00096
AB42410445-isolate1	ATOD00000000.1	NA114*	CP002797
AB42602061-isolate1	ATOB00000000.1	PMV_1	HG428755
APEC O78	CP004009	REL606	CP000819
ATCC8739	CP000946	RM12579	CP003109
BL21(DE3)	CP001509	S88	CU928161
CB9615	CP001846	SE11	AP009240
CE10	CP003034	SE15	AP009378
CFT073	AE014075	SMS-3-5	CP000970
clone D i2	CP002211	TW14359	CP001368
EC096/10	AONF00000000	UMN026*	CU928163
EC302/04*	AMFM00000000	UTI89	CP000243
ED1a	CU928162	W (ATCC9637)	CP002185
HM46	APNY00000000.1	W3110	AP009048

* reference genomes used in constructing phylogenetic tree

6.2.3.2 *In-silico* identification of multi-locus sequence type, phylogenetic group, virulence gene profile, resistotype, serotype and CRISPR region

Seven housekeeping gene sequences [*adk* (adenylate kinase), *fumC* (fumerase isozyme C), *gyrB* (DNA gyrase subunit B), *icd* (isocitrate hydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate synthetase), and *recA* (recombinase A)] of the 40 *E. coli* genomes listed in Table 6.1 (including EC302/04 and EC096/10) were extracted and then assigned to a specific allelic type and sequence type (ST) (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/) (Wirth et al., 2006). All *E. coli* genomes in the GenBank databases (nr/nt and WGS) were further screened for the presence of ST349, ST2914 and ST70 using BLASTn (accessed date: November, 2014). *In-silico* phylo-grouping for all studied genomes was performed based on quadruplex phylo-group assignment method (Clermont et al., 2013). The virulence profiles of 40 *E. coli* genomes were determined with BLASTn using 142 nucleotide sequences of extraintestinal-associated virulence factors obtained from a previous study (Salipante et al., 2015) as query sequences (detailed in Appendix V; d). These virulence factors include adhesins, invasins, iron acquisition systems, toxins, protectins and others. Genomes with matching BLAST hits with e-value ≤1e–10, query coverage ≥80% and nucleotide identity ≥80% were considered as positive hits.

In silico serotyping was also carried out for both EC302/04 and EC096/10 and closely related strains (ie. KTE146, HM46, 5-366-08_S1_C3, AB42410445-isolate1, AB42602061-isolate1 and MC6003) using SeroTypeFinder (https://cge.cbs.dtu.dk/services/SerotypeFinder/) (Joensen et al., 2015). The resistance genes, CRISPR clusters and plasmid replicon(s) for strains EC302/04, EC096/10, KTE146, HM46, 5-366-08_S1_C3 were then determined using ARDB and ResFinder webservers (Liu & Pop, 2009; Zankari et al., 2012), CRISPRFinder (Grissa et al., 2007)

and PlasmidFinder (Carattoli et al., 2014) with default parameters, respectively.

6.2.4 Nucleotide sequence accession number

Whole genome sequence data of *E. coli* EC302/04 and EC096/10 were deposited in GenBank with the accession numbers AMFM00000000 and AONF00000000, respectively. The GenBank accession numbers for the remaining 38 *E. coli* genomes included in this study are listed in Table 6.1.

6.2.5 Sequence analysis of plasmid pEC302/04 gap closing, annotation and analysis

Putative plasmid pEC302/04 contigs were determined by subjecting the assembled contigs of EC302/04 to BLASTN (http://blast.ncbi.nlm.nih.gov/) against the GenBank non-redundant nucleotide collection (nr/nt). Two plasmids sequences of *E. coli* (pEC_L46 and pEC_L8) and a genome sequence of *Salmonella enterica* serovar Typhimurium (T000240) showing the highest coverage and identities with the pEC302/04 contig sequences were used as the reference sequences for plasmid gap closing. Bacterial genomes finishing tool (Galardini et al., 2011) was used for mapping the contigs over the reference genomes and primers for PCR were generated for gap closing (Appendix III; k). PCR-amplified products were purified and submitted to a commercial facility (First BASE Laboratories) for conventional Sanger dideoxy sequencing. Toxin-antitoxin systems were identified using TAfinder using default parameters (http://202.120.12.133/ TAfinder/index.php) and the identified toxins and antitoxins were further validated using BLASTp.

Virulence genes and resistance genes and in plasmid pEC302/04 were identified as described in section 6.2.3.2. Insertion-sequence (IS) elements were identified by tBLASTx searches against IS-finder (https://www-is.biotoul.fr/). All annotations were then manually curated. The multilocus sequence type of plasmid pEC302/04 was determined *in silico* using the PubMLST server (http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_plasmid_ seqdef&page=sequenceQuery). Due to the multi-replicon nature of IncF plasmids, the FAB formula (Villa et al., 2010) was applied to assign allelic profiles for three types of IncF replicons, where "F" represents IncFII replicon, "A" represents FIA and "B" represents FIB. For instance, if a plasmid X has an FAB formula of F1:A2:B-, it indicates that plasmid X carries an FII replicon with assigned allelic profile of "1," a FIA replicon with allelic profile "2"; while the FIB replicon is absent.

6.2.6 Phenotypic characterization of *E. coli* strains

6.2.6.1 Biolog phenotype microarray assay and analysis

Phenotype microarray (PM) assay was carried out for both EC302/04 and EC096/10 using the OmniLog® automated incubator/ reader (Biolog Inc., Hayward, USA) as described previously (Shea et al., 2012). A total of 8 PM plates containing different carbon sources (PM1-2A), nitrogen sources (PM3B), phosphorous and sulphur sources (PM4A), nutrient supplements (PM5), and peptide nitrogen sources (PM6, PM7, PM8) were used to determine the phenotypes of EC302/04 and EC096/10. All PM assays were carried out in duplicates.

In brief, *E. coli* strains were streaked on LBA and incubated for 18 hours at 37°C. Cell suspension of each strain was then prepared by picking multiple single colonies with a cotton swab and then resuspended in 15 ml of Biolog inoculating fluid (IF-O).

The cell density was adjusted to 85% transmittence and 1% of Biolog dye A (vol/vol) was then added subsequently. Each well of the 96-well PM microplates was then filled with 100 µl of the freshly prepared cell suspension. All PM microplates were incubated at 37°C in an OmniLog reader for 48 hours. The colour changes in the wells (formation of purple colour indicates positive results) were recorded using the automated Omnilog system. PM analysis was performed using OmniLog® phenotype microarray software v1.2 with selected option of A1 zero to substract background signals with reference to the A1 negative control that was present in each and every PM microplate. The phenotypic differences between the two strains (EC302/04 and EC096/10) were determined using the OmniLog® phenotype microarray software v1.2. ORFs with non-synonymous SNPs (nsSNPs) which may be involved in the metabolic pathway of substrates where the substrate utilization can be determined using PM assay were further investigated.

6.2.6.2 Conjugation assay and confirmation of transconjugants

The transmissibility of plasmid pEC302/04 was determined using liquid conjugation experiment at 37ºC as described previously (Johnson et al., 2010). EC302/04 and a nalidixic acid-resistant strain of *E. coli* DH5α were used as donor cell and recipient cell, respectively. In brief, 1.8 ml of an overnight culture of the recipient *E. coli* strain (i.e., DH5α) was mixed with 0.2 ml of donor cells (i.e., EC302/04) in exponential growth phase in LBB. Mixtures were then incubated overnight without shaking at 37°C. Transconjugants were confirmed by: a) selective plating on LB agar plates containing a donor-inhibiting concentration of ampicillin (100 µg/ml; Sigma-Aldritch) and a recipient-inhibiting concentration of nalidixic acid (30µg/ml; Sigma-Aldritch); b) PCR detection of the bla_{TEM} gene which was only present in plasmid pEC302/04 and; c) PFGE profiling of the donor, recipient cells and transconjugants. The conjugation

frequency was calculated using the averages of two biological and three technical replicates.

Conjugation frequency = $\frac{number\ of\ transconjugants}{number\ of\ donor\ cells}$

6.2.6.3 Bacterial growth assays in iron-limited media and analysis

E. coli EC096/10, EC302/04, the transconjugant of EC302/04 (DH5α (pEC302/04)) and DH5α were grown in LB broth overnight (16-18 hours) at 37ºC. Cultures were diluted to OD_{600} of 0.05-0.1 into LB broth and LB broth supplemented with different concentrations of the iron chelator 2′2-dipyridyl (DIP) (Sigma-Aldrich) (100 μM, 200 μM, and 300 μM). The growth assay was performed in 96-well plates at 37ºC with sample reading taken every 15 min for 24 h using a Spectramax spectrophotometer (Molecular Devices, USA). The results are presented as the averages of three independent biological replicates with two technical replicates each. The growth rates for *E. coli* strains across all tested media were determined using GrowthRates (Hall et al., 2014). The output from the Spectramax spectrophotometer (readings of cell density at an interval of 15 min for 24 h) in the format of tab-delimited text was used directly as the input files for GrowthRates. The growth rates during the exponential phase and maximum OD were then determined using GrowthRates.

6.2.6.4 Statistical analysis

The differences in growth rates between strains harbouring the plasmid and without plasmid [(EC302/04 versus EC096/10) and (DH5α (pEC302/04) versus DH5α)] in different media were determined using unpaired t-test (http://www.physics.csbsju.edu/stats/t-test_bulk_form.html). A p-value < 0.05 is considered as statistically significant.

6.3 RESULTS

6.3.1 General features of EC302/04, EC096/10 and plasmid pEC302/04

The general genomic features and background data of EC302/04 and EC096/10 are detailed in Table 6.2.

Table 6.2: Genomic features of EC096/10, EC302/04 and pEC302/04

Feature		*E. coli* strain	
		EC096/10	**EC302/04**
Sequence		WGS	WGS
Accession number		AONF00000000	AMFM00000000
Isolate origin	**Country**	Malaysia	Malaysia
	Year	2010	2004
	Source	stool	tracheal aspirate
General genomic features	**Accumulated length (bp)**	4745284	4846195
	Contigs	63	82
	N50 (bp)	247444	201460
	Longest assembled contig	676,708 bp	350,759 bp
	Fold-coverage	204	312
	G+C content (%)	51	51
	CDS	4563	4687
Plasmid	**Plasmid**	absent	pEC302/04
	Size (bp)		140232
	G+C content (%)		52
	CDS		185
	Accession number		CP011493
RNA	**tRNA**	n=71	n=70
	rRNA	n=4	n=4
Molecular type	**Serotype**	O166:H15.	O166:H15
	MLST	ST 349	ST 349
	Phylogenetic group	D	D

EC302/04 and EC096/10 which were initially found to be closely related based on their respective PFPs were indeed clonal as they shared the same serotype (O166:H15), sequence type (ST349) and phylogenetic group (D). *E. coli* EC302/04 harboured a large plasmid of 140,232 bp that was absent from EC096/10. The genome sequences of all three replicons (EC302/04, EC096/10 and pEC302/04) have also been deposited into

GenBank and the accession numbers are listed in Table 6.2.

6.3.2 Similar virulence determinants and pathogenicity islands in EC096/10 and EC302/04

Iron uptake from the host can be carried out directly or indirectly. A system responsible for the direct uptake of iron encoded by the *chuAS-chuTWXYUV* genes (Panina et al., 2001) was found in both strains. Several indirect iron acquisition systems via high-affinity ferric iron chelators, the siderophores (Miethke & Marahiel, 2007) were also identified. A large locus consisting of genes for siderophore enterobactin biosynthesis (*entABCDEF*) and uptake (*fepABCDEG*) was identified in both EC302/04 and EC096/10. Another siderophore aerobactin (*iut/iucABCD*) was found encoded next to the *sitABCD* operon (manganese and iron transporter), and were collectively found in the plasmid pEC302/04 in strain EC302/04 while such plasmid is absent EC096/10 (detailed in section 6.3.3).

Pili and fimbriae can act as adherence factors for both biotic and abiotic surfaces (Schembri & Klemm, 2001). The *E. coli* common pilus (*ecp*) that is encoded by *ecpRABCDE,* and the *fim* operon (*fimA, fimB, fimC, fimD, fimE, fimF, fimG, fimH and fimI*), which is responsible for expression and assembly of the type I fimbriae, were identified in both strains. Genes encoding for outer membrane protein A (OmpA) which has been reported to be involved in invasion and serum resistance (Prasadarao et al., 2002), and AslA, an arylsulfatase which plays a role in invasion (Hoffman et al., 2000), were also found in both EC302/04 and EC096/10.

Type III secretion system (T3SS) serves as a virulence mechanism enabling cells to secrete bacterial proteins called effectors into the host, which is important for the virulence and survival of bacteria (Coburn et al., 2007). T3SS proteins are generally

grouped into four classes, namely the structural operon that is responsible for expression proteins, and genes that encode translocon proteins, effector proteins and the chaperones (Niu et al., 2013). However, an *E. coli* type III secretion system 2 (ETT2) that is not involved in the classic secretion of effectors (Ideses et al., 2005) (as confirmed using the SIEVE server) (Samudrala et al., 2009), was found in both EC302/04 and EC096/10. Complete ETT2 loci have been reported to be widely distributed in *E. coli* especially in IPEC (especially EHEC O157:H7) and to a lesser extent, in ExPEC and non-pathogenic *E. coli* (Zhou et al., 2014). The ETT2 locus best resembles the pathogenicity island (PAI) designated ETT2sepsis, previously identified from a septicemic *E. coli* (Ideses et al., 2005). Furthermore, ETT2 has also been reported to be involved in bacterial virulence such as biofilm formation, bacterial adhesion and invasion (Zhou et al., 2014). The presence of a complete ETT2 locus in both EC302/04 and EC096/10 strains indicate the potential virulence of both strains despite their difference in isolation source (intestinal and extraintestinal sites). Both EC302/04 and EC096/10 shared similar fitness traits except for those that are located in plasmid pEC302/04.

6.3.3 Characterization of plasmid pEC302/04

E. coli EC302/04 harboured a 140,232 bp plasmid designated pEC302/04 that was absent in *E. coli* EC096/10. pEC302/04 is a circular multi-replicon IncFIIA plasmid with an average GC content of 52% (Figure 6.2). A total of 184 ORFs were predicted and annotated using RAST (Aziz et al., 2008) which were then manually curated to improve the annotations (Appendix V; e).

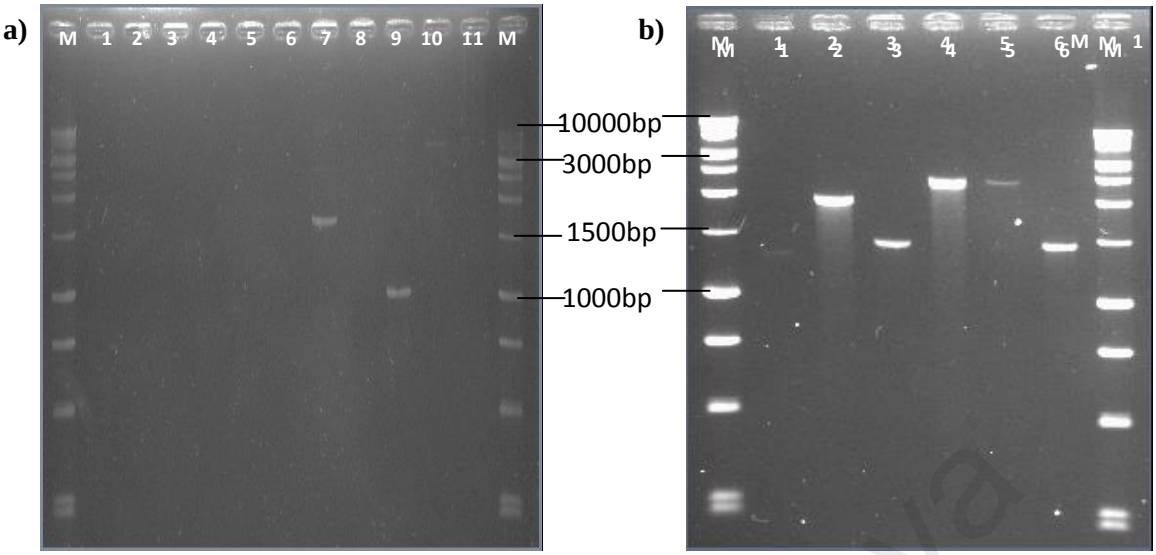


Figure 6.1: Representative agarose gel pictures of PCR- based plasmid gap closing. **a) PCR determination of the linkage between two putative plasmid contigs of EC302/04, i.e contigs number 63 and 67 (accession number: AMFM00000000).** Lane M: 1 kb DNA ladder (Promega); lanes 1, 3, 5, 7, 9, 11: EC302/04; lanes 2, 4, 6, 8, 10: EC096/10. **b) Purified PCR amplified regions between two putative plasmid contigs with EC302/04 used as the DNA template.** Lane M: 1 kb DNA ladder (Promega); lanes 1, 2, 3: linkage between contigs 17 and 63; lane 4, 5: linkage between contigs 68 and 30; lane 6: linkage between contigs 22 and 42.

A total of three replicons were found for pEC302/04, namely RepFIA, RepFIIA and RepFIB with the multireplicon F plasmid FAB formula of F2:A1:B1. A complete *tra* region that codes for the conjugative transfer components of plasmid were identified in pEC302/04 and it spans approximately a quarter (36 kb) of the plasmid. This *tra* region comprised of 24 *tra* genes (*traM*, *traJ*, *traY*, *traA*, *traL*, *traE*, *traK*, *traB*, *traP*, *traV*, *traR*, *traC*, *traW*, *traU*, *traN*, *traF*, *traQ*, *traH*, *traG*, *traS*, *traT*, *traD*, *traI* and *traX*), 9 *trb* genes (*trbD*, *trbG*, *trbI*, *trbC*, *trbE*, *trbA*, *trbB*, *trbJ* and *trbF*) and the regulatory fertility inhibition gene (*finO*) which act as a conjugal transfer repressor (Yoshioka et al., 1987).

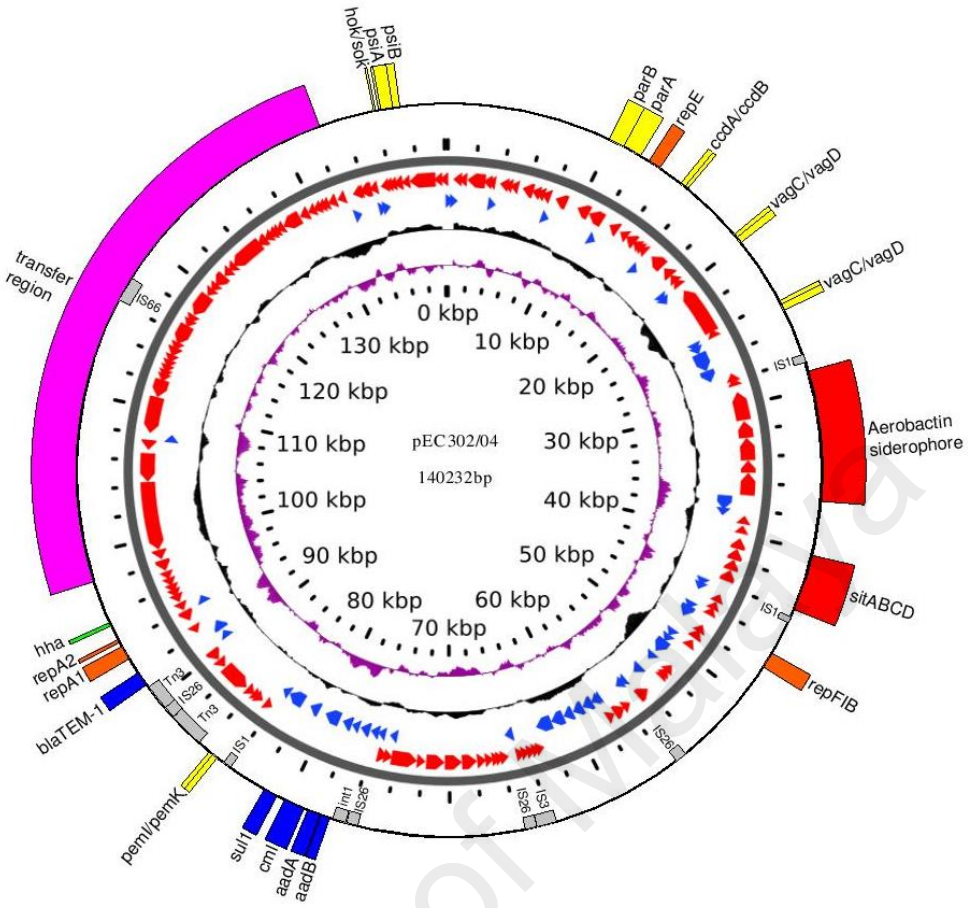


Figure 6.2: Circular map of plasmid pEC302/04. The 1st (innermost ring) and the 6th rings show the plasmid size in kbp. The 2nd and 3rd rings (from inner to outer) show the GC skew (purple) and GC content (black), respectively. The 4th and 5th rings show the ORFs in forward (blue) and reverse orientation (red). The outermost ring shows the important genes encoding regions highlighted with different colours according to its function: orange, replication; pink, transfer region; yellow, addictive systems; red, iron acquisition systems; blue, antimicrobial resistance; green, other virulence factors; grey, IS elements.

To investigate if pEC302/04 is self-transmissible, conjugation experiments were carried out for EC302/04 as the donor with a spontaneous nalidixic acid-resistant *E. coli* DH5α as the recipient. Transconjugants were detected at a frequency of 5×10^{-4}. Subsequent PCR detection of the bla_{TEM} gene carried on pEC302/04 in the transconjugants and PFGE of the donor, recipient and transconjugant cells (Figures 6.3 & 6.4) conclusively showed that pEC302/04 is indeed self-transmissible.

Comparison of the resistance phenotypes of EC302/04, DH5α and the DH5α

(pEC302/04) transconjugant revealed that genes conferring resistance to ampicillin, kanamycin, streptomycin, spectinomycin, compound sulfonamides, gentamicin, tobramycin and chloramphenicol have been transferred to the transconjugant (Tables 6.3 & 6.4).

In silico analysis of resistance genes carried on pEC302/04 also showed concordant results with that of the phenotypic antimicrobial susceptibility testing. A class I integron harbouring the *aadA1* gene (conferring resistance to streptomycin and spectinomycin), *aadB* gene (conferring resistance to gentamicin, kanamycin and tobramycin) and *cmlA* (non-enzymatic chloramphenicol resistance gene that encodes for an efflux pump) were identified in pEC302/04. Besides the resistance determinants, *sulI* (conferring resistance to sulphonamide) and *qacEΔ* (quatenary compound resistance gene) which are genes found in the 3'-conserved segment of class I integrons (Mazel, 2006) were also found in the pEC302/04-encoded integron. Notably, pEC302/04 also carried a $bla_{\text{TEM-1}}$ gene (conferring ampicillin resistance) within a Tn*3* transposon which, in turn, is interrupted with a copy of the IS*26* element (Figure 6.2, Appendix V; d). Collectively, the findings revealed that the non-susceptibility to amoxicillin/clavulanic acid and cefoperazone, which are associated with bla_{AmpC} (Jacoby, 2009) is not plasmid pEC302/04-mediated in EC302/04.

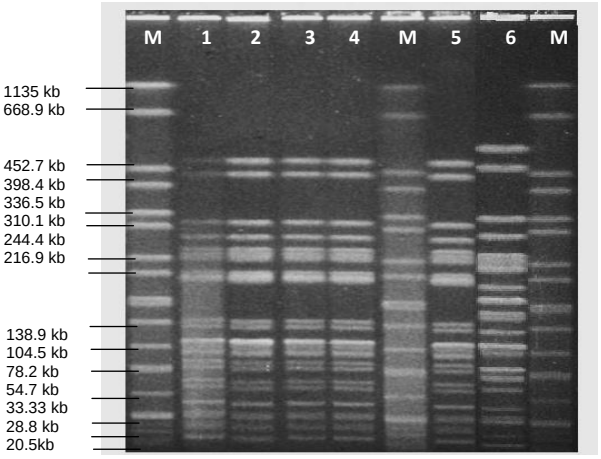


Figure 6.3: PFGE gel of the DH5α (pEC302/04) transconjugants, DH5α and EC302/04. Lane M: *XbaI* digested H9812 DNA marker; lanes 1-4: DH5α (pEC302/04) transconjugants; lane 5: DH5α; lane 6: EC302/04.

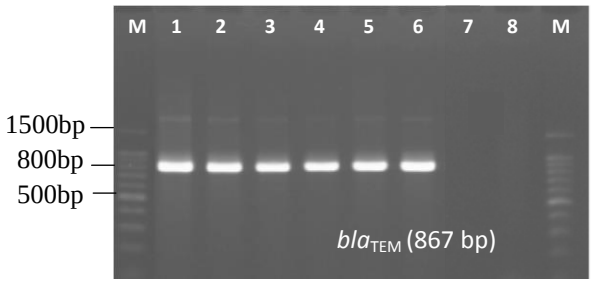


Figure 6.4: *bla*$_{TEM}$ detection in DH5α (pEC302/04) transconjugants, DH5α and EC302/04. Lane M: 100 bp DNA ladder (Promega); Lane l: positive control, lanes 2-5: DH5α (pEC302/04); lane 6: EC302/04; Lane 7: DH5α; Lane 8: negative control.

Table 6.3 : Resistance phenotypes of DH5α (pEC302/04) transconjugants, DH5α and EC302/04.

Strain	RESISTANCE PHENOTYPE										
	AMP	AMC	CFP	GEN	KAN	CHL	STR	SPC	TOB	S300	NAL
EC302/04[1]	R	R	R	R	R	R	R	R	R	R	S
DH5α (pEC302/04)[1]	R	S	S	R	R	R	R	R	R	R	R
DH5α[2]	S	S	S	S	S	S	S	S	S	S	R

[1]strain harboring plasmid pEC302/04; [2] strain without plasmid pEC302/04; R= non-susceptible (including resistant and intermediate resistant); S= susceptible.

Table 6.4: *In silico* analysis of pEC302/04-encoded resistance genes.

Gene name	Locus tag[1]	Best hit to ResFinder[2]		Resistance phenotype		
		% Identity[2]	Accession no. [2]	Predicted phenotype[2]	Reported phenotype[3]	Actual phenotype[4]
aadB	pEC302/04_107	100	JN119852	Aminoglycoside resistance	GEN; KAN; TOB	GEN; KAN; TOB
aadA1	pEC302/04_108	100	JQ414041	Aminoglycoside resistance	STR; SPC	STR; SPC
cmlA	pEC302/04_109	99.05	AB212941	Beta-lactam resistance	CHL	CHL
sul1	pEC302/04_110	100	CP002151	Sulphonamide resistance	S300	S300
$bla_{\text{TEM-1}}$	pEC302/04_121	100	JF910132	Phenicol resistance	AMP	AMP

[1]refer to Appendix V; d. [2]In silico results obtained from ResFinder. [3]Resistance phenotypes reported by other studies. [4]Resistance phenotypes determined using antimicrobial susceptibility testing according to CLSI (CLSI, 2015)

The *parABS* segregation system, which is responsible for an active plasmid partition system, was found upstream of the *tra* region of pEC302/04. The partition system is also important in ensuring each daughter cell receives a plasmid following cell division (Kroll et al., 2010). Besides, several toxin/antitoxin (TA)-based addiction systems were also found on pEC302/04: (i) PemI-PemK (Pem for plasmid emergency maintenance) TA system is found associated with a transfer inhibition protein (Tir); (ii) CcdA-CcdB (Ccd for coupled cell division) TA system is found near to the FIA replicon; (iii) two copies of VagC-VagD (Vag for virulence-associated gene) TA systems are separated by 2 ORFs encoding hypothetical proteins and are found near to another addiction system, CcdA-CcdB and (iv) Hok-Sok (Hok for host killing and Sok for suppression of killing) TA system is found next to the gene that codes for PsiA-PsiB (plasmid SOS inhibition protein) (Figure 6.2). All these segregation and addiction systems are important to ensure stable plasmid inheritance (Kroll et al., 2010).

Ten ORFs were identified as belonging to insertion sequences (ISs) and transposons with five main IS families (IS*1*, IS*3*, IS*6*, IS*66* and Tn*3*) being found in pEC302/04. Interestingly, among the five IS families, pEC302/04 contained relatively higher abundance of IS*6* (4 copies) and IS*1* (3 copies). Genes encoding the two putative iron acquisition systems (*iuc/iutABCD* and *sitABCD*) were collectively flanked by two copies of IS*1* in an inverted orientation, a genetic organization which resembled a composite transposon (Figure 6.2). Although the IS*1*-flanked aerobactin siderophore encoding DNA fragment has been demonstrated experimentally to be transposable in *E. coli* (Perez-casal & Crosa, 1984; Lorenzo et al., 1988), the target site duplication (evidence of a tranposition event) was not found in pEC302/04, indicating that this IS*1*-flanked fragment may not have transposed recently. Besides virulence genes, transposable elements are also often associated with resistance genes (Kiiru et al., 2013). The $bla_{\text{TEM-1}}$ resistance gene was found within the Tn*3* transposon, in which the

transposase was disrupted by a copy of IS*26*. The integrase gene of the class I integron was also interrupted by a copy of IS*26*. These findings suggest that the integrase and transposase are likely non-functional, rendering the class I integron and Tn*3* transposon non-mobile. Interestingly, the complete transfer regions of pEC302/04 was also interrupted by an IS element, in this case IS*66*, which was inserted into the intergenic region between the *traN* and *trbE* genes, without affecting the self-transmissibility of pEC302/04.

Twenty-four putative virulence factors (VFs) were identified in pEC302/04 with majority of the VFs being associated with iron acquisition systems. The region encoding two iron acquisition systems constitutes the putative virulence region which spans approximately 16 kb of pEC302/04. The aerobactin siderophore-encoding gene clusters (*iutA-iucABC*D) were found adjacent to *sitABCD* (encoding for iron and manganese transport system) and were separated by genes encoding ShiF (putative membrane transport protein) and Eno (phosphopyruvate hydratase). Both systems were reportedly required for growth under iron depleted condition (Boyer et al., 2002) as well as increased virulence for the *E. coli* strain both *in vivo* and *in vitro* (Boyer et al., 2002; Sabri et al., 2006).

Besides iron acquisition systems, virulence determinants with other functions were also identified on pEC302/04. The *traJ* gene, which codes for a plasmid conjugation transfer protein, may play an important role in invasion (Hill et al., 2004). A complete transfer region which encodes for proteins corresponding to type IV pilus (Lawley et al., 2003; Zahrl et al., 2006), that has been described to be urovirulence determinants for ExPEC (UPEC in particular) (Kulkarni et al., 2009) was also identified.

6.3.4 Limited core SNPs and little to no variation for phenomic characteristics were observed for both EC302/04 and EC096/10

Besides the plasmid pEC302/04, which was only found in EC302/04, the two genomes were highly similar in terms of genetic contents (Figure 6.5). Both genomes differed by a very limited number of core SNPs, with 61 non-synonymous SNPs (nsSNPs) and 33 synonymous SNPs (sSNPs) identified in 34 shared ORFs. Despite the variations being found for both strains, EC096/10 and EC302/04 can be considered as clonally related since in general, a "clone threshold" value of 200 SNPs was proposed to define clones for *E. coli* although different outbreak clones may have different SNP cutoff values, subjected to factors such as outbreak duration (Kaas, 2014).

A total of 34 ORFs with non-synonymous mutations which may be involved in the metabolic pathway of substrates were further investigated by phenotype microarray assay. Fructose-1,6-bisphosphatase II, arginine decarboxylase, and glutamate racemase were proteins with nsSNPs identified in their respective genes. No significant differences were observed for the growth kinetics of EC302/04 and EC096/10 supplemented with fructose-1,6-bisphosphate [substrate of fructose-1,6-bisphosphatase II; PM1 (well 52)], arginine [substrate of arginine decarboxylase; PM2 (well 76), PM3 (well 8), PM5 (well 4)], agmatine [product of arginine decarboxylase; PM3 (well 48)] and glutamate [substrate of glutamate racemase; PM1 (well 24), PM3 (well 12, 30) PM5 (well 8, 44)] in the PM assay (Figure 6.5). Despite the genetic variation found in both strains, no observable phenotypic differences (substrate utilization of carbon, nitrogen, phosphorus and sulfur, nutrient supplements, and peptide nitrogen) were found between the two strains using PM assay that was carried out at 37ºC, the habitat temperature for both EC302/04 and EC096/10 (detailed in Figure 6.6).

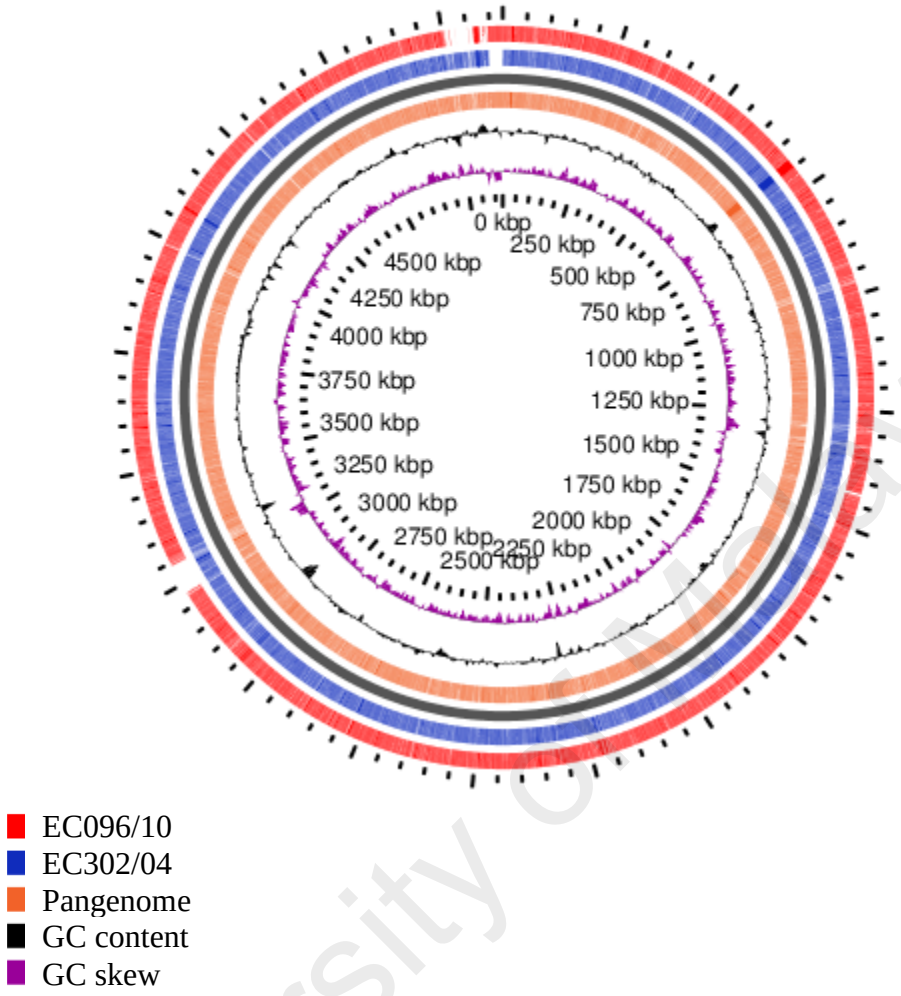


Figure 6.5: Circular map of EC302/04 and EC096/10. High similarity between genomes of EC096/10 (in red) and EC302/04 (in blue).

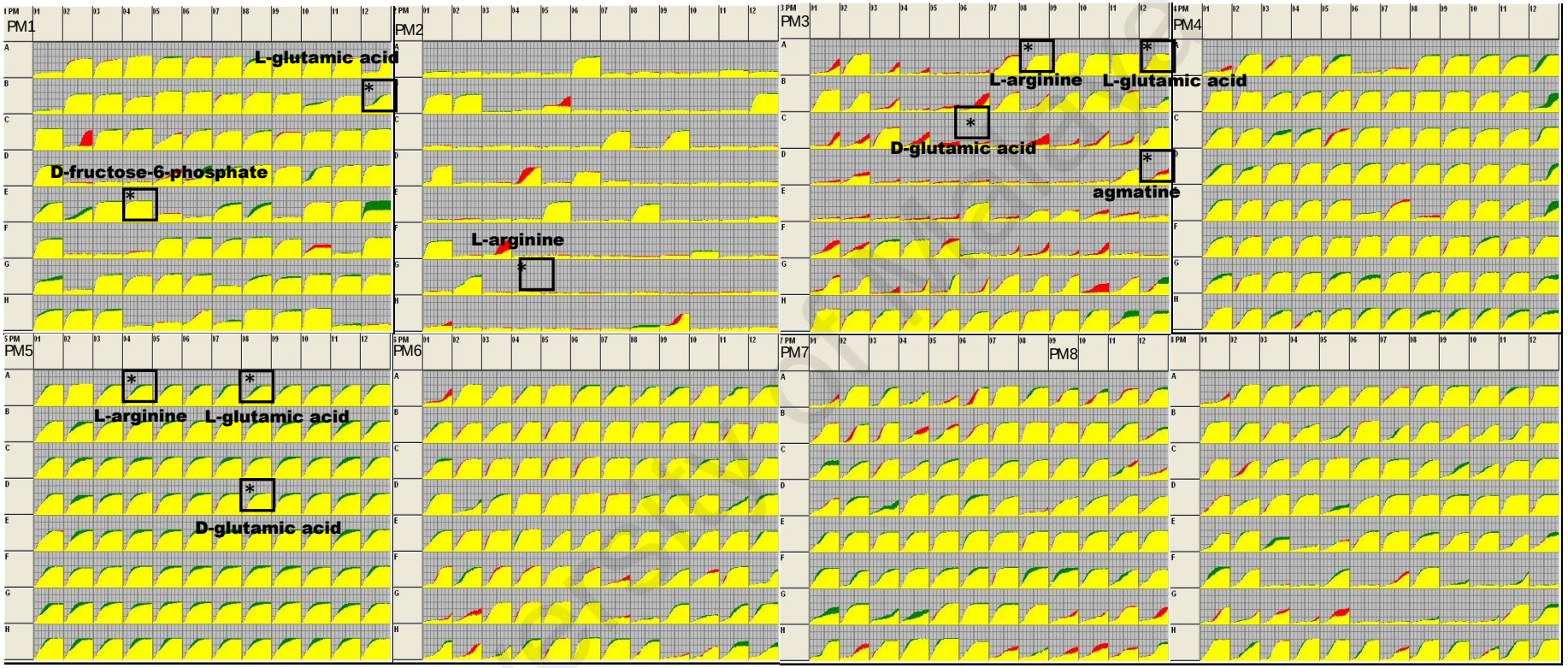


Figure 6.6: PM analysis of EC302/04 and EC096/10. Growth curves are red for EC302/04, green for EC302/04, and yellow for the overlapping region. Black boxes indicate the growth curves for substrates that may be involved in the metabolic pathways and which may be affected by ORFs with non-synonymous mutations.

6.3.5 Growth of EC302/04 and EC096/10 in iron-limited medium

Clonally related *E. coli* EC302/04 and EC096/10 were isolated from extraintestinal and intestinal sites, respectively, and it is known that extraintestinal sites contain very low concentrations of iron that is essential for bacterial growth (Hagan et al., 2010). Plasmid pEC302/04, as the major genetic difference found between the clonally related EC302/04 and EC096/10 strains, encode additional iron acquisition systems. This led to the further comparison on the bacterial growth of *E. coli* EC302/04, EC096/10, DH5α and DH5α transconjugant harbouring pEC302/04 in rich media and iron-limited media to investigate if pEC302/04 could confer growth advantage in an environment with low iron concentrations.

When all 4 strains were grown in LB broth supplemented with iron chelator (DIP) at different concentrations, their respective growth were attenuated and the maximum levels of OD_{600nm} of all strains were reduced as the concentration of DIP increased (Figure 6.6). Nevertheless, there was no difference in growth rates between DH5α (pEC302/04) versus DH5α in all rich and iron-limited media tested (Figure 6.6). On the other hand, EC302/04 has slightly better growth compared to EC096/10 in LB broth across all media tested. These results indicated that the carriage of pEC302/04 in the DH5α (pEC302/04) transconjugant did not have a significant impact on their growth in iron-limited media *in vitro*.

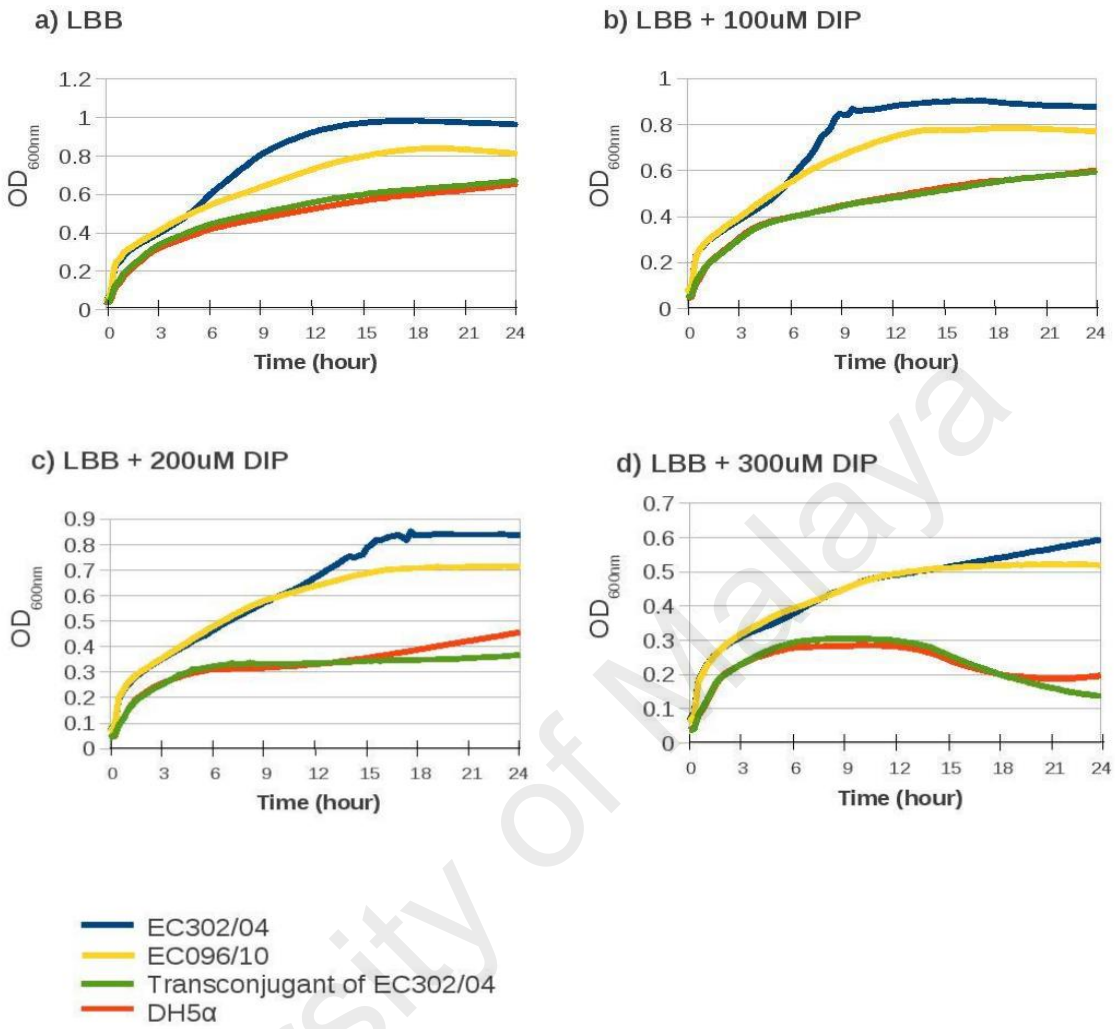


Figure 6.7: Growth curves of EC302/04, EC096/10, DH5α and the DH5α (pEC302/04) transconjugant in rich and iron-limited medium. (A) LB broth; (B) LB broth supplemented with 100 µM of DIP; (C) LB broth supplemented with 200 µM of DIP; (D) LB broth supplemented with 300 µM of DIP.

6.3.6 Phylogenomic analysis of 40 *E. coli* strains from various sources

Phylogenomic analysis of 40 *E. coli* strains from various sources was performed using the accumulated alignment length of 3,227,599 bp, which covers 70.8% and 59% of the smallest and largest *E. coli* genome sizes that were included in this comparative genomic analysis. Phylogenomic analysis revealed that *E. coli* of different pathotypes that are closely grouped together were found to have similar serotypes, STs as well as virulence profiles, hence sharing similar phylogenetic signals. In fact, some of these strains may be clonally related (Cluster B2, Figure 6.7). Both *E. coli* EC302/04 (extraintestinal strain) and EC096/10 (intestinal strain) were found to be closely related to strains KTE147, HM46, 5-366-08_S1_C3, AB42410445-isolate1, AB42602061-isolate 1 and MC6003, which belonged to ExPEC or intestinal strains from patients with extraintestinal infections (Figure 6.7, Cluster B2). *E. coli* strains EC096/10, EC302/04, KTE147, HM46, and 5-366-08_S1_C3 also belonged to the same sequence type (ST349) and serotype (O166:H15), with the exception of strain 5-366-08_S1_C3 which belonged to ST2914, a single locus variant of ST349 (Figure 6.7). The remaining three strains in Cluster B2 (AB42410445-isolate1, AB42602061-isolate 1 and MC6003) were obtained from bacteremic patients, and all three are triple locus variant of ST349 and belonged to serotype O7:H15. Further screening for STs 349, 2914 and 70 in the NCBI database led to the discovery of more *E. coli* strains belonging to these rare STs with the majority of these strains categorized as ExPECs or intestinal strains obtained from patients with extraintestinal infections (Table 6.5). In fact, more than half of the *E. coli* strains in Cluster B2 share highly similar virulence profiles, regardless of intestinal or ExPEC strains, indicating the genetic propensity of intestinal strains to cause extraintestinal infections.

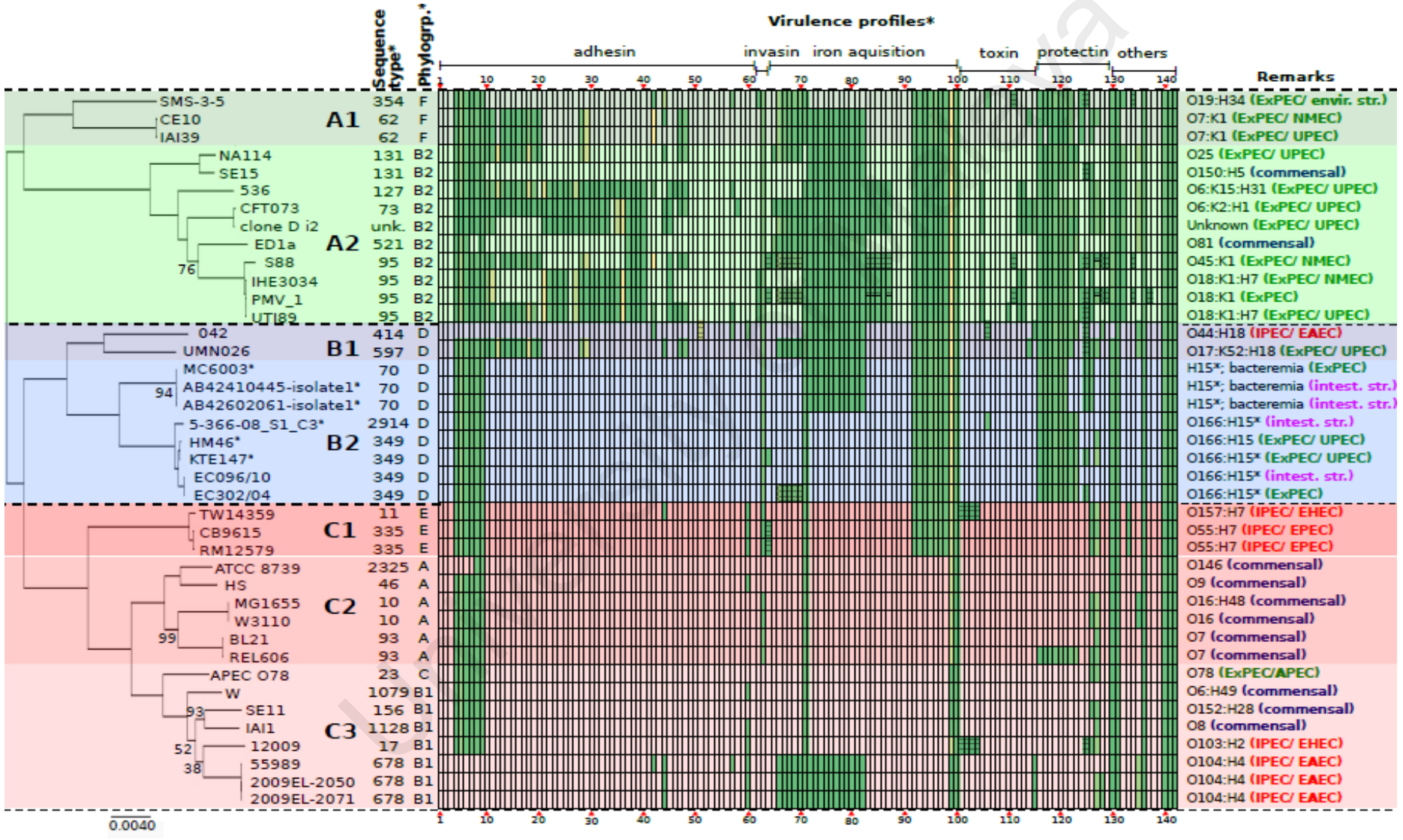

Virulence profiles*
adhesin
invasin
iron aquisition
toxin
protectin
others
Sequence type*
Phylogrp.*
Remarks
A1
A2
B1
B2
C1
C2
C3
SMS-3-5 354 F O19:H34 (ExPEC/ envir. str.)
CE10 62 F O7:K1 (ExPEC/ NMEC)
IAI39 62 F O7:K1 (ExPEC/ UPEC)
NA114 131 B2 O25 (ExPEC/ UPEC)
SE15 131 B2 O150:H5 (commensal)
536 127 B2 O6:K15:H31 (ExPEC/ UPEC)
CFT073 73 B2 O6:K2:H1 (ExPEC/ UPEC)
clone D i2 unk. B2 Unknown (ExPEC/ UPEC)
ED1a 521 B2 O81 (commensal)
S88 95 B2 O45:K1 (ExPEC/ NMEC)
IHE3034 95 B2 O18:K1:H7 (ExPEC/ NMEC)
PMV_1 95 B2 O18:K1 (ExPEC)
UTI89 95 B2 O18:K1:H7 (ExPEC/ UPEC)
042 414 D O44:H18 (IPEC/ EAEC)
UMN026 597 D O17:K52:H18 (ExPEC/ UPEC)
MC6003* 70 D H15*; bacteremia (ExPEC)
AB42410445-isolate1* 70 D H15*; bacteremia (intest. str.)
AB42602061-isolate1* 70 D H15*; bacteremia (intest. str.)
5-366-08_S1_C3* 2914 D O166:H15* (intest. str.)
HM46* 349 D O166:H15 (ExPEC/ UPEC)
KTE147* 349 D O166:H15* (ExPEC/ UPEC)
EC096/10 349 D O166:H15* (intest. str.)
EC302/04 349 D O166:H15* (ExPEC)
TW14359 11 E O157:H7 (IPEC/ EHEC)
CB9615 335 E O55:H7 (IPEC/ EPEC)
RM12579 335 E O55:H7 (IPEC/ EPEC)
ATCC 8739 2325 A O146 (commensal)
HS 46 A O9 (commensal)
MG1655 10 A O16:H48 (commensal)
W3110 10 A O16 (commensal)
BL21 93 A O7 (commensal)
REL606 93 A O7 (commensal)
APEC O78 23 C O78 (ExPEC/APEC)
W 1079 B1 O6:H49 (commensal)
SE11 156 B1 O152:H28 (commensal)
IAI1 1128 B1 O8 (commensal)
12009 17 B1 O103:H2 (IPEC/ EHEC)
55989 678 B1 O104:H4 (IPEC/ EAEC)
2009EL-2050 678 B1 O104:H4 (IPEC/ EAEC)
2009EL-2071 678 B1 O104:H4 (IPEC/ EAEC)
76
94
99
93
52
38
0.0040

Figure 6.8: Phylogenomic tree inferred by maximum likelihood method using REALPHY. Dotted lines and coloured boxes were used to facilitate visualization according to clusters. Asterisk (*) indicates *in-silico* results. Red, blue, green and purple fonts represent IPEC, commensal, ExPEC and intestinal strains, respectively. Bootstrap values below 100 are shown at their respective branch points. Coloured boxes to facilitate visualization of the phylogenomic clustering.

Table 6.5: Details of *E. coli* genomes with sequence types 349, 2914 and 70

Strain	Accession number	*E. coli* group	Country	Allelic profile							Sequence type
				adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	
EC096/10	AONF00000000	intestinal strain	Malaysia	34	36	39	87	67	16	4	349
EC302/04	AMFM00000000	ExPEC/ tracheal aspirate	Malaysia	34	36	39	87	67	16	4	349
KTE147	ANWN00000000.1	ExPEC/ UPEC	Denmark	34	36	39	87	67	16	4	349
HM46	APNY00000000.1	ExPEC/ UPEC	Michigan, USA	34	36	39	87	67	16	4	349
TOP2396-2	AOQR00000000.1	stool sample from individual with recurrent urinary tract Infection	Seattle, USA	34	36	39	87	67	16	4	349
TOP2396-3	AOQS00000000.1	stool sample from individual with recurrent urinary tract Infection	Seattle, USA	34	36	39	87	67	16	4	349
5-366-08_S1_C3	JONE00000000.1	intestinal strain	Tanzania	34	36	207	87	67	16	4	2914
5-366-08_S1_C1	JOQU00000000.1	intestinal strain	Tanzania	34	36	207	87	67	16	4	2914
AB42410445-isolate1	ATOD00000000.1	stool sample of individual with bacteremia	USA	34	36	28	25	28	16	4	70
AB42602061-isolate1	ATOB00000000.1	stool sample of individual with bacteremia	USA	34	36	28	25	28	16	4	70
AB42554418-isolate1	ATOC00000000.1	stool sample of individual with bacteremia	USA	34	36	28	25	28	16	4	70
MC6003	ATNV00000000.1	blood sample of individual with bacteremia	USA	34	36	28	25	28	16	4	70
Upec-105	JSOV00000000.1	ExPEC/ UPEC	USA	34	36	28	25	28	16	4	70
Upec-154	JSMW00000000.1	ExPEC/ UPEC	USA	34	36	28	25	28	16	4	70
Upec-126	JSNZ01000001.1	ExPEC/ UPEC	USA	34	36	28	25	28	16	4	70

6.3.7 Genetic variations in *E. coli* belonging to the same clonal complex

The identification of several *E. coli* strains, namely the EC096/10, EC302/04, KTE147, HM46, and 5-366-08_S1_C3 that belong to the same and yet uncommon clonal complex (ST349) which possess intrinsic extraintestinal virulence potential led to further investigation and comparison with respect to their resistotypes, virulence profiles, plasmid replicon types and CRISPR clusters.

CRISPR-cas system is reportedly to act as a defense system to recognize and destroy specific foreign DNA such as the phage genomes and is comprised of clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR associated (*cas*) genes (Bondy-Denomy & Davidson, 2014). A CRISPR locus is identified in all 5 strains, namely the EC096/10, EC302/04, KTE147, HM46, and 5-366-08_S1_C3. A total of 5 repeats (32 bp; 4 identical repeats and 1 repeat with a 4 bp difference), interspaced by 4 different spacers were detected (29 bp) (Table 6.6). The direct repeats and spacers for EC096/10, EC302/04, KTE147, and HM46 are identical with 2 SNPs found in the CRISPR locus of strain 5-366-08_S1_C3, indicating high similarity in the CRISPR regions of these *E. coli* strains belonging to the same clonal complex ST349. Although CRISPR is deemed useful for genotyping and population structure analysis for different bacterial species including *Mycobacterium tuberculosis*, *Mycobacterium canettii*, *Yersinia pestis* and *Pseudomonas aeruginosa*, CRISPR typing is not used as a general *E. coli* typing tool as there is no exact correlation between the CRISPR with other genotyping method such as the MLST (Pourcel & Drevet, 2013).

Table 6.6: Sequences of direct repeats and spacers of the CRISPR locus for EC096/10, EC302/04, KTE147, HM46, and 5-366-08_S1_C3.

Direct repeat sequence*	**Size (bp)**
ATGGTTATCCCCGCTGACGCGGGGAACTC	29
CGGTTTATCCCCGCTCGCGCGGGGAACTC	29
CGGTTTATCCCCGCTCGCGCGGGGAACTC	29
CGGTTTATCCCCGCTCGCGCGGGG(A/**G**)ACTC	29
CGGTTTATCCCCGCTCGCGCGGGGAACTC	29
Spacer DNA sequence*	**Size (bp)**
GCCAGTGCCTGGTCATATGTGTAACACCTATC	32
TTGCTGTATAGAGATTATCGGCGTGGGATTGT	32
GTTTGCGGCTT(G/**T**)CAAACATTTTAATGCGTATT	32
CCAGCCCGACACAAATAATGCAGCATTGCGCG	32

All the direct repeats and spacers identified were identical for EC096/10, EC302/04, KTE147 and HM46. *The underlined and bolded nucleotide is SNP observed in strain 5-366-08_S1_C3.

Genome analyses of EC096/10, EC302/04, HM46, KTE147 and 5-366-08_S1_C3 revealed genes that confer resistance to various antimicrobials (Table 6.7). All five genomes harboured the similar chromosomal mediated resistance determinants and also multidrug efflux pumps. Class C beta-lactamase (AmpC) gene can be chromosomal- or plasmid-encoded (Jacoby, 2009). An AmpC gene that belongs to periplasmic AmpC beta-lactamase is identified in all five genomes (Table 6.7). The five *E. coli* strains also harboured several multidrug efflux pumps encoding genes that are associated with decreased susceptibility to a range of antimicrobials including macrolide, puromycin, fosfomycin, enoxacin and norfloxacin, as predicted by ARDB (Liu & Pop, 2009). Several acquired antimicrobial resistance genes that are associated with aminoglycoside, phenicol, sulphonamide, trimethoprim and ampicillin resistance were also identified in EC302/04 and/or 5-366-08_S1_C3 but not EC096/10, KTE147 and HM46 (Table 6.7), where the acquired resistance genes of EC302/04 are plasmid pEC302/04 mediated

(detailed in section 6.3.3). Overall, the phenotypic and genotypic resistance profiles of EC302/04 and EC096/10 are in agreement with each other, with the exception of antimicrobials that are not commonly used for treatment of *E. coli* infections (MOH, 2014, 2013) as their susceptibility were not determined. On the other hand, the phenotypic resistance profiles of HM46, KTE147 and 5-366-08_S1_C3 have not been reported before.

Based on Figure 6.8, it is observable that EC096/10, EC302/04, HM46, KTE147 and 5-366-08_S1_C3 had distinct virulence profiles compared to ExPEC of other sequence types (Cluster B2). Only minor differences were seen among the virulence profiles of the 5 strains, with the most notable dissimilarity found for EC302/04, where the plasmid mediated iron acquisition system encoding genes were only found in EC302/04 (detailed in 6.3.3) but not other 4 *E. coli* strains of the same clonal complex ST349. Several other genes encoding plasmid transfer region (Tra), enteroaggregative heat-stable enterotoxin (AstA) and Increased serum survival (Iss) were the dispensable virulence factors identified among EC096/10, EC302/04, HM46, KTE147 and 5-366-08_S1_C3. Nonetheless, the location of these dispensable genes is unknown as only draft genome sequences were available for all 5 strains, except for EC302/04, where the genome sequence of plasmid pEC302/04 is completed.

Although EC302/04 and EC096/10 are assumed to be clonally related, EC302/04 harboured an additional IncFII plasmid pEC302/04 which carries extra virulence and resistance determinants compared to EC096/10. Hence, it is interesting to determine the plasmid replicon(s) that may be present in strains HM46, KTE147 and 5-366-08_S1_C3. Further interrogation revealed no plasmid replicon in HM46. On the other hand, strains KTE147 harboured IncFIA and IncI1 while 5-366-08_S1_C3 also carried two plasmid replicons, namely the IncB/O/K/Z and IncFII. Nevertheless, as mentioned previously,

only draft genome sequences were available for these strains. Hence, the presence of plasmids in these *E. coli* strains could not be ascertained as certain plasmid segment may be integrated into the bacterial chromosome yielding plasmid-derived chromosomal segment, an event which has been reported for *dnd* clusters (Ou et al., 2009).

Table 6.7: Resistance genes in *E. coli* strains belonging to the clonal complex ST349

Resistance genes*	Definition*	Predicted resistance phenotype*	EC096 /10	EC302 /04	HM 46	KTE 147	5-366-08_S1_C3
acrA	Macrolide-specific efflux system	aminoglycoside; glycylcycline; macrolide; beta_lactam; acriflavin	√	√	√	√	√
acrB	Macrolide-specific efflux system						
arnA	Nucleoside-diphosphate-sugar epimerases	polymyxin	√	√	√	√	√
bacA	Undecaprenyl pyrophosphate phosphatase	bacitracin	√	√	√	√	√
bl1_EC	Class C beta-lactamase	cephalosporin	√	√	√	√	√
emrE	Multidrug resistance efflux pump	aminoglycoside			√	√	
ksgA	Dimethyladenosine transferase	kasugamycin	√	√	√	√	√
macA	Macrolide-specific efflux system	macrolide	√	√	√	√	√
macB	Macrolide-specific efflux system						
mdtF	Multidrug resistance efflux pump	doxorubicin; erythromycin	√	√	√	√	√
mdtG	Multidrug resistance efflux pump	deoxycholate; fosfomycin	√	√	√	√	√
mdtH	Multidrug resistance efflux pump						
mdtK	Multidrug resistance efflux pump	enoxacin; norfloxacin	√	√	√	√	√
mdtI	Multidrug resistance efflux pump	chloramphenicol	√	√	√	√	√
mdtM	Multidrug resistance efflux pump	chloramphenicol; acriflavine; norfloxacin			√		

mdtN	Multidrug resistance efflux pump	t_chloride; acriflavine; puromycin	√	√	√	√	√
mdtO	Multidrug resistance efflux pump						
mdpP	Multidrug resistance efflux pump						
tolC	Multidrug resistance efflux pump	aminoglycoside; glycylcycline; macrolide; beta_lactam; acriflavin	√	√	√	√	√
aadB	aminoglycoside O-nucleotidylyltransferase	tobramycin; kanamycin; sisomicin; dibekacin; gentamicin		√			
aadA1	aminoglycoside O-nucleotidylyltransferase	spectinomycin; streptomycin		√			
blaTEM-1B	beta-lactamase	beta-lactam		√			√
cml_e1	chloramphenicol efflux pump	chloramphenicol		√			
dfrA7	dihydrofolate reductase	trimethoprim					√
sul1	sulfonamide-resistant dihydropteroate synthase	sulfonamide		√			√
sul2	sulfonamide-resistant dihydropteroate synthase	sulfonamide					√

*indicates results obtained from ARDB and ResFinder. √ indicates presence of resistance gene in particular *E. coli* genome. Resistance genes in grey boxes are acquired antimicrobial resistance genes

6.4 DISCUSSION

The encounter of two clonal strains obtained from different sites in human host(s) is indeed interesting as the environment of intestinal and extraintestinal sites were known to be different especially with regards to the availability of iron and nitrogen sources (Hagan et al., 2010; Sandy & Butler, 2010). Multireplicon plasmid pEC302/04, which was only found in EC302/04, carried additional iron acquisition system and was thought to be important for the extraintestinal strain EC302/04 to survive better in iron-limited environment. In fact, ExPEC carrying IncFIIA plasmids is known to be the leading cause of several extraintestinal infections in humans and has been deemed to be a major global health threat (Fricke et al., 2008; Johnson et al., 2010; Lemaître et al., 2013; Mellata et al., 2009; Peigne et al., 2009; Pitout, 2012; Skyberg et al., 2006; Smet et al., 2010; Villa et al., 2010; Woodford et al., 2009). However, our study revealed that growth rates in iron-limited medium under *in vitro* conditions were similar for the strains with plasmid pEC302/04 (EC302/04 and the DH5α (pEC302/04) transconjugant) and strains without pEC302/04 (EC096/10 and *E. coli* DH5α). Nonetheless, the contribution of the iron acquisition systems encoded in pEC302/04 towards growth adaptation in iron-limited environments could not be ruled out until a derivative of EC302/04 cured of pEC302/04 could be obtained and the growth rates in iron-limited medium determined and compared. The EC302/04 and EC096/10 also harbored two iron acquisition systems, namely the enterobactin and Chu-type iron uptake systems. On the other hand, *E. coli* DH5α only contained genes for the enterobactin iron acquisition system. The redundancy of iron acquisition systems in the EC302/04 and EC096/10 suggests that these strains may have better adaptability in iron- limited media when compared to *E. coli* DH5α, hence their better growth rate in media supplemented with DIP.

There was also no observable difference in nitrogen utilization for both *E. coli* EC302/04 and EC096/10. Both strains displayed 61 non-synonymous SNPs (nsSNPs) and 33 synonymous SNPs (sSNPs) identified in 34 shared ORFs, thereby indicating the clonal relatedness of EC302/04 and EC096/10, taking a suggested "clone threshold" value of 200 SNPs to define clones for *E. coli* (Kaas, 2014). At this juncture, it is tempting to speculate that EC302/04 and EC096/10 were derived from the same *E. coli* clone that has somehow persisted in that particular hospital environment over the 5-year period in between their respective isolations. Nonetheless, much more clinical and epidemiological links and investigations are required to support such hypothesis. On the other hand, both strains may have a common source in the community and may not necessarily be hospital related. Disregarding their unkown epidemiological link, it is still interesting to note that despite their clonality, both strains were isolated from different isolation sites indicating the likelihood of an intestinal strain (EC096/10) causing extraintestinal infections (EC302/04).

Not surprisingly, our phylogenomic analysis which included other *E. coli* genomes from the public database also revealed that intestinal and ExPEC strains can be clonally related (Cluster B2, Figure 6.8), concurring with previous studies where intestinal tracts were shown to often serve as the reservoir for *E. coli* strains that cause extraintestinal infections (UTI in particular) (Chen et al., 2013; Foxman, 2010; Jakobsen et al., 2011; Moreno et al., 2006). Chen et al., (2013) also reported that the same ExPEC strain (determined using WGS and PM assay) was capable of colonizing and persisting in both the intestinal and urinary tracts while maintaining similar fitness characteristics and phenomic profiles at both sites. Remarkably, all 8 strains in Cluster B2 (Figure 6.8) (regardless of whether they were ExPEC or intestinal strains) belonged to uncommon STs, namely ST349, ST2914 and ST70 with only 1 to 3 loci differences being identified

among the three STs. To the best of our knowledge, studies reporting on *E. coli* belonging to these three sequence types were scarce, with only two each for ST349 and ST70 (from avian, clinical and rat samples) while no published report is available for ST2914 (Chattaway et al., 2014; Guenther et al., 2012; Haque et al., 2014; Hernandez et al., 2013; Kim et al., 2011). Five out of eight strains in Cluster B2 also belonged to an uncommon serotype O166:H15, an *E. coli* serotype that is often associated with EAEC incidence (Chattaway et al., 2014; Zhou et al., 2002). The identification of the same *E. coli* serotype or ST associated with different pathotypes (i.e., EAEC, ExPEC, and intestinal strains) highlighted the complexity of *E. coli* strains since certain STs and serotypes were often found to be associated with specific pathotypes (Köhler & Dobrindt, 2011; Russo & Johnson, 2000). Further screening for STs 349, 2914 and 70 in the public database led to the discovery of more ExPEC or intestinal strains belonging to these STs (Table 6.5), providing evidence that further support the intrinsic extraintestinal virulence potential of intestinal *E. coli* strains belonging to these rare genotypes.

The genetic similarities and dissimilarities among the pathogenic *E. coli* strains of the same sequence types, such as the ST131, are also often of global interest due to its frequent association with public health issues (Mcnally et al., 2013; Price et al., 2013). The discovery of *E. coli* strains of rare genotypes (clonal complex ST349 and serotypes O166:H15) with intrinsic extraintestinal virulence potential in this study prompted further investigations in regards to their fitness traits such as the virulence and resistance determinants. Overall, all 5 strains of the clonal complex ST349 (i.e. EC096/10, EC302/04, HM46, KTE147 and 5-366-08_S1_C3) have genome sizes and GC content ranging from 4.75 to 5.44 Mb and 50.6 – 50.8%, respectively. Besides EC302/04 and EC096/10 that were obtained from Malaysia, the remaining 3 strains

were isolated from geographically distinct regions including United States of America (USA), Tanzania and Denmark (Subashchandrabose et al., 2013), indicating the sporadic nature of these strains. These 5 strains are also predicted to have different antimicrobial susceptibility profiles based on the resistance genotypes, and may be multidrug- or non-multidrug-resistant; indicating that *E. coli* strains of the same clonal complex ST349 may not associate with specific resistance phenotypes, a phenomenon which contradicts with the cephalosporin and fluoroquinolone resistant ST131 clones (Lee et al., 2010; Platell et al., 2011). In addition, the five strains also differed in their respective plasmid replicon profiles. Despite belonging to the same serotypes and clonal complex, marked genetic variation such as the resistance and plasmid replicon profiles were observed among the 5 *E. coli* strains of the same rare genotypes. Nonetheless, the intrinsic virulence potential in these strains could not be disregarded.

6.5 CONCLUSION

In this study, the first *E. coli* genome, EC302/04 that was obtained from tracheal aspirate, an extraintestinal site from human host was presented. Mobile genetic elements, such as the plasmid pEC302/04 which harboured resistance and virulence determinants appeared to be the main genetic material contributing to the diversity of two clonally related *E. coli* strains, i.e. the EC302/04 and EC096/10. Nonetheless, the phenotypic characteristics of both strains were highly similar. Our study showed an overall agreement between the phylogenies inferred using the phylogenomic tree with *E. coli* serotypes, STs and virulence profiles, with the exception of *E. coli* pathotypes as *E. coli* of different pathotypes are found to cluster together in the phylogenomic tree. Using the same approach, two distinct ExPEC clades which encompass classical ExPEC and intestinal strains that shared similar phylogenetic signals were unveiled. Nonetheless, it

is necessary to take note that since infection is a host-pathogen interaction, it very much dependent on the predisposition of an individual to infection.

Fortuitously, this study also unveiled ExPECs with rare genotypes, indicating the importance of a genomic approach in risk assessment and detection of possible emergence of bacterial pathogens. Nonetheless, the *E. coli* strains of the same rare genotypes (clonal complex ST349 and serotype O166:H15) with intrinsic extraintestinal virulence potential seems to be sporadic in nature due to the identified genetic variations and the epidemiological data of the five strains. Overall, the high-resolution genomic approach is deemed useful for investigating ExPECs and may be an essential microbiological diagnostic tool in the near future.

CHAPTER 7: CONCLUSION

Antimicrobial resistance poses serious public health threat as it causes therapeutic failure and hence reduces the efficiency in combating infectious diseases. Increasing cases of antimicrobial resistance have been reported worldwide and Malaysia is no exception (MOH, 2014), with *E. coli* resistant to extended-spectrum cephalosporins (ESC) being recognized as one of the drug-resistant threats (CDC, 2013; MOH, 2014). One of the steps needed for effective infection control is to monitor the resistance trends of the infectious agents of interest in the respective region (CDC, 2013). Epidemiological investigation is also a must in tracking the source of resistant *E. coli* as well as determining the genetic diversity of various clinical *E. coli* strains (Chapters 3 and 4).

In Chapter 3, the detailed characterization of 110 *E. coli* strains isolated from pediatric wards of a tertiary hospital in the state of Johor, Malaysia revealed that CTX-M is indeed the most prevalent ESBL genes that confer resistance to extended-cephalosporins in the local *E. coli* strains, with CTX-M-15 identified as the predominant subtype. Our findings concurred with that of other studies done in Malaysia (Lim et al., 2009; Sekawi et al., 2008) as well as globally (Al-Agamy et al., 2014; Sheng et al., 2013; Trang et al., 2013), where CTX-M-15 gene is indeed the predominant ESBL subtype. Although class 1 integron remained the major class of integron in the Malaysian *E. coli* clinical strains, no gene cassette encoding ESBL genes was found. Nonetheless, only certain CTX-M subtypes, such as the CTX-M-2 and CTX-M-9 are associated with integrons, in particular the complex integrons (Eckert et al., 2006). Using PFGE as a molecular typing tool, few clonal CTX-M-15-positive strains with indistinguishable pulsotypes were identified. Nonetheless, the CTX-M-15-producing pandemic O25-ST131 clone was not detected among the CTX-M-15-producing clinical

strains of *E. coli*. This is not surprising as *E. coli* of other ST clones may also play a role in the clonal dissemination of CTX-M-15-producing *E. coli*, although this may take place to a lesser extent compared to ST131 (Pitout, 2012; Riley, 2014). This is also the first report of a CTX-M-27-producing *E. coli* strain from Malaysia, a CTX-M subtype that is rarely found and has been reported in certain countries including Saudi Arabia, Canada and Korea (Al-Agamy et al., 2014; Pitout et al., 2007; Song et al., 2009).

To have a better understanding on the mechanisms of dissemination of CTX-M genes, the study was further expanded to characterize and determine the prevalence of CTX-M-encoding genes in *E. coli* strains isolated from different sources (clinical, zoonotic and environment samples) between 2002 to 2011 from Malaysia (Chapter 4). It was found that all CTX-M-producing *E. coli* strains were isolated from clinical samples, indicating the importance of clinical specimens as the reservoir of CTX-M genes. By expanding the pool of *E. coli* strains and CTX-M subtypes to be examined, we have detected the presence of the pandemic *E. coli* O25-ST131 clone, which is the first report of this clone in Malaysia. Besides the ST131, *E. coli* of other sequence types, such as the ST349, ST69, ST354, ST1171, ST10, ST117 and ST405 were also found to harbour CTX-M genes, with some of these STs recognized as the dominant clonal groups in other countries responsible for the dissemination of CTX-M genes (Kim et al., 2011; Coque et al., 2008; Nicolas-Chanoine et al., 2013). Besides being disseminated via clonal expansion, CTX-M genes were also found associated with various mobile genetic elements. Further analysis of the genetic environment of CTX-M genes revealed a variety of mobile elements, with insertion sequences IS*Ecp1*, IS*26* and IS*903* as the main genetic elements associated with CTX-M genes, concurring with reports from previous studies (Dhanji et al., 2011; Eckert et al., 2006; Poirel et al., 2012). Nonetheless, the genetic environment of some CTX-M genes were non-determinable as

PCR detection of specific neighbouring regions yielded no amplicon, indicating the likelihood of these CTX-M genes associated with rare or unreported genetic environments.

Whole genome sequencing enabled high resolution genomic analyses to be carried out, whether it is for a specific genomic feature of interest (Chapter 5) or as part of a phylogenomic analysis of *E. coli* populations (Chapter 6). In Chapter 5, whole genome sequence analysis was used to investigate the genetic environment of *dnd* operons (*dndBCDE*), a cluster of genes which is responsible for the phosphorothioation modification of DNA (Xu et al., 2010). Such DNA modification yielded a DNA degradation (Dnd) phenotype in several *E. coli* strains, rendering these strains nontypeable by PFGE and thus this could result in incomplete molecular epidemiological data (Chapter 3). A PCR assay targeting the *dndBCDE* genes was then developed to specifically detect the presence of these genes in *E. coli*. The PCR assay coupled with the use of thiourea in PFGE was shown to improve the detection and typeability of Dnd$^+$ *E. coli* strains. Large amounts of whole genome sequence data that are available for *E. coli* enabled us to perform a comprehensive comparative genomic analysis of the *E. coli*-encoded *dnd* operon leading to the discovery of their location within genomic islands (GIs). Unlike *Mycobacterium abscessus* which was reported to have conserved regions flanking their *dnd*-encoding GIs (Howard et al., 2013), the immediate genetic environment and *dnd*-encoding GIs in the Dnd$^+$ *E. coli* strains were highly diverse, with a total of 7 types of genetic organizations identified so far. Fortuitously, *dnd* operons were more often found in pathogenic *E. coli,* providing a possible linkage of the presence of the *dnd* operons with pathogenicity in *E. coli* strains.

In chapter 6, whole genome sequencing was carried out on two *E. coli* strains, EC302/04 and EC096/10, which displayed nearly identical pulsed-field profiles (with

only a single band difference) and were isolated from the same hospital over a gap of 5 years. To the best of our knowledge, EC302/04 genome is also the first whole genome sequence of an *E. coli* obtained from tracheal aspirate, an extraintestinal site from the human host. Detailed genome sequence analysis revealed that a large conjugative plasmid, designated pEC302/04, appeared to be the main genetic material contributing to the diversity of two clonally related *E. coli* strains. Besides carrying antibiotic resistance genes which conferred multidrug resistance phenotype to its host, pEC302/04 also harboured two iron acquisition systems (SitABCD and IutA-IucABCD) which may be important for the adaptation of the EC302/04 host in the iron-limited extraintestinal sites. Nonetheless, our study revealed no significant differences in the growth rates in iron-limited medium under *in vitro* conditions for both EC096/10 and EC302/04. High resolution phylogenomic analysis was then used to investigate the genetic relationship of EC302/04 and EC096/10 with 38 genomes of different *E. coli* pathotypes that were extracted from the public database. Two distinct ExPEC clades, which encompassed classical ExPEC as well as intestinal strains that shared similar phylogenetic signals with existing ExPECs were identified. Although the opportunistic nature of ExPEC had been reported (Guenther et al., 2012; Russo & Johnson, 2000), the identification of intestinal strains with phylogenetic signals similar to existing ExPEC enabled us to identify possible commensal-like strains with extraintestinal virulence potential. Using the same approach, some virulence factors were more commonly found in the ExPEC lineages but not in the commensal and IPEC, which may be useful in identifying potential vaccine candidates which are effective against ExPEC and intestinal strains with extraintestinal virulence potential for future studies. The analyses also unveiled *E. coli* strains of rare genotypes (specifically, clonal complex ST349 and serotype O166:H15) that possessed intrinsic extraintestinal virulence potential. Despite the tight phylogenetic relationship of these *E. coli* strains sharing the same rare genotypes (as

observed in the phylogenomic tree), they are sporadic in nature as they were isolated in different countries and periods and had substantial variations in the resistotypes. This finding is in contrast with *E. coli* ST131 clones which are often associated with fluoroquinolone and extended-cephalosporin resistance (Lee et al., 2010; Platell et al., 2011). Overall, the high-resolution genomic approach is useful for investigating ExPECs and also serves as a proof of concept in detecting possible emergence of ExPEC clones. However, this method may not be able to identify the `true' *E. coli* pathotypes in causing extraintestinal infections as different pathotypes may share similar phylogenetic signals and are clustered in the same clade.

Overall, this study has provided further insights into the resistance mechanisms and genetic diversity of *E. coli* strains using both phenotypic and genotypic approaches. Nonetheless, there are several limitations in the study. The resistance rates in this study should be interpreted with caution as the *E. coli* samples may not represent the general resistance situation at the national level due to limited and possible skewed samples (isolates obtained from a single hospital, for instance), which may lead to underestimation or overestimation of resistance rates. Due to computational limitations, a limited number of global *E. coli* genomes have also been included into the phylogenomic analysis. Although this study revealed limitations with regards on the samples and representativeness of data, it has nevertheless provided useful insights into the resistance mechanisms and genetic diversity of *E. coli*, creating a basis for further development in the field.

For future resistance and epidemiological investigations, a large number of Malaysia *E. coli* strains with proper sampling approach from various sources and regions is necessary to provide data that can represent the situation in Malaysia. The biological function(s) of phosphorothioated DNA are also worthy of further exploration in

particular the possible linkage between the carriage of the *dnd* operons and *E. coli* pathogenesis. In addition, large scale WGS studies of the highly versatile *E. coli* can also be carried out, to expand our current knowledge on various *E. coli* pathotypes, for the interest of both fundamental research and public health concerns.

REFERENCES

Al-Agamy, M. H., Shibl, A. M., Hafez, M. M., Al-Ahdal, M. N., Memish, Z. A., & Khubnani, H. (2014). Molecular characteristics of extended-spectrum β-lactamase-producing *Escherichia coli* in Riyadh: emergence of CTX-M-15-producing E. coli ST131. *Annals of Clinical Microbiology and Antimicrobials*, *13*(1), 4.

Alqasim, A., Emes, R., Clark, G., Newcombe, J., La Ragione, R., & McNally, A. (2014). Phenotypic microarrays suggest *Escherichia coli* ST131 is not a metabolically distinct lineage of extra-intestinal pathogenic *E. coli*. *PloS One*, *9*(2), e88374.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Alves, M. S., Pereira, A., Araújo, S. M., Castro, B. B., Correia, A. C. M., & Henriques, I. (2014). Seawater is a reservoir of multi-resistant *Escherichia coli*, including strains hosting plasmid-mediated quinolones resistance and extended-spectrum beta-lactamases genes. *Frontiers in Microbiology*, *5*, 426.

An, X., Xiong, W., Yang, Y., Li, F., Zhou, X., Wang, Z., … Liang, J. (2012). A Novel Target of IscS in *Escherichia coli*: Participating in DNA Phosphorothioation. *PLoS ONE*, *7*.

Archambault, M., Petrov, P., Hendriksen, R. S., Asseva, G., Bangtrakulnonth, A., Hasman, H., & Aarestrup, F. M. (2006). Molecular characterization and occurrence of extended-spectrum beta-lactamase resistance genes among *Salmonella enterica* serovar Corvallis from Thailand, Bulgaria, and Denmark. *Microbial Drug Resistance (Larchmont, N.Y.)*, *12*(3), 192–8.

Aziz, R. K., Bartels, D., Best, A. A, DeJongh, M., Disz, T., Edwards, R. A, … Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(9), 75.

Baraniak, A., Fiett, J., Hryniewicz, W., Nordmann, P., & Gniadkowski, M. (2002). Ceftazidime-hydrolysing CTX-M-15 extended-spectrum beta-lactamase (ESBL) in Poland. *The Journal of Antimicrobial Chemotherapy*, *50*, 393–396.

Bauernfeind, A., Stemplinger, I., Jungwirth, R., & Giamarellou, H. (1996). Characterization of the plasmidic beta-lactamase CMY-2, which is responsible for cephamycin resistance. *Antimicrobial Agents and Chemotherapy*, *40*(1), 221–4.

Bekal, S., Brousseau, R., Masson, L., Prefontaine, G., Fairbrother, J., & Harel, J. (2003). Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. *Journal of Clinical Microbiology*, *41*(5), 2113–25.

Belkum, A. Van, Tassios, P. T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N. K., … Struelens, M. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and Infectious Diseases*, *13*, 1–46.

Bellanger, X., Payot, S., Leblond-Bourget, N., & Guédon, G. (2014). Conjugative and

mobilizable genomic islands in bacteria: Evolution and diversity. *FEMS Microbiology Reviews*, *38*, 720–760.

Bertelli, C., & Greub, G. (2013). Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clinical Microbiology and Infection*, *19*(9), 803–13.

Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B., & van Nimwegen, E. (2014). Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Molecular Biology and Evolution*, *31*(5), 1077–88. h

Blount, Z. D. (2015). The unexhausted potential of E. coli. *eLife*, *4*, 1–12.

Blum, S. E., Heller, E. D., Sela, S., Elad, D., Edery, N., & Leitner, G. (2015). Genomic and Phenomic Study of Mammary Pathogenic *Escherichia coli*. *PloS One*, *10*(9), e0136387.

Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001). Phenotype MicroArrays for High-Throughput Phenotypic Testing and Assay of Gene Function. *Genome Research*, *11*, 1246–1255.

Bondy-Denomy, J., & Davidson, A. R. (2014). To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends in Microbiology*, *22*(4), 218–25.

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

Bonnet, R., Recule, C., Baraduc, R., Chanal, C., Sirot, D., De Champs, C., & Sirot, J. (2003). Effect of D240G substitution in a novel ESBL CTX-M-27. *The Journal of Antimicrobial Chemotherapy*, *52*(1), 29–35.

Boyd, E. F., Almagro-Moreno, S., & Parent, M. A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in Microbiology*, *17*, 47–53.

Boyer, E., Bergevin, I., Malo, D., Gros, P., & Cellier, M. F. M. (2002). Acquisition of Mn (II) in Addition to Fe (II) Is Required for Full Virulence of Salmonella enterica Serovar Typhimurium. *Infection and Immunity*, *70*(11), 6032–6042.

Bradford, P. A. (2001). Extended-Spectrum β-Lactamases in the 21st Century : Characterization , Epidemiology , and Detection of This Important Resistance Threat. *Clinical Microbiology Reviews*, *14*(4), 933–951.

Brolund, A., Franzén, O., Melefors, O., Tegmark-Wisell, K., & Sandegren, L. (2013). Plasmidome-analysis of ESBL-producing *Escherichia coli* using conventional typing and high-throughput sequencing. *PloS One*, *8*(6), e65793.

Brown, T. (2002). *Understanding a Genome Sequence. In Genome.* (2nd ed.). Oxford: Wiley-Liss.

Bush, K., Jacoby, G. a, & Medeiros, a a. (1995). A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrobial Agents and Chemotherapy*, *39*, 1211–1233.

Cambray, G., Guerout, A.-M., & Mazel, D. (2010). Integrons. *Annual Review of Genetics*, *44*, 141–66.

Cantón, R., & Coque, T. M. (2006). The CTX-M beta-lactamase pandemic. *Current Opinion in Microbiology*, *9*(5), 466–75.

Cantón, R., González-Alba, J. M., & Galán, J. C. (2012). CTX-M enzymes: origin and diffusion. *Frontiers in Microbiology*, *3*(110), 1–19.

Cao, B., Chen, C., DeMott, M. S., Cheng, Q., Clark, T. a, Xiong, X., … Dedon, P. C. (2014). Genomic mapping of phosphorothioates reveals partial modification of short consensus sequences. *Nature Communications*, *5*, 3951.

Cao, X., Cavaco, L., Lv, Y., Li, Y., Zheng, B., Wang, P., … Aarestrup, F. (2011). Molecular characterization and antimicrobial susceptibility test- ing of Escherichia coli isolates from patients with urinary tract infections in 20 Chinese hospitals. *Journal of Clinical Microbiology*, *49*, 2496–2501.

Carattoli, A. (2009). Resistance plasmid families in Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy*, *53*(6), 2227–38.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., … Hasman, H. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*, *58*(7), 3895–903.

Carlos, C., Pires, M. M., Stoppe, N. C., Hachich, E. M., Sato, M. I. Z., Gomes, T. A. T., … Ottoboni, L. M. M. (2010). *Escherichia coli* phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. *BMC Microbiology*, *10*(1), 161.

Centre for Disease Control and Prevention (CDC). (2011). *E. coli* Infection and Food Safety. Retrieved from http://www.cdc.gov/features/ecoliinfection/

Centre for Disease Control and Prevention (CDC) (2013). Antibiotic resistance threats in the United States 2013. Retrieved from http://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf

Centre for Disease Control and Prevention (CDC) (2014). General Information | *E.coli* | CDC. Retrieved October 10, 2015, from http://www.cdc.gov/ecoli/general/

Chan, W. T., Moreno-Córdoba, I., Yeo, C. C., & Espinosa, M. (2012). Toxin-antitoxin genes of the Gram-positive pathogen *Streptococcus pneumoniae*: so few and yet so many. *Microbiology and Molecular Biology Reviews : MMBR*, *76*, 773–91.

Chanawong, A., Lulitanond, A., Kaewkes, W., Lulitanond, V., Srigulbutr, S., &

Homchampa, P. (2007). CTX-M extended-spectrum β-lactamases among clinical isolates of Enterobacteriaceae in a Thai university hospital. *Southeast Asian Journal of Tropical Medicine and Public Health*, *38*, 493–500.

Chattaway, M. A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., … Wain, J. (2014). Evidence of evolving extraintestinal enteroaggregative *Escherichia coli* ST38 Clone. *Emerging Infectious Diseases*, *20*(11), 1935–1937.

Chattaway, M. A., Jenkins, C., Rajendram, D., Cravioto, A., Talukder, K. A., Dallman, T., … Wain, J. (2014). Enteroaggregative *Escherichia coli* have evolved independently as distinct complexes within the *E. coli* population with varying ability to cause disease. *PloS One*, *9*(11), e112967.

Chen, L., Xiong, Z., Sun, L., Yang, J., & Jin, Q. (2012). VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research*, *40*(Database issue), D641–5.

Chen, S. L., Wu, M., Henderson, J. P., Hooton, T. M., Hibbing, M. E., Hultgren, S. J., & Gordon, J. I. (2013). Genomic diversity and fitness of *E . coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Science Translational Medicine*, *5*(184), 1–14.

Claeys, G., Verschraegen, G., de Baere, T., & Vaneechoutte, M. (2000). PER-1 beta-lactamase-producing *Pseudomonas aeruginosa* in an intensive care unit. *The Journal of Antimicrobial Chemotherapy*, *45*(6), 924–5.

Clermont, O., Bonacorsi, S., & Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Applied and Environmental Microbiology*, *66*(10), 4555–4558.

Clermont, O., Christenson, J. K., Denamur, E., & Gordon, D. M. (2013). The Clermont *Escherichia coli* phylo-typing method revisited : improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports*, *5*(1), 58–65.

Clermont, O., Dhanji, H., Upton, M., Gibreel, T., Fox, A., Boyd, D., … Denamur, E. (2009). Rapid detection of the O25b-ST131 clone of *Escherichia coli* encompassing the CTX-M-15-producing strains. *The Journal of Antimicrobial Chemotherapy*, *64*(2), 274–7.

CLSI. (2015). Performance standards for antimicrobial susceptibility testing; twenty-fifth informational supplement. CLSI document M100-S25. Wayne, PA: Clinical and Laboratory Standard Institute.

Coburn, B., Sekirov, I., & Finlay, B. B. (2007). Type III secretion systems and disease. *Clinical Microbiology Reviews*, *20*, 535–549.

Coimbra, R. S., Grimont, F., Lenormand, P., Burguière, P., Beutin, L., & Grimont, P. A. (2000). Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). *Research in Microbiology*, *151*(8), 639–54.

Conceição, T., Brízio, A., Duarte, A., Lito, L. M., Cristino, J. M., & Salgado, M. J. (2005). First description of CTX-M-15-producing *Klebsiella pneumoniae* in Portugal. *Antimicrobial Agents and Chemotherapy*, *49*(1), 477–8.

Coque, T. M., Novais, Â., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., … Nordmann, P. (2008). Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum β-Lactamase CTX-M-15. *Emerging Infectious Diseases*, *14*(2), 195–200.

Corkill, J. E., Graham, R., Hart, C. A., & Stubbs, S. (2000). Pulsed-field gel electrophoresis of degradation-sensitive DNAs from *Clostridium difficile* PCR ribotype 1 strains. *Journal of Clinical Microbiology*, *38*(7), 2791–2.

Cronan, J. E. (2014). *Escherichia coli as an Experimental Organism*. Chichester, UK: John Wiley & Sons, Ltd.

D'Andrea, M. M., Arena, F., Pallecchi, L., & Rossolini, G. M. (2013). CTX-M-type β-lactamases: a successful story of antibiotic resistance. *International Journal of Medical Microbiology*, *303*(6-7), 305–17.

Dancer, S. J. (2001). The problem with cephalosporins. *Journal of Antimicrobial Chemotherapy*, *48*(4), 463–478.

Datta, N., & Kontomichalou, P. (1965). Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature*, *208*, 239–241.

De, C. C., Poirel, L., Bonnet, R., Sirot, D., Chanal, C., Sirot, J., & Nordmann, P. (2002). Prospective survey of beta-lactamases produced by ceftazidime- resistant *Pseudomonas aeruginosa* isolated in a French hospital in 2000. *Antimicrob.Agents Chemother.*, *46*, 3031–3034.

Dhanji, H., Patel, R., Wall, R., Doumith, M., Patel, B., Hope, R., … Woodford, N. (2011). Variation in the genetic environments of bla(CTX-M-15) in *Escherichia coli* from the faeces of travellers returning to the United Kingdom. *The Journal of Antimicrobial Chemotherapy*, *66*(5), 1005–12.

Dhillon, B. K., Laird, M. R., Shay, J. A, Winsor, G. L., Lo, R., Nizam, F., … Brinkman, F. S. L. (2015). IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Research*, *43*(W1), W104–8.

Donnenberg, M. S., & Whittam, T. S. (2001). Bacterial polymorphisms pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. *The Journal of Clinical Investigation*, *107*(5), 539–548.

Eckert, C., Gautier, V., & Arlet, G. (2006). DNA sequence analysis of the genetic environment of various blaCTX-M genes. *The Journal of Antimicrobial Chemotherapy*, *57*(1), 14–23.

Edelstein, M., Pimkin, M., Palagin, I., Edelstein, I., & Stratchounski, L. (2003). Prevalence and molecular epidemiology of CTX-M extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* in Russian

hospitals. *Antimicrobial Agents and Chemotherapy*, *47*(12), 3724–32.

EFSA. (2011). *Shiga toxin-producing E . coli (STEC) O104 : H4 2011 outbreaks in Europe : European Food Safety Authority* (Vol. 9). Parma, Italy.

Ensor, V. M., Livermore, D. M., & Hawkey, P. M. (2007). A novel reverse-line hybridization assay for identifying genotypes of CTX-M-type extended-spectrum beta-lactamases. *The Journal of Antimicrobial Chemotherapy*, *59*(3), 387–95.

Escherich, T. (1886). Die Darmbakterien des Säuglings und ihre Beziehungen zur Physiologie der Verdauung. *Stuttgart : F. Enke*.

Fontecave, M., & Ollagnier-de-Choudens, S. (2008). Iron-sulfur cluster biosynthesis in bacteria: Mechanisms of cluster assembly and transfer. *Archives of Biochemistry and Biophysics*, *474*, 226–237.

Fournier, P.-E., Dubourg, G., & Raoult, D. (2014). Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Medicine*, *6*(11), 114.

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

Fricke, W. F., Wright, M. S., Lindell, A. H., Harkins, D. M., Baker-Austin, C., Ravel, J., & Stepanauskas, R. (2008). Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5. *Journal of Bacteriology*, *190*(20), 6779–94.

Galardini, M., Biondi, E. G., Bazzicalupo, M., & Mengoni, A. (2011). CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code for Biology and Medicine*, *6*(1), 11.

Germon, P., Roche, D., Melo, S., Mignon-Grasteau, S., Dobrindt, U., Hacker, J., … Moulin-Schouleur, M. (2007). tDNA locus polymorphism and ecto-chromosomal DNA insertion hot-spots are related to the phylogenetic group of *Escherichia coli* strains. *Microbiology*, *153*, 826–837.

Gilmour, M., Thomson, N., Sanders, M., Parkhill, J., & Taylor, D. (2004). The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid*, *52*, 182–202.

Gómez-Duarte, O. G., Bai, J., & Newell, E. (2009). Detection of *Escherichia coli*, *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica*, *Vibrio cholerae*, and *Campylobacter* spp. enteropathogens by 3-reaction multiplex polymerase chain reaction. *Diagnostic Microbiology and Infectious Disease*, *63*(1), 1–9.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, *35*(Web Server issue), W52–7.

Guenther, S., Bethe, A., Fruth, A., Semmler, T., Ulrich, R. G., Wieler, L. H., & Ewers, C.

(2012). Frequent combination of antimicrobial multiresistance and extraintestinal pathogenicity in *Escherichia coli* isolates from urban rats (Rattus norvegicus) in Berlin, Germany. *PloS One*, *7*(11), e50331.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–21.

Hagan, E. C., Lloyd, A. L., Rasko, D. a, Faerber, G. J., & Mobley, H. L. T. (2010). *Escherichia coli* global gene expression in urine from women with urinary tract infection. *PLoS Pathogens*, *6*(11), e1001187.

Hall, B. G., Acar, H., Nandipati, A., & Barlow, M. (2014). Growth rates made easy. *Molecular Biology and Evolution*, *31*(1), 232–8.

Hanson, N. D. (2003). AmpC beta-lactamases: what do we need to know for the future? *The Journal of Antimicrobial Chemotherapy*, *52*(1), 2–4.

Haque, A., Yoshizumi, A., Saga, T., Ishii, Y., & Tateda, K. (2014). ESBL-producing Enterobacteriaceae in environmental water in Dhaka, Bangladesh. *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy*, *20*(11), 735–7.

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C. A., Frimodt-Møller, N., & Aarestrup, F. M. (2014). Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *Journal of Clinical Microbiology*, *52*(1), 139–46.

Hawkey, P. M. (2008). Prevalence and clonality of extended-spectrum beta-lactamases in Asia. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *14 Suppl 1*, 159–65.

Hawkey, P. M., & Jones, A. M. (2009). The changing epidemiology of resistance. *The Journal of Antimicrobial Chemotherapy*, *64 Suppl 1*, i3–10.

Hazen, T. H., Sahl, J. W., Fraser, C. M., Donnenberg, M. S., Scheutz, F., & Rasko, D. A. (2013). Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), 12810–5.

He, X., Ou, H. Y., Yu, Q., Zhou, X., Wu, J., Liang, J., … Deng, Z. (2007). Analysis of a genomic island housing genes for DNA S-modification system in *Streptomyces lividans* 66 and its counterparts in other distantly related bacteria. *Molecular Microbiology*, *65*, 1034–1048.

Heiman, K. E., Mody, R. K., Johnson, S. D., Griffin, P. M., & Gould, L. H. (2015). Escherichia coli O157 Outbreaks in the United States, 2003–2012. *Emerging Infectious Diseases*, *21*(8), 1293–1301.

Hernandez, J., Johansson, A., Stedt, J., Bengtsson, S., Porczak, A., Granholm, S., … Drobni, M. (2013). Characterization and comparison of extended-spectrum β-lactamase (ESBL) resistance genotypes and population structure of *Escherichia coli* isolated from Franklin's gulls (Leucophaeus pipixcan) and humans in Chile. *PloS One*, *8*(9), e76150.

Hill, V. T., Townsend, S. M., Arias, R. S., Jenabi, J. M., Gomez-gonzalez, I., Shimada, H., & Badger, J. L. (2004). TraJ-Dependent *Escherichia coli* K1 Interactions with Professional Phagocytes Are Important for Early Systemic Dissemination of Infection in the Neonatal Rat. *Infection and Immunity*, *72*(1), 478–488.

Ho, W. S., Gan, H. M., Yap, K. P., Balan, G., Yeo, C. C., & Thong, K.-L. (2012). Genome sequence of multidrug-resistant *Escherichia coli* EC302/04, isolated from a human tracheal aspirate. *Journal of Bacteriology*, *194*(23), 6691–2.

Ho, W. S., Tan, L. K., Ooi, P. T., Yeo, C. C., & Thong, K. L. (2013). Prevalence and characterization of verotoxigenic-*Escherichia coli* isolates from pigs in Malaysia. *BMC Veterinary Research*, *9*, 109.

Hoffman, J. A., Badger, J. L., Zhang, Y., Huang, S. H., & Kim, K. S. (2000). *Escherichia coli* K1 aslA contributes to invasion of brain microvascular endothelial cells in vitro and in vivo. *Infection and Immunity*, *68*, 5062–5067.

Honish, L., Predy, G., Hislop, N., Chui, L., Kowalewska-Grochowska, K., Trottier, L., … Zazulak, I. (2005). An outbreak of *E. coli* O157:H7 hemorrhagic colitis associated with unpasteurized gouda cheese. *Canadian Journal of Public Health =*, *96*(3), 182–4.

Howard, S. T., Newman, K. L., McNulty, S., Brown-Elliott, B. A., Vasireddy, R., Bridge, L., & Wallace, R. J. (2013). Insertion site and distribution of a genomic island conferring DNA phosphorothioation in the *Mycobacterium abscessus* complex. *Microbiology (Reading, England)*, *159*(Pt 11), 2323–32.

Huang, Z., Mao, P., Chen, Y., Wu, L., & Wu, J. (2004). Study on the molecular epidemiology of SHV type beta-lactamase-encoding genes of multiple-drug-resistant acinetobacter baumannii. *Zhonghua Liu Xing Bing Xue Za Zhi*, *25*(5), 425–7.

Ideses, D., Gophna, U., Paitan, Y., Chaudhuri, R. R., Pallen, M. J., & Ron, E. Z. (2005). A degenerate type III secretion system from septicemic *Escherichia coli* contributes to pathogenesis. *Journal of Bacteriology*, *187*, 8164–8171.

Jacoby, G. A. (2009). AmpC beta-lactamases. *Clinical Microbiology Reviews*, *22*(1), 161–82, Table of Contents. http://doi.org/10.1128/CMR.00036-08

Jacoby, G. A., & Medeiros, A. A. (1991). More extended-spectrum beta-lactamases. *Antimicrobial Agents and Chemotherapy*, *35*(9), 1697–704.

Jacoby, G. A., & Munoz-Price, L. S. (2005). The new beta-lactamases. *The New England Journal of Medicine*, *352*(4), 380–391.

Jakobsen, L., Garneau, P., Kurbasic, A., Bruant, G., Stegger, M., Harel, J., … Frimodt-Møller, N. (2011). Microarray-based detection of extended virulence and antimicrobial resistance gene profiles in phylogroup B2 *Escherichia coli* of human, meat and animal origin. *Journal of Medical Microbiology*, *60*, 1502–11.

Jarlier, V., Nicolas, M. H., Fournier, G., & Philippon, A. (1988). Extended broad-spectrum beta-lactamases conferring transferable resistance to newer beta-lactam agents in Enterobacteriaceae: hospital prevalence and susceptibility patterns. *Reviews of Infectious Diseases*, *10*(4), 867–78.

Jauréguy, F., Carbonnelle, E., Bonacorsi, S., Clec'h, C., Casassus, P., Bingen, E., … Lortholary, O. (2007). Host and bacterial determinants of initial severity and outcome of *Escherichia coli* sepsis. *Clinical Microbiology and Infection*, *13*(9), 854–62.

Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *Journal of Clinical Microbiology*, *53*(8), 2410–26.

Johnson, J. R., Goullet, P., Picard, B., Moseley, S. L., Roberts, P. L., & Stamm, W. E. (1991). Association of carboxylesterase B electrophoretic pattern with presence and expression of urovirulence factor determinants and antimicrobial resistance among strains of *Escherichia coli* that cause urosepsis. *Infection and Immunity*, *59*, 2311–5.

Johnson, J. R., Johnston, B., Clabots, C., Kuskowski, M. a, & Castanheira, M. (2010). *Escherichia coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *51*(3), 286–94.

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. *Antimicrobial Agents and Chemotherapy*, *53*(7), 2733–9.

Johnson, J. R., & Russo, T. a. (2002). Extraintestinal pathogenic *Escherichia coli* : "The other bad E coli ." *Journal of Laboratory and Clinical Medicine*, *139*(3), 155–162.

Johnson, T. J., Jordan, D., Stell, A. L., Bell, N. P., Yvonne, M., Alarcón, C. F., … Ferna, C. (2010). Sequence analysis and characterization of a transferable hybrid plasmid encoding multidrug resistance and enabling zoonotic potential for extraintestinal *Escherichia coli* sequence analysis and characterization of a transferable hybrid plasmid encoding mul. *Infection and Immunity*, *78*(5), 1931–1942.

Johnson, T. J., & Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiology and Molecular Biology Reviews : MMBR*, *73*(4), 750–74.

Johnson, T. J., Thorsness, J. L., Anderson, C. P., Lynne, A. M., Foley, S. L., Han, J., … Singer, R. S. (2010). Horizontal gene transfer of a ColV plasmid has resulted in a dominant avian clonal type of *Salmonella enterica* serovar Kentucky. *PloS One*,

5(12), e15524.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, *33*(2), 376–93.

Kaas, R. S. (2014). *Whole Genome Epidemiological Typing of Escherichia coli*.

Kanamori, H., Navarro, R. B., Yano, H., Sombrero, L. T., Capeding, M. R. Z., Lupisan, S. P., … Kaku, M. (2011). Molecular characteristics of extended-spectrum β-lactamases in clinical isolates of Enterobacteriaceae from the Philippines. *Acta Tropica*, *120*, 140–145.

Kaper, J., Nataro, J., & Mobley, H. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, *2*(2), 123–40.

Kiiru, J., Butaye, P., Goddeeris, B. M., & Kariuki, S. (2013). Analysis for prevalence and physical linkages amongst integrons, IS*Ecp1*, IS*CR1*, Tn*21* and Tn*7* encountered in *Escherichia coli* strains from hospitalized and non-hospitalized patients in Kenya during a 19-year period (1992-2011). *BMC Microbiology*, *13*(1), 109.

Kim, J., Bae, I. K., Jeong, S. H., Chang, C. L., Lee, C. H., & Lee, K. (2011). Characterization of IncF plasmids carrying the blaCTX-M-14 gene in clinical isolates of *Escherichia coli* from Korea. *The Journal of Antimicrobial Chemotherapy*, *66*(6), 1263–8.

Kiratisin, P., Apisarnthanarak, A., Laesripa, C., & Saifon, P. (2008). Molecular characterization and epidemiology of extended-spectrum-beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolates causing health care-associated infection in Thailand, where the CTX-M family is endemic. *Antimicrobial Agents and Chemotherapy*, *52*(8), 2818–24.

Knothe, H., Shah, P., Krcmery, V., Antal, M., & Mitsuhashi, S. (1983). Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. *Infection*, *11*, 315–317.

Köhler, C., & Dobrindt, U. (2011). What defines extraintestinal pathogenic *Escherichia coli* ? *International Journal of Medical Microbiology*, *301*, 642–647.

Kong, R. Y. ., So, C. ., Law, W. ., & Wu, R. S. . (1999). A Sensitive and Versatile Multiplex PCR System for the Rapid Detection of Enterotoxigenic (ETEC), Enterohaemorrhagic (EHEC) and Enteropathogenic (EPEC) Strains of *Escherichia coli*. *Marine Pollution Bulletin*, *38*(12), 1207–1215.

Köser, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., … Peacock, S. J. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogens*, *8*(8), e1002824.

Kroll, J., Klinter, S., Schneider, C., Voss, I., & Steinbüchel, A. (2010). Plasmid addiction systems: perspectives and applications in biotechnology. *Microbial Biotechnology*,

3(6), 634–57.

Kulkarni, R., Dhakal, B. K., Slechta, E. S., Kurtz, Z., Mulvey, M. a, & Thanassi, D. G. (2009). Roles of putative type II secretion and type IV pilus systems in the virulence of uropathogenic *Escherichia coli*. *PloS One*, *4*(3), e4752.

Kwon, N. Y., Kim, J. D., & Pai, H. J. (2002). The resistance mechanisms of b-lactam antimicrobials in clinical isolates of *Acinetobacter baumannii*. *The Korean Journal of Internal Medicine*, *17*(2), 94–9.

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–8.

Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S., & Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Research*, *41*(17), 8034–44.

Langille, M. G. I., Hsiao, W. W. L., & Brinkman, F. S. L. (2012). Detecting genomic islands using bioinformatics approaches. *Nature Reviews Microbiology*, *8*, 373–382.

Larsen, M. V, Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, L., … Lund, O. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology*, *50*(4), 1355–1361.

Lartigue, M.-F., Poirel, L., & Nordmann, P. (2004). Diversity of genetic environment of bla(CTX-M) genes. *FEMS Microbiology Letters*, *234*(2), 201–7.

Lawley, T. ., Klimke, W. ., Gubbins, M. ., & Frost, L. . (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters*, *224*(1), 1–15.

Lee, M. Y., Choi, H. J., Choi, J. Y., Song, M., Song, Y., Kim, S.-W., … Ko, K. S. (2010). Dissemination of ST131 and ST393 community-onset, ciprofloxacin-resistant *Escherichia coli* clones causing urinary tract infections in Korea. *The Journal of Infection*, *60*, 146–53.

Leflon-Guibout, V., Blanco, J., Amaqdouf, K., Mora, A., Guize, L., & Nicolas-Chanoine, M.-H. (2008). Absence of CTX-M enzymes but high prevalence of clones, including clone ST131, among fecal *Escherichia coli* isolates from healthy subjects living in the area of Paris, France. *Journal of Clinical Microbiology*, *46*(12), 3900–5.

Lemaître, C., Mahjoub-Messai, F., Dupont, D., Caro, V., Diancourt, L., Bingen, E., … Bonacorsi, S. (2013). A conserved virulence plasmidic region contributes to the virulence of the multiresistant *Escherichia coli* meningitis strain S286 belonging to phylogenetic group C. *PloS One*, *8*(9), e74423.

Lim, K.-T., Yasin, R., Yeo, C.-C., Puthucheary, S., & Thong, K.-L. (2009). Characterization of multidrug resistant ESBL-producing *Escherichia coli* isolates from hospitals in Malaysia. *Journal of Biomedicine & Biotechnology*, *2009*,

165637.

Literacka, E., Bedenic, B., Baraniak, A., Fiett, J., Tonkic, M., Jajic-Bencic, I., & Gniadkowski, M. (2009). blaCTX-M genes in *Escherichia coli* strains from Croatian Hospitals are located in new (blaCTX-M-3a) and widely spread (blaCTX-M-3a and blaCTX-M-15) genetic structures. *Antimicrobial Agents and Chemotherapy*, *53*, 1630–5.

Liu, B., & Pop, M. (2009). ARDB — Antibiotic Resistance Genes Database, *37*(October 2008), 443–447.

Livermore, D. M. (1995). Beta-lactamases in laboratory and clinical resistance. *Clinical Microbiology Reviews*, *8*, 557–584.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE : a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964.

Lukjancenko, O., Wassenaar, T. M., & Ussery, D. W. (2010). Comparison of 61 Sequenced *Escherichia coli* Genomes. *Microbial Ecology*, *60*, 708–720.

Luo, C., Walk, S. T., Gordon, D. M., Feldgarden, M., Tiedje, J. M., & Konstantinidis, K. T. (2011). Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 7200–5.

Luzzaro, F., Mantengoli, E., Perilli, M., Lombardi, G., Orlandi, V., Orsatti, a., … Toniolo, a. (2001). Dynamics of a nosocomial outbreak of multidrug-resistant Pseudomonas aeruginosa producing the PER-1 extended-spectrum β-lactamase. *Journal of Clinical Microbiology*, *39*, 1865–1870.

Machado, E., Canton, R., Baquero, F., Galan, J.-C., Rollan, A., Peixe, L., & Coque, T. M. (2005). Integron content of extended-spectrum β-lactamase-producing *Escherichia coli* strains over 12 years in a single hospital in Madrid, Spain. *Antimicrobial Agents and Chemotherapy*, *49*(5), 1823–1829.

Machado, E., Ferreira, J., Novais, A., Peixe, L., Cantón, R., Baquero, F., & Coque, T. M. (2007). Preservation of integron types among Enterobacteriaceae producing extended-spectrum beta-lactamases in a Spanish hospital over a 15-year period (1988 to 2003). *Antimicrobial Agents and Chemotherapy*, *51*(6), 2201–4.

Machado, J., Grimont, F., & Grimont, P. A. D. (2000). Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. *Research in Microbiology*, *151*, 535–546.

Mamlouk, K., Boutiba-Ben Boubaker, I., Gautier, V., Vimont, S., Picard, B., Ben Redjeb, S., & Arlet, G. (2006). Emergence and outbreaks of CTX-M beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* strains in a Tunisian hospital. *Journal of Clinical Microbiology*, *44*(11), 4049–56.

Marshall, B. M., & Levy, S. B. (2011). Food animals and antimicrobials: impacts on human health. *Clinical Microbiology Reviews*, *24*(4), 718–33.

Matsumoto, Y., Ikeda, F., Kamimura, T., Yokota, Y., & Mine, Y. (1988). Novel plasmid-mediated β-lactamase from *Escherichia coli* that inactivates oxyimino-cephalosporins. *Antimicrobial Agents and Chemotherapy*, *32*, 1243–1246.

Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nature Reviews. Microbiology*, *4*(8), 608–20.

Mcnally, A., Alhashash, F., Collins, M., Alqasim, A., Paszckiewicz, K., Weston, V., & Diggle, M. (2013). Genomic analysis of extra-intestinal pathogenic *Escherichia coli* urosepsis. *Clinical Microbiology and Infection*, *19*, E328–E334.

Mellata, M., Touchman, J. W., & Curtiss, R. (2009). Full sequence and comparative analysis of the plasmid pAPEC-1 of avian pathogenic *E. col*i chi7122 (O78:K80:H9). *PloS One*, *4*(1), e4232.

Miethke, M., & Marahiel, M. A. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiology and Molecular Biology Reviews : MMBR*, *71*, 413–451.

Ministry of Health (MOH). (2008). *Nantional Antibiotic Guideline 2008*. Available at http://apps.who.int/medicinedocs/documents/s17800en/s17800en.pdf

Ministry of Health (MOH). (2014a). *National Surveillance of Antibiotic Resistance (NSAR) Report*. Available at http://www.imr.gov.my/images/uploads/NSAR/NSAR_2014/Summary_of_Antibiotic_Resistance_2014_report.pdf

Ministry of Health (MOH). (2011). National Surveillance of Antibiotic Resistance (NSAR) Report, 1–9. Available at http://www.imr.gov.my/images/uploads/NSAR%20Archives/Summary%20of%20antibiotic%20resistance%202011.pdf

Ministry of Health (MOH). (2012). *National Surveillance of Antibiotic Resistance (NSAR) Report*. Available at http://www.imr.gov.my/images/uploads/NSAR/Summary_of_antibiotic_resistance_2012_for_website.pdf

Ministry of Health (MOH). (2013). *National Surveillance of Antibiotic Resistance (NSAR) Report*. Available at http://www.imr.gov.my/images/uploads/NSAR/NSAR_2013/Summary_of_NSAR_2013_report.pdf

Ministry of Health (MOH). (2014b). *National Antibiotic Guideline 2014*. Kuala Lumpur. Retrieved from http://www.pharmacy.gov.my/v2/sites/default/files/document-upload/national-antibiotic-guideline-2014-full-versionjun2015_1.pdf

Molina-lópez, J., Aparicio-ozores, G., Ribas-aparicio, R. M., Chávez-berrocal, M. E., Hernández-castro, R., & Ángel, H. (2010). Drug resistance, serotypes, and phylogenetic groups among uropathogenic *Escherichia coli* including O25-ST131

in Mexico City. *The Journal of Infection in Developing Countries*, *5*(12), 840–849.

Mora, A., Herrera, A., Mamani, R., López, C., Alonso, M., Blanco, J., … Blanco, J. (2010). Recent emergence of clonal group O25b:K1:H4-B2-ST131 *ibeA* strains among *Escherichia coli* poultry isolates, including CTX-M-9-producing strains, and comparison with clinical human isolates. *Applied and Environmental Microbiology*, *76*, 6991–7.

Moreno, E., Andreu, A., Pérez, T., Sabaté, M., Johnson, J. R., & Prats, G. (2006). Relationship between *Escherichia coli* strains causing urinary tract infection in women and the dominant faecal flora of the same hosts. *Epidemiology and Infection*, *134*(5), 1015–23.

Moriel, D. G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., … Serino, L. (2010). Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9072–7.

Murray, P. (1999). *Manual of clinical microbiology*. Washington D.C.: ASM Press.

Naas, T., Mikami, Y., Imai, T., Poirel, L., & Nordmann, P. (2001). Characterization of In53, a class 1 plasmid- and composite transposon-located integron of *Escherichia coli* which carries an unusual array of gene cassettes. *Journal of Bacteriology*, *183*(1), 235–49.

Nataro, J. P., & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, *11*(1), 142–201.

Nicolas-Chanoine, M. H., Robert, J., Vigan, M., Laouénan, C., Brisse, S., Mentré, F., & Jarlier, V. (2013). Different factors associated with CTX-M-producing ST131 and non-ST131 *Escherichia coli* clinical isolates. *PLoS ONE*, *8*, 5–10.

Niu, C., Yu, D., Wang, Y., Ren, H., Jin, Y., Zhou, W., … Liang, L. (2013). Common and pathogen-specific virulence factors are different in function and structure. *Landes Bioscience*, *4*, 473–482.

Niumsup, P. R., Tansawai, U., Boonkerd, N., Polwichai, P., & Dejsirilert, S. (2008). Dissemination of extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* and *Escherichia coli* in Thai hospitals. *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy*, *14*(6), 404–8.

Norton, J. P., & Mulvey, M. A. (2012). Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *Plos Pathogen*, *8*(10).

Novais, A., Canto´ n, R., Valverde, A., Machado, E., Carattoli, A., Baquero, F., & Coque, T. (2006). Dissemination and persistence of *bla*CTX-M-9 are linked to class 1 integrons containing CR1 associated with defective transposon derivatives from Tn402 located in early antibiotic resistance plasmids of IncHI2, IncP1-a, and IncFI Groups. *Antimicrobial Agents and Chemotherapy*, *50*, 2741–2750.

Oliver, A., Coque, T. M., Alonso, D., Valverde, A., Baquero, F., & Cantón, R. (2005). CTX-M-10 linked to a phage-related element is widely disseminated among Enterobacteriaceae in a Spanish hospital. *Antimicrobial Agents and Chemotherapy*, *49*, 1567–1571.

Oliver, A., Weigel, L. M., Rasheed, J. K., McGowan Jr, J. E., Raney, P., & Tenover, F. C. (2002). Mechanisms of decreased susceptibility to cefpodoxime in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *46*(12), 3829–36.

Orskov, F., & Orskov, I. (1984). Serotyping of *Escherichia coli*. *Methods in Molecular Biology*, *Volume 14*, 43–112.

Ou, H. Y., He, X., Shao, Y., Tai, C., Rajakumar, K., & Deng, Z. (2009). DndDB: A database focused on phosphorothioation of the DNA backbone. *PLoS ONE*, *4*.

Ou, H.-Y., Chen, L.-L., Lonnen, J., Chaudhuri, R. R., Thani, A. Bin, Smith, R., … Rajakumar, K. (2006). A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Research*, *34*(1), e3.

Pagani, L., Dell'Amico, E., Migliavacca, R., D'Andrea, M. M., Giacobone, E., Amicosante, G., … Rossolini, G. M. (2003). Multiple CTX-M-type extended-spectrum beta-lactamases in nosocomial isolates of Enterobacteriaceae from a hospital in northern Italy. *Journal of Clinical Microbiology*, *41*(9), 4264–9.

Pagani, L., Mantengoli, E., Migliavacca, R., Nucleo, E., Pollini, S., Spalla, M., … Rossolini, G. M. (2004). Multifocal detection of multidrug-resistant *Pseudomonas aeruginosa* producing the PER-1 extended-spectrum β-lactamase in northern Italy. *Journal of Clinical Microbiology*, *42*, 2523–2529.

Pai, H., Kang, C.-I., Byeon, J.-H., Lee, K.-D., Park, W. B., Kim, H.-B., … Choe, K.-W. (2004). Epidemiology and clinical features of bloodstream infections caused by AmpC-type-beta-lactamase-producing *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, *48*(10), 3720–8.

Pallecchi, L., Malossi, M., Mantella, A., Gotuzzo, E., Trigoso, C., Bartoloni, A., … Rossolini, G. M. (2004). Detection of CTX-M-type β-lactamase genes in fecal *Escherichia coli* isolates from healthy children in Bolivia and Peru. *Antimicrobial Agents and Chemotherapy*, *48*, 4556–4561.

Panina, E. M., Mironov, A. A., & Gelfand, M. S. (2001). Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Research*, *29*, 5195–5206.

Park, S., You, X., & Imlay, J. A. (2005). Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx- mutants of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 9317–9322.

Paterson, D. L., & Bonomo, R. A. (2005). Extended-spectrum β-lactamases : a clinical update. *Clinical Microbiology Reviews*, *18*(4), 657–686.

Paterson, D. L., Hujer, K. M., Hujer, A. M., Yeiser, B., Bonomo, M. D., Rice, L. B., & Bonomo, R. A. (2003). Extended-Spectrum β-Lactamases in *Klebsiella pneumoniae* Bloodstream Isolates from Seven Countries: Dominance and Widespread Prevalence of SHV- and CTX-M-Type β-Lactamases. *Antimicrobial Agents and Chemotherapy*, *47*(11), 3554–3560.

Peigne, C., Bidet, P., Mahjoub-Messai, F., Plainvert, C., Barbe, V., Médigue, C., … Bonacorsi, S. (2009). The plasmid of *Escherichia coli* strain S88 (O45:K1:H7) that causes neonatal meningitis is closely related to avian pathogenic *E. coli* plasmids and is associated with high-level bacteremia in a neonatal rat meningitis model. *Infection and Immunity*, *77*(6), 2272–84.

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

Perez-casal, J. F., & Crosa, J. H. (1984). Aerobactin iron uptake sequences in plasmid ColV-K30 are flanked by inverted IS*J*-like elements and replication regions. *Journal of Bacteriology*, *160*(1), 256–265.

Pfeifer, Y., Cullik, A., & Witte, W. (2010). Resistance to cephalosporins and carbapenems in Gram-negative bacterial pathogens. *International Journal of Medical Microbiology*, *300*(6), 371–379.

Philippon, A., Arlet, G., & Jacoby, G. A. (2002). Plasmid-determined AmpC-type beta-lactamases. *Antimicrobial Agents and Chemotherapy*, *46*(1), 1–11.

Picard, B., Garcia, J. S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., … Denamur, E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection? *Infection and Immunity*, *67*, 546–553.

Pitout, J. D. (2009). Recent changes in the epidemiology and management of extended-spectrum β-lactamase-producing Enterobacteriaceae. *F1000 Medicine Reports*, *1*, 16–19.

Pitout, J. D. D. (2012). Extraintestinal pathogenic *Escherichia coli*: A combination of virulence with antibiotic resistance. *Frontiers in Microbiology*, *3*(9).

Pitout, J. D. D., Church, D. L., Gregson, D. B., Chow, B. L., McCracken, M., Mulvey, M. R., & Laupland, K. B. (2007). Molecular epidemiology of CTX-M-producing *Escherichia coli* in the Calgary Health Region: emergence of CTX-M-15-producing isolates. *Antimicrobial Agents and Chemotherapy*, *51*(4), 1281–6.

Pitout, J. D. D., & Laupland, K. B. (2008). Extended-spectrum β-lactamase-producing Enterobacteriaceae : an emerging public-health concern. *Lancet*, *8*(March), 159–166.

Platell, J. L., Cobbold, R. N., Johnson, J. R., Heisig, A., Heisig, P., Clabots, C., … Trott, D. J. (2011). Commonality among fluoroquinolone-resistant sequence type ST131 extraintestinal *Escherichia coli* isolates from humans and companion animals in

Australia. *Antimicrobial Agents and Chemotherapy*, *55*(8), 3782–7.

Poirel, L., Bonnin, R. a, & Nordmann, P. (2012). Genetic support and diversity of acquired extended-spectrum β-lactamases in Gram-negative rods. *Infection, Genetics and Evolution :*, *12*(5), 883–93.

Poirel, L., Decousser, J., & Nordmann, P. (2003). Insertion Sequence IS*Ecp1B* Is Involved in Expression and Mobilization of a *bl*a CTX-M β-Lactamase Gene. *Antimicrobial Agents and Chemotherapy*, *47*(9), 2938–2945.

Poirel, L., Decousser, J., & Nordmann, P. (2005). IS*Ecp1B* -mediated transposition of *bla* CTX-M in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *49*(1), 1–4.

Poirel, L., Gniadkowski, M., & Nordmann, P. (2002). Biochemical analysis of the ceftazidime-hydrolysing extended-spectrum β-lactamase CTX-M-15 and of its structurally related β-lactamase CTX-M-3. *Journal of Antimicrobial Chemotherapy*, *50*, 1031–1034.

Poirel, L., Lebessi, E., Castro, M., Fèvre, C., Foustoukou, M., & Nordmann, P. (2004). Nosocomial outbreak of extended-spectrum beta-lactamase SHV-5-producing isolates of *Pseudomonas aeruginosa* in Athens, Greece. *Antimicrobial Agents and Chemotherapy*, *48*, 2277–2279.

Poirel, L., Naas, T., Guibert, M., Chaibi, E. B., Labia, R., & Nordmann, P. (1999). Molecular and biochemical characterization of VEB-1, a novel class A extended-spectrum beta-lactamase encoded by an *Escherichia coli* integron gene. *Antimicrobial Agents and Chemotherapy*, *43*(3), 573–81.

Poirel, L., Naas, T., & Nordmann, P. (2008). Genetic support of extended-spectrum β-lactamases. *Clinical Microbiology and Infection*, *14*, 75–81.

Pourcel, C., & Drevet, C. (2013). Occurrence, Diversity of CRISPR-Cas systems and genotyping implications. In R. Barrangou & J. van der Oost (Eds.), *CRISPR-Cas Systems* (pp. 33–60). Berlin, Heidelberg: Springer Berlin Heidelberg.

Prasadarao, N. V, Blom, A. M., Villoutreix, B. O., & Linsangan, L. C. (2002). A novel interaction of outer membrane protein A with C4b binding protein mediates serum resistance of *Escherichia coli* K1. *Journal of Immunology (Baltimore, Md. : 1950)*, *169*, 6352–6360.

Price, L. B., Johnson, J. R., Aziz, M., Clabots, C., Johnston, B., Tchesnokova, V., … Sokurenko, E. V. (2013). The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *mBio*, *4*(6), 1–10.

Pupo, G., Karaolis, D., Lan, R., & Reeves, P. (1997). Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infection and Immunity*, *65*, 2685–92.

R Development Core Team. (2013). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing Vienna Austria*. Retrieved from http://www.r-project.org/

Ranjbar, R., Karami, A., Farshad, S., Giammanco, G. M., & Mammina, C. (2014). Typing methods used in the molecular epidemiology of microbial pathogens : a how-to guide. *New Microbiologica*, *37*, 1–15.

Ray, T., Mills, A., & Dyson, P. (1995). Tris-dependent oxidative DNA strand scission during electrophoresis. *Electrophoresis*, *16*, 888–894.

Riley, L. W. (2014). Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clinical Microbiology and Infection*, *20*(5), 380–90.

Roesch, P. L. (2003). Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract. *Mol. Microbiol.*, *49*, 55–67.

Römling, U., & Tümmler, B. (2000). Achieving 100% typeability of *Pseudomonas aeruginosa* by pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, *38*(1), 464–5.

Rossolini, G. M., Andrea, M. M. D., & Mugnaioli, C. (2008). The spread of CTX-M-type extended-spectrum b -lactamases. *Clinical Microbiology and Infection*, *14*(1), 33–41.

Sabri, M., Léveillé, S., & Dozois, C. M. (2006). A SitABCD homologue from an avian pathogenic *Escherichia coli* strain mediates transport of iron and manganese and resistance to hydrogen peroxide. *Microbiology*, *152*, 745–58.

Saladin, M., Cao, V. T. B., Lambert, T., Donay, J.-L., Herrmann, J.-L., Ould-Hocine, Z., … Arlet, G. (2002). Diversity of CTX-M beta-lactamases and their promoter regions from Enterobacteriaceae isolated in three Parisian hospitals. *FEMS Microbiology Letters*, *209*(2), 161–8.

Salipante, S. J., Roach, D. J., Kitzman, J. O., Snyder, M. W., Stackhouse, B., Butler-wu, S. M., … Shendure, J. (2015). Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Research*, 1–10.

Samudrala, R., Heffron, F., & McDermott, J. E. (2009). Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type iii secretion systems. *PLoS Pathogens*, *5*.

Sandy, M., & Butler, A. (2010). Microbial Iron Acquisition: Marine and Terrestrial Siderophores. *Chemical Reviews*, *109*(10), 4580–4595.

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., … Griffin, P. M. (2011). Foodborne illness acquired in the United States-Major pathogens. *Emerging Infectious Diseases*, *17*(1), 7–15.

Schembri, M. A., & Klemm, P. (2001). Biofilm formation in a hydrodynamic environment by novel FimH variants and ramifications for virulence. *Infection and Immunity*,

69, 1322–1328.

Sekawi, Z., Yusof, R., Shamsudin, & Nor, M. (2008). Extended-spectrum β-lactamases-producing *Escherichia coli* from a tertiary hospital in Malaysia: Emergence of CTX-M-type β-Lactamases variation. *Research Journal of Microbiology*, *3*(6), 489.

Sengupta, M., & Austin, S. (2011). Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. *Infection and Immunity*, *79*(7), 2502–9.

Severin, J. A, Mertaniasih, N. M., Kuntaman, K., Lestari, E. S., Purwanta, M., Lemmens-Den Toom, N., … Goessens, W. H. (2010). Molecular characterization of extended-spectrum beta-lactamases in clinical *Escherichia coli* and *Klebsiella pneumoniae* isolates from Surabaya, Indonesia. *The Journal of Antimicrobial Chemotherapy*, *65*, 465–469.

Severin, J. A., Lestari, E. S., Kloezen, W., Lemmens-den Toom, N., Mertaniasih, N. M., Kuntaman, K., … Keuter, M. (2012). Faecal carriage of extended-spectrum β-lactamase-producing Enterobacteriaceae among humans in Java, Indonesia, in 2001-2002. *Tropical Medicine and International Health*, *17*, 455–461.

Shakil, S., Akram, M., Ali, S. M., & Khan, A. U. (2010). Acquisition of extended-spectrum beta-lactamase producing *Escherichia coli* strains in male and female infants admitted to a neonatal intensive care unit: molecular epidemiology and analysis of risk factors. *Journal of Medical Microbiology*, *59*(Pt 8), 948–54.

Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H.-Y., … Deng, Z. (2011). TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Research*, *39*(Database issue), D606–11.

Shao, Y., He, X., Harrison, E. M., Tai, C., Ou, H. Y., Rajakumar, K., & Deng, Z. (2010). mGenomeSubtractor: A web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. *Nucleic Acids Research*, *38*, W194–200.

Shea, A., Wolcott, M., Daefler, S., & Rozak, D. A. (2012). *Microbial Systems Biology*. (A. Navid, Ed.) (Vol. 881). Totowa, NJ: Humana Press.

Sheng, W.-H., Badal, R. E., & Hsueh, P.-R. (2013). Distribution of extended-spectrum β-lactamases, AmpC β-lactamases, and carbapenemases among Enterobacteriaceae isolates causing intra-abdominal infections in the Asia-Pacific region: results of the study for Monitoring Antimicrobial Resistance Trends (SMA. *Antimicrobial Agents and Chemotherapy*, *57*(7), 2981–8.

Sidjabat, H. E., & Paterson, D. L. (2015). Multidrug-resistant *Escherichia coli* in Asia: epidemiology and management. *Expert Review of Anti-Infective Therapy*, *13*(5), 575–91.

Skyberg, J. a, Johnson, T. J., Johnson, J. R., Clabots, C., Logue, C. M., & Nolan, L. K. (2006). Acquisition of avian pathogenic *Escherichia coli* plasmids by a commensal

E. coli isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infection and Immunity*, *74*(11), 6287–92.

Smet, A., Van Nieuwerburgh, F., Vandekerckhove, T. T. M., Martel, A., Deforce, D., Butaye, P., & Haesebrouck, F. (2010). Complete nucleotide sequence of CTX-M-15-plasmids from clinical *Escherichia coli* isolates: insertional events of transposons and insertion sequences. *PloS One*, *5*(6), e11202.

Soares, S. C., Abreu, V. A. C., Ramos, R. T. J., Cerdeira, L., Silva, A., Baumbach, J., … Azevedo, V. (2012). PIPS: pathogenicity island prediction software. *PloS One*, *7*(2), e30848.

Song, W., Lee, H., Lee, K., Jeong, S. H., Bae, I. K., Kim, J.-S., & Kwak, H.-S. (2009). CTX-M-14 and CTX-M-15 enzymes are the dominant type of extended-spectrum beta-lactamase in clinical isolates of *Escherichia coli* from Korea. *Journal of Medical Microbiology*, *58*(Pt 2), 261–6.

Stürenburg, E., Kühn, A., Mack, D., & Laufs, R. (2004). A novel extended-spectrum beta-lactamase CTX-M-23 with a P167T substitution in the active-site omega loop associated with ceftazidime resistance. *The Journal of Antimicrobial Chemotherapy*, *54*(2), 406–9.

Stürenbug, E., & Mack, D. (2003). Extended-spectrum β-lactamases: implications for the clinical microbiology laboratory, therapy, and infection control. *Journal of Infection*, *47*, 273–295.

Subashchandrabose, S., Hazen, T. H., Rasko, D. a, & Mobley, H. L. T. (2013). Draft genome sequences of five recent human uropathogenic *Escherichia coli* isolates. *Pathogens and Disease*, *69*, 66–70.

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig : a genome comparison visualiser . *Bioinformatics (Oxford, England)*.

Takayanagi, M., Nagasawa, Z., Kusaba, K., & Fukutomi, Y. (2005). The addition of thiourea is very useful for DNA restriction fragments to be seen clearly in PFGE analysis. *Japanese Journal of Medical Technology*, *54*, 1207–1211.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739.

Tan, T. Y., Ng, L. S. Y., He, J., & Hsu, L. Y. (2010). CTX-M and *ampC* beta-lactamases contributing to increased prevalence of ceftriaxone-resistant *Escherichia coli* in Changi General Hospital, Singapore. *Diagnostic Microbiology and Infectious Disease*, *66*, 210–3.

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. *Journal of Clinical Microbiology*, *33*(9), 2233–9.

Thong, K. L., Lai, K. S., Puthucheary, S. D., Koh, Y. T., Ahmad, N., & Yassin, R. M. (2007). Subtyping of *Salmonella enterica* serovar Muenchen by pulsed field gel electrophoresis. Plasmid profiling and antimicrobial susceptibility testing. *Malaysian Journal Science*, *26*, 1–13.

Toleman, M. A., Bennett, P. M., & Walsh, T. R. (2006). ISCR Elements: Novel Gene-Capturing Systems of the 21st Century? *Microbiology and Molecular Biology Reviews*, *70*(2), 296–316.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., … Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, *5*(1), e1000344.

Trang, N. H. T., Nga, T. V. T., Campbell, J. I., Hiep, N. T., Farrar, J., Baker, S., & Duy, P. T. (2013). Original Article The characterization of ESBL genes in *Escherichia coli* and Klebsiella pneumoniae causing nosocomial infections in Vietnam. *The Journal of Infection in Developing Countries*, *7*(12), 922–928.

Tzouvelekis, L. S., Tzelepi, E., Tassios, P. T., & Legakis, N. J. (2000). CTX-M-type β-lactamases: An emerging group of extended-spectrum enzymes. *International Journal of Antimicrobial Agents*, *14*, 137–142.

Victor de Lorenzo, Herrero, M., & Neilands, J. B. (1988). IS1-mediated mobility of the aerobactin system of pColV-K30 in *Escherichia coli*. *Molecular and General Genetics*, *213*(2-3), 487–490.

Villa, L., García-Fernández, A., Fortini, D., & Carattoli, A. (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *The Journal of Antimicrobial Chemotherapy*, *65*(12), 2518–29.

Vinué, L., Sáenz, Y., Martínez, S., Somalo, S., Moreno, M. A., Torres, C., & Zarazaga, M. (2009). Prevalence and diversity of extended-spectrum beta-lactamases in faecal *Escherichia coli* isolates from healthy humans in Spain. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *15*(10), 954–7.

Wang, L., Chen, S., Vergin, K. L., Giovannoni, S. J., Chan, S. W., DeMott, M. S., … Dedon, P. C. (2011). DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 2963–2968.

Waters, V. L., & Crosa, J. H. (1991). Colicin V virulence plasmids. *Microbiological Reviews*, *55*(3), 437–50.

Weldhagen, G. F., Poirel, L., & Nordmann, P. (2003). Ambler Class A Extended-Spectrum -Lactamases in *Pseudomonas aeruginosa*: Novel Developments and Clinical Impact. *Antimicrobial Agents and Chemotherapy*, *47*(8), 2385–2392.

White, P. A., McIver, C. J., & Rawlinson, W. D. (2001). Integrons and gene cassettes in the Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy*, *45*(9), 2658–61.

World Health Organization (WHO). (2014). *Antimicrobial Resistance: Global Report on Surveillance*. Available at http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf?ua=1.

Winokur, P. L., Vonstein, D. L., Hoffman, L. J., & Uhlenhopp, E. K. (2001). Evidence for Transfer of CMY-2 AmpC β-Lactamase Plasmids between *Escherichia coli* and Salmonella Isolates from Food Animals and Humans. *Antimicrobial Agents and Chemotherapy*, *45*(10), 2716–2722.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., … Achtman, M. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular Microbiology*, *60*(5), 1136–51.

Woodford, N., Carattoli, A., Karisik, E., Underwood, A., Ellington, M. J., & Livermore, D. M. (2009). Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. *Antimicrobial Agents and Chemotherapy*, *53*(10), 4472–82.

Woodford, N., Ward, M. E., Kaufmann, M. E., Turton, J., Fagan, E. J., James, D., … Livemore, D. M. (2004). Community and hospital spread of *Escherichia coli* producing CTX-M extended-spectrum β-lactamases in the UK. *Journal of Antimicrobial Chemotherapy*, *54*, 735–743.

Xie, X., Liang, J., Pu, T., Xu, F., Yao, F., Yang, Y., … Wang, Z. (2012). Phosphorothioate DNA as an antioxidant in bacteria. *Nucleic Acids Research*, *40*, 9115–9124.

Xu, T., Yao, F., Zhou, X., Deng, Z., & You, D. (2010). A novel host-specific restriction system associated with DNA backbone S-modification in *Salmonella*. *Nucleic Acids Research*, *38*, 7133–7141.

Yan, J.-J., Hsueh, P.-R., Lu, J.-J., Chang, F.-Y., Shyr, J.-M., Wan, J.-H., … Wu, J.-J. (2006). Extended-spectrum beta-lactamases and plasmid-mediated AmpC enzymes among clinical isolates of *Escherichia coli* and *Klebsiella pneumoniae* from seven medical centers in Taiwan. *Antimicrobial Agents and Chemotherapy*, *50*(5), 1861–4.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134.

Yong, D., Shin, J. H., Kim, S., Lim, Y., Yum, J. H., Lee, K., … Bauernfeind, A. (2003). High prevalence of PER-1 extended-spectrum beta-lactamase-producing *Acinetobacter* spp . in Korea. *Antimicrobial Agents and Chemotherapy*, *47*, 1749–1751.

Yoon, S. H., Han, M.-J., Jeong, H., Lee, C. H., Xia, X.-X., Lee, D.-H., … Kim, J. F. (2012). Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biology*, *13*(5), R37.

Yoon, S. H., Park, Y.-K., & Kim, J. F. (2015). PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. *Nucleic Acids Research*, *43*(Database issue), D624–30.

Yoshioka, Y., Ohtsubo, H., & Ohtsubo, E. (1987). Repressor gene *finO* in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS3 into F *finO*. *Journal of Bacteriology*, *169*(2), 619–23.

Yu, K. X., & Thong, K. L. (2009). Multiplex PCR for simultaneous detection of virulence genes in *Escherichia coli*. *Malaysian Journal of Science*, *28*(1), 1–14.

Yu, Y., Ji, S., Chen, Y., Zhou, W., Wei, Z., Li, L., & Ma, Y. (2007). Resistance of strains producing extended-spectrum beta-lactamases and genotype distribution in China. *The Journal of Infection*, *54*(1), 53–7.

Yuan, M., Aucken, H., Hall, L. M. C., Pitt, T. L., & Livermore, D. M. (1998). Epidemiological typing of *Klebsiellae* with extended-spectrum β-lactamases from European intensive care units. *Journal of Antimicrobial Chemotherapy*, *41*(5), 527–539.

Yusof, R. (2011, January 1). *Molecular Characterisation of Extended Spectrum Beta-Lactamase-Producing Escherichia Coli in Selayang Hospital, Malaysia*. Universiti Putra Malaysia.

Zahrl, D., Wagner, M., Bischof, K., & Koraimann, G. (2006). Expression and assembly of a functional type IV secretion system elicit extracytoplasmic and cytoplasmic stress responses in *Escherichia coli*. *Journal of Bacteriology*, *188*(18), 6611–21.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., … Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *The Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–4.

Zhao, W.-H., & Hu, Z.-Q. (2013). Epidemiology and genetics of CTX-M extended-spectrum β-lactamases in Gram-negative bacteria. *Critical Reviews in Microbiology*, *39*(1), 79–101.

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., & Yu, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics (Oxford, England)*, *28*(3), 416–8.

Zhou, L., Lei, X.-H., Bochner, B. R., & Wanner, B. L. (2003). Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *Journal of Bacteriology*, *185*(16), 4956–4972.

Zhou, M., Guo, Z., Duan, Q., Hardwidge, P. R., & Zhu, G. (2014). *Escherichia coli* type III secretion system 2: a new kind of T3SS? *Veterinary Research*, *45*, 32.

Zhou, X., Deng, Z., Firmin, J. L., Hopwood, D. A., & Kieser, T. (1988). Site-specific degradation of *Streptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. *Nucleic Acids Research*, *16*, 4341–4352.

Zhou, X., He, X., Liang, J., Li, A., Xu, T., Kieser, T., … Deng, Z. (2005). A novel DNA

modification by sulphur. *Molecular Microbiology*, *57*, 1428–1438.

Zhou, Z., Ogasawara, J., Nishikawa, Y., Seto, Y., Helander, A., Hase, A., … Haruki, K. (2002). An outbreak of gastroenteritis in Osaka , Japan due to *Escherichia coli* serogroup O166 : H15 that had a coding gene for enteroaggregative *E . coli* heat-stable enterotoxin 1. *Epidemiology and Infection*, *128*, 363–371.

List of publications and papers presented

List of publications

- Ho, W. S., Balan, G., Puthucheary, S., Kong, B. H., Lim, K. T., Tan, L. K., Koh, X. P., Yeo, C.C., and Kwai Lin Thong Thong, K. L. (2012). Prevalence and Characterization of Multidrug-Resistant and Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* from Pediatric Wards of a Malaysian Hospital. *Microbial Drug Resistance*, 18(4), 408–416. http://doi.org/10.1089/mdr.2011.0222.

- Ho, W.S., Thong, K.L., Yeo, C.C. (2012). Multilocus sequence typing of clinical ESBL-producing *E. coli* strains. *International Journal of Infectious Disease*s 16, e416-e417.

- Ho, W. S., Gan, H. M., Yap, K. P., Balan, G., Yeo, C. C., & Thong, K.L. (2012). Genome sequence of multidrug-resistant *Escherichia coli* EC302/04, isolated from a human tracheal aspirate. *Journal of Bacteriology*, 194(23), 6691–2. http://doi.org/10.1128/JB.01804-12.

- Ho, W. S., Ou, H.Y., Yeo, C. C., & Thong, K. L. (2015). The *dnd* operon for DNA phosphorothioation modification system in *Escherichia coli* is located in diverse genomic islands. *BMC Genomics*, 16, 199. http://doi.org/10.1186/s12864-015-1421-8.

- Ho, W.S., Yap, K.P., Yeo, C.C., Ganeswrie, R., and Thong, K.L. Complete sequence and comparative analysis of a multidrug-resistance and virulence multireplicon IncFIIA plasmid from *Escherichia coli* EC302/04 isolated from a patient with extraintestinal infection. *Frontiers in Microbiology*, 6, 1547. http://dx.doi.org/10.3389/fmicb.2015.01547

- Ho, W.S., Yap, K.P., Yeo, C.C., Ganeswrie, R., and Thong, K.L. High resolution phylogenomic analysis coupled with extensive virulence profiling: a useful approach for identifying *Escherichia coli* with underlying threat to cause extraintestinal infections. (In preparation, to be submitted to Clinical Infectious Diseases)

- Ho, W.S., Yeo, C.C., Ganeswrie, R., Yusof, M.Y.M, and Thong, K.L. Emergence and characterization of *Escherichia coli* clone O25-ST131 in Malaysia. (In preparation, to be submitted to Antimicrobial Agents and Chemotherapy)

List of publications (not included in PhD work)

- Ho, W. S., Tan, L. K., Ooi, P. T., Yeo, C. C., & Thong, K. L. (2013). Prevalence and characterization of verotoxigenic-*Escherichia coli* isolates from pigs in Malaysia. *BMC Veterinary Research*, 9, 109. http://doi.org/10.1186/1746-6148-9-109.

- Ghaderpour, A., Ho, W. S., Chew, L.L., Bong, C. W., Chong, V. C., Thong, K.L., & Chai, L. C.(2015). Diverse and abundant multi-drug resistant *E. coli* in Matang mangrove estuaries, Malaysia. *Frontiers in Microbiology*, 6, 977. http://doi.org/10.3389/fmicb.2015.00977.

- Yap, K.P., Ho, W.S., Gan, H.M., Chai L.C., and Thong, K.L. Global MLST of *Salmonella* Typhi Revisited in Post-Genomic Era: Genetic conservation, Population Structure and Concordance with Whole Genome Phylogeny. *Frontiers in Microbiology*, 7, 270. http://dx.doi.org/10.3389/fmicb.2016.00270.

List of presentations

- Ho, W.S., Thong, K.L. (2010, October). Prevalence of extended-spectrum beta-lactamases, class 1 and 2 integrons in *Escherichia coli* isolated from pediatrics' stool samples in Malaysia. Paper presented at My1Bio Conference 2010, Kuala Lumpur, Malaysia.

- Ho, W.S., Yeo, C.C. and Thong, K.L. (2011, December). Genetic characterization of integrons in Malaysian *E. coli* strains. Paper presented at International Congress of Malaysian Society for Microbiology (ICMSM) 2011, Penang, Malaysia.

- Ho, W.S., Yeo, C.C. and Thong, K.L. (2012, July). Characterization of Malaysian extended-spectrum beta-lactamase-producing *Escherichia coli.* Paper presented at National POstgraduate Seminar (NPS) 2012, Kuala Lumpur, Malaysia.

- Ho, W.S., Thong, K.L., Yeo, C.C. (2012, June). Multilocus sequence typing of clinical ESBL-producing *E. coli* strains. Paper presented at 15th International Congress of Infectious Diseases (ICID), Bangkok, Thailand.

- Ho, W.S., Yeo, C.C. Yusof, M.Y.M., and Thong, K.L. (2013, March). Association between plasmid replicon types with cephalosporin resistance traits in *Escherichia coli*. Paper presented at 9th International Symphosium on Antimicrobial Agents and Resistance (ISAAR) 2013, Kuala Lumpur, Malaysia.

- Ho, W.S., Tan, L.K., Ooi, P.T., Yeo, C.C. and Thong, K.L. (2013, June). Pathotyping of swine *Escherichia coli* strains. Paper presented at 20th Malaysian Society for Molecular Biology and Biotechnology (MSMBB) Scientific Meeting, Kuala Lumpur, Malaysia.

- Ho, W.S., Yeo, C.C. Yap, K.P., Thong, K.L. (2015, September). Clonally related *Escherichia coli* strains from different isolation sites are differentiated by the presence of plasmid pEC302/04. Paper presented at Malaysian Society for Molecular Biology and Biotechnology (MSMBB) 22nd Annual Scientific Meeting 2015, Kuala Lumpur, Malaysia.