COMPARATIVE GENOMICS ANALYSIS OF *Escherichia coli*: INSIGHT INTO ITS PATHOGENICITY

AHMED FARAH MOHAMED IBRAHIM

UNIVERSITI SAINS MALAYSIA

2022

COMPARATIVE GENOMICS ANALYSIS OF *Escherichia coli*: INSIGHT INTO ITS PATHOGENICITY

by

AHMED FARAH MOHAMED IBRAHIM

Dissertation submitted in partial fulfilment of
the requirements of the degree of
Master of Science (Biomedicine) Mixed Mode

SEPTEMBER 2022

ACKNOWLEDGEMENT

All praise is due to almighty Allah, the creator, who makes it possible to complete my thesis easily and successfully. Many thanks due to Allah for all things, specifically for giving me the strength, health, and intellectual capability to accomplish this task on time. I would like to acknowledge with deep reverence, sincerity and feel utmost pleasure in expressing my heartiest gratitude to the many people who helped me to make this thesis possible.

First, of all, I am greatly thankful to my supervisor **Dr. Shuhaila binti Mat Sharani** for her prodigious support, encouragement, excellent advice, valuable comments, and professional guidance till the completion of this dissertation. I would also like to express my deepest and heartfelt gratitude to my co-supervisor, **Dr. Nik Yusnoraini Yusof**, who has provided me with excellent advice, timely help, and support during my entire thesis.

Many thanks go to the course coordinator (GTB540) Dr Wong Weng Kin for providing with me timely help and moral support. I would like to thank all my lecturers at Universiti Sains Malaysia (USM) for their guidance, and advice to the dissertation, not forgetting my fellow friends especially Mr. Dahir Ali, Dr Mohamed Ali Mohamud and Abdulkadir Elmi Barre who dedicated their time to equip me with knowledge and skills to provide me excellent support and encouragement during my thesis.

Lastly, I wish to express my deep and sincere thanks to my family especially my beloved mother for her love, affection, and constant prayers. Big thanks, in fact, go to my entire family who always provides inspiration and support to get success in my education career. Words are lacking behind to say thanks to my beloved late father Farah Mohamed who could not see me getting success until this stage. Finally, big thanks goes to my

loving and caring wife Hiba Farah Ali and my two beloved sons who always supported me mentally and physically during my education career I always feel proud to have such a nice, caring and loving wife.

TABLE OF CONTENTS

CERTIFICATE **ii**

DECLARATION **iii**

ACKNOWLEDGEMENT **iv**

TABLE OF CONTENTS **vi**

LIST OF TABLES **viii**

LIST OF FIGURES **ix**

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMNS **x**

ABSTRAK **xv**

ABSTRACT **xvii**

CHAPTER ONE INTRODUCTION ***1***

1.1 Background 1

1.2 Rational of the study 4

1.3 Problem statement 5

1.4 Objective of the study 5

1.5 Overview of the Study 6

CHAPTER TWO LITERATURE REVIEW ***7***

2.1 *Escherichia coli* 7

2.2 Genome of *E. coli* strains 10

2.3 Pathogenicity of *Escherichia coli* 13

2.3.1 Enteropathogenic *Escherichia coli* (EPEC) 14

2.3.2 Enterohemorrhagic *Escherichia coli* (EHEC) 17

2.3.3 Enterotoxigenic *Escherichia coli* (ETEC) 18

2.3.4 Enteroinvasive *Escherichia coli* (EIEC) 20

2.3.5 Enteroaggregative *Escherichia coli* (EAEC) 23

2.3.6 Diffusely Adherent *Escherichia coli* (DAEC) 24

2.3.7 Adherence invasive *Escherichia coli* (AIEC) 26

2.3.8 Uropathogenic *Escherichia coli* (UPEC) 26

2.3.9 Neonatal meningitis *Escherichia coli* (NMEC) 28

2.3.10 Avian pathogenic *Escherichia coli* (APEC) 29

2.4 Mobile gene elements of Escherichia coli (MGE) 30

2.5 Phylogenetic analysis of *E. coli* 41

2.6 Multi-Locus Sequence Typing (MLST) 44

CHAPTER THREE MATERIALS AND METHODS ***46***

3.1 Materials 46

3.2 Methods 46

CHAPTER FOUR RESULTS ***51***

4.1 Collection of 25 *Escherichia coli* Genome Strain 51

4.2 Genomic islands 53

CHAPTER FIVE DISCUSSION ***64***

CHAPTER SIX CONCLUSION ***71***

REFERENCES ***72***

APPENDICES ***88***

APPENDIX 1 HUMAN ETHICAL APPROVAL **88**

APPENDIX 2 MSLT RESULT FOR 25 *E. COLI* STRAINS **89**

APPENDIX 3 THE 16s rRNA **98**

APPENDIX 4 16S RRNA ALIGNMENT **103**

APPENDIX 5 THE GENOMIC ISLAND OUTPUT 1 **109**

APPENDIX 6 THE GENOMIC ISLAND OUTPUT 2 **122**

LIST OF TABLES

Table 2.1 ExPEC and InPEC *E. coli* pathotypes, abbreviation, disease presentations, and associated virulence factors (VFs)…….....15

Table 4.1 25 *E. coli* genome data 52

Table 4.2 Number of GIs in each strain for all 25 strains. 54

Table 4.3 Genomic islands for the most pathogenic and local strains 56

Table 4.4 Sequence types of 25 *E. coli* strains 60

Table 4.5 The alleles and sequence types of the strains 62

LIST OF FIGURES

Figure 1.1 Study flow chart ……6

Figure 2.1 Hub genes have high connectivity in the accessory genome. ……..12

Figure 2.2 Description of four mechanism that EPEC interact with host cell. ……16

Figure 2.3 The pathophysiology of ETEC.....20

Figure 2.4 Interaction of *Shigella*/EIEC during invasion with immune cells and the epithelium.....23

Figure 2.5 Distribution of virulence factors in UPEC pathotypes and their role in the physiopathology of infections.....28

Figure 2.6 The procedures involved in the intracellular or intercellular transfer of antibiotic resistance genes as well as mobile genetic elements (MGE) 31

Figure 2.7 Major features and functions of genomic islands… 32

Figure 2.8 Structure and motility functions of PAIs. 34

Figure 2.9 Maximum-likelihood phylogenetic tree showing genetic relatedness of *E. coli* strain F2_18C. 43

Figure 3.1 Methodology of GIs by using frequency numbers.....47

Figure 3.2 Methodology of the phylogenetic tree.....49

Figure 3.3 The procedure of MLST.....50

Figure 4.1 The number of GIs in each strain.....55

Figure 4.2 Phylogenetic tree of 25 *E. coli* strains using neighbor-joining method with 1000 bootstraps. 59

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMNS

≥	Greater than or equal
A/E	Attaching and effacing
AAF	Aggregative adherence fimbriae
ADP	Adenosine diphosphate
aEPEC	atypical enteropathogenic *E. coli*
aer	Aerobactin
Afa/Dr	afimbrial and fimbrial adhesins
AIDA-I	Adhesin involved in diffuse adherence
AIEC	Adherent-Invasive *E. coli*
AMR	Antimicrobial resistance
APEC	Avian pathogenic *E. coli*
BBB	Blood-brain barrier
bfpA	bundlin pilin protein of bundle-forming pilus
CAECAM	Carcino-embryogenic antigen-related cell adhesion molecule
cAMP	Cyclic adenosine monophosphate
CD	Crohn's disease
CFAs	Colonization factor antigens
CFTR	Cystic fibrosis transmembrane conductance regulator
cGMP	Cyclic guanosine monophosphate
CNF	Cytotoxic necrotizing factor
CNS	Central nervous system
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSV	Comma-separated values

DAEC	Diffusely adherent *E. coli*
DAF	Decay accelerating factor
DC	Dendritic cell
DEC	Diarrheagenic *E. coli* (DEC)
DNA	deoxyribonucleic acid
DRS	Direct repeat sequences
E. coli	*Escherichia coli*
Eae	cell-surface intimin
EAEC	Enteroaggregative *E. coli*
ECP	*E. coli* common pilus
EHEC	Enterohemorrhagic *E. coli*
ehxA	Enterohemolysin
EIEC	Enteroinvasive *E. coli*
EOM	Early-onset meningitis
EPEC	Enteropathogenic *E. coli*
Esp	*E. coli*-secreted protein
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
GIs	Genomic islands
GIT	Gastrointestinal Tract
GTPase	Guanosine triphosphatase
HAP	Hospital-acquired Pneumonia
HC	haemorrhagic colitis
HGT	Horizontal gene transfer
HPI	high-pathogenicity island

HUS	Haemolytic-Uremic Syndrome
ICEs	Integrative and conjugative elements
IECs	Intestinal epithelial specialized cell
IL-8	Interleukin 8
Inc	Incompatible
InPEC	Intestinal pathogenic *E. coli*
iss	Increased serum survival protein
Kbp	Kilobase pairs
LEE	Locus of enterocyte effacement
LeoA	Labile enterotoxin output
LOM	Late-onset meningitis
LPF	Long polar fimbriae
LPS	Lipopolysaccharide
LT	Heat-labile enterotoxins
M cell	Microfold
Mbp	Millions of base pairs
MDR	Multi-drug resistance
MGE	Mobile genetic elements
MPF	Mating pair formation
MUSCLE	Multiple sequences comparison by log expectation
NCBI	National center for biotechnology information
NIS	Non-ribosomal peptide synthetase-independent siderophore
nle	non-LEE
NMEC	Neonatal meningitis *E. coli*
PAG	Pathogenicity association genes

PAIs	Pathogenicity islands
pEAF	adherence factor plasmid
Pic	Protein involved in intestinal colonisation
RG	Resistance genes
rRNA	Ribosomal ribonucleic acid
Sat	Secreted autotransporter toxin
SEPEC	Sepsis-associated *E. coli*
ShET1	Shiga enterotoxin 1
ShET2	Shiga enterotoxin 2
ShMu	Shigella mucinase
SigA	Shigella immunoglobulin A protease
SLT	Shiga-like toxin
SNP	single nucleotide polymorphism
SPATEs	Serine Protease Autotransporters of Enterobacteriaceae
SSI	Surgical Site Infection
ST	Heat-stable enterotoxins
STEC	Shiga toxin-producing *E. coli*
STs	Sequence types
stx	Shiga toxins
T2SS	Type II secretion system
T3SS	type III secretion system
tEPEC	typical enteropathogenic *E. coli*
Tir	Translocated intimin receptor
tmRNA	transfer-messenger ribonucleic acid
TNF	Tumour necrosis factor

tRNA	transfer ribonucleic acid
Tsh	temperature-sensitive hemagglutinin
TTP	Thrombotic thrombocytopenic purpura
UPEC	Uropathogenic *E. coli*
UTI	Urinary Tract Infection
VAGs	Virulence-associated genes (VAGs)
VG	Virulence genes
VTEC	Verotoxin-producing *E. coli*

PERBANDINGAN ANALISI GENOM *Escherichia coli*: PENDEKATAN TERHADAP KEPATOGENANNYA

ABSTRAK

Escherichia coli (*E. coli*) ialah bakteria bacilli gram-negatif daripada keluarga Enterobacteriaceae yang boleh menjangkiti dan mampu hidup sebagai komensal dalam saluran gastrousus bayi (GIT) beberapa jam selepas kelahiran, termasuklah haiwan dan persekitaran. *E. coli* boleh menyebabkan jangkitan pada usus atau organ luar usus bagi dikedua-dua negara berpendapatan tinggi dan rendah termasuklah penyakit invasif yang serius seperti bakteremia dan sepsis. Tambahan pula, *E. coli* ialah mikroorganisma rintangan antibiotik terkemuka seperti *E. coli* jenis jujukan 131 (JJ131), adalah contoh kejayaan pelbagai rintangan ubat bagi klon pandemik. Matlamat kajian ini adalah untuk membandingkan jujukan genom 22 strain *E. coli* genom lengkap dengan 3 strain draf tempatan dan mendapatkan gambaran tentang kepatogenannya. Perbandingan kepatogenan antara strain tempatan dan strain awam telah disiasat berdasarkan analisis jujukan genom kesemua 25 strain yang diperoleh daripada pangkalan data NCBI. Islandviewer 4 digunakan untuk meramalkan kepulauan genom. OmicsBox 2.1.14 digunakan sebagai alat mencari jenis jujukan berbilang lokus (JJBL). Hubungan antara patogen dan bukan patogen dianalisis dengan membina pokok filogeni menggunakan kaedah *neighbor joining* daripada perisian MEGA dengan 1000 bootstraps berdasarkan 16S rRNA gen yang mewakili setiap strain. Hasil analisis perbandingan genom mendapati bahawa 11 daripada 25 strain itu adalah patogen dan selebihnya bukan patogen. Tujuh daripada 11 strain patogen ini sangat berbahaya. Kebanyakan strain adalah dari lokasi yang sama dalam analisis filogenetik didapati berkumpul bersama, dan dua strain

tempatan (INF13/18/A dan INF191/17/A) berkongsi asal usul yang sama dengan strain tinggi kepatogenan yang menghasilkan toksin Shiga (Al Ain dan AR -0427), manakala strain tempatan ketiga (INF32/16/A) berkongsi asal usul dengan strain makmal yang bukan patogen iaitu K-12 substr. MG1655. Kebanyakan strain patogen untuk kajian MLST masuk ke dalam kategori ST11, manakala strain bukan patogen masuk ke dalam klasifikasi ST10. Strain tempatan kami dikaitkan dengan pelbagai jenis jujukan (*Sequence type*; ST)) iaitu INF13/18/A dengan ST156, INF32/16/A dengan ST354 dan INF191/17/A dengan ST131, semuanya dicadangkan sebagai patogen, Extraintestinal *E. coli* (ExPEC). Sebagai kesimpulan, kepatogenen dua strain tempatan (INF13/18/A dan INF191/17/A) adalah sangat tinggi, tetapi satu (INF32/16/A) mempunyai kepatogenan rendah. Secara umumnya, kepatogenan kesemua strain tempatan adalah kurang berbanding dengan 22 strain yang lain.

COMPARATIVE GENOMICS ANALYSIS OF *Escherichia coli*: INSIGHT INTO ITS PATHOGENICITY

ABSTRACT

Escherichia coli (*E. coli*) is a gram-negative bacilli bacterium from the Enterobacteriaceae family that may colonise and survive as a commensal in the infant's gastrointestinal tract (GIT) a few hours after birth, as well as in animals and the environment. *E. coli* may cause infections in the intestine or extraintestinal organs in both high and low-income nations including in serious invasive diseases such as bacteraemia and sepsis. Furthermore, *E. coli* is the leading antibiotic resistance organisms such as *E. coli* sequence type 131 (ST131), an example of a successful multidrug-resistant pandemic clone. The aim of this study was to compare the 22 complete genome strains against 3 local draft genome strains of *E. coli* and gain insight into their pathogenicity. Pathogenicity comparison between local strains and public strains was investigated based on analysis of 25 strains genomic sequences that were retrieved from the NCBI database. Islandviewer 4 was used to forecast genomic islands. OmicsBox 2.1.14 was employed as a multi-locus sequence typing tool (MLST) to screen for sequence type. The pathogenicity and non-pathogenicity relationship was analysed by constructing a phylogenetic tree using neighbour joining method from MEGA software with 1000 bootstraps based on representative strains of 16S rRNA genes. Result from genome comparison found that 11 of the 25 stains are pathogenic while others non-pathogenic. Seven out of 11 of these pathogenic strains are extremely high pathogenic. Most strains from the same locations in the phylogenetic analysis are found to be clustered together, and two of the local strains (INF13/18/A and INF191/17/A) share a common ancestor with the highly pathogenic strains that produce Shiga toxin (Al Ain and AR-0427), while

the third local strain (INF32/16/A) shares a common ancestor with the non-pathogenic laboratory strain K-12 substr. MG1655. Most pathogenic strains for the MLST study came into the ST11 category, while non-pathogenic strains went into the ST10 classification. Our local strains seem to be associated with various sequence types; INF13/18/A is associated with ST156, INF32/16/A with ST354, and INF191/17/A with ST131, which all suggested as pathogen, Extraintestinal *E. coli* (ExPEC). As conclusion, two of the local strains (INF13/18/A and INF191/17/A) have been identified to be highly pathogenic while INF32/16/A is less pathogenic. In general, since they have less pathogenicity and virulence genes, pathogenicity of local strains is less compared to other 22 strains.

CHAPTER ONE

INTRODUCTION

1.1 Background

Escherichia coli (E. coli) is a gram-negative bacilli bacterium from the Enterobacteriaceae family. In 1885, Theodor Escherich isolated and identified them for the first time from neonatal faeces. *E. coli* may colonise and survive as a commensal in the infant's gastrointestinal tract (GIT) a few hours after birth (Pakbin, Brück & Rossen, 2021). However, *E. coli* may cause infections in the intestine or extraintestinal organs, including serious invasive diseases such as bacteraemia and sepsis. In addition, *E. coli* is the most common cause of bacteraemia in high-income nations (Bonten *et al.*, 2021). Furthermore, many serious disorders are caused by *E. coli*, including urinary tract infection (UTI), hospital-acquired pneumonia (HAP), sepsis, surgical site infection (SSI), gastrointestinal tract infections, hemolytic-uremic syndrome (HUS), meningitis, and meningeal inflammation (Sarowska *et al.*, 2019).

Pathogenic *E. coli* are called pathovars because they have virulence factors that allowed them to facilitate an organism's pathogenicity. Different *E. coli* strains have different sets of virulence genes that generate pathogenesis in both diarrheagenic and extra-intestinal infections. Typically, pathogenicity islands and plasmids are used to encode them (Oktay Gultekin, Tezcan Ulger & Delialioğlu, 2022). Pathovars are classified into intestinal and extra-intestinal *E. coli* (Abuelhassan *et al.*, 2016). Intestinal *E. coli* are called Diarrheagenic *E. coli* (DEC) and six pathovars of Diarrheagenic *E. coli* (DEC) were identified as described by Li *et al.*, (2018) they are EnteroPathogenic *E. coli* (EPEC), EnteroToxigenic *E. coli* (ETEC), VeroToxin- or

Shigsa Toxin-producing E. *coli* (VTEC or STEC), EnteroInvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), and Diffusely Adherent *E. coli* (DAEC). This classification is based on distinct virulence factors and the clinical symptoms that result from them (Jarocki *et al.*, 2020). While extra-intestinal, affects the outside of the gastrointestinal tract (GIT) and causes serious infections such as meningitis, bacteraemia, skin and soft tissue infection, and urinary tract infection (UTI) (Čurová *et al.*, 2020). Because of the versatility of *E. coli* genomes, certain strains may live as commensals in the gastrointestinal tract, while the acquisition of different combinations of virulence-associated genes (VAGs) has resulted in clades that cause a wide range of intestinal and extraintestinal disorders (Jarocki *et al.*, 2020).

Notably, prior to the discovery of virulence factors in pathogenic strains, *E. coli* was classified primarily based on the serologic detection of O (lipopolysaccharide, LPS), H (flagellar) antigens, K (capsular) with the O157:H7 (Shumi Gebisa, 2019). It was later discovered that the most common contributions to *E. coli* pathogenicity include phage-associated Shiga toxins (stx), which contribute to haemorrhagic colitis and haemolytic uremic syndrome (HUS) associated with enterohemorrhagic *E. coli* (EHEC) infections, enterohemolysin (ehxA), heat-stable enterotoxin (astA) (Jarocki *et al.*, 2020), in addition to the locus of enterocyte effacement (LEE), which can cause attaching and effacing (A/E), various non-LEE (nle)-encoded effector genes, which are frequently carried on prophage elements and contribute to virulence by interfering with host signalling pathways, apoptosis, and phagocytosis and others (Li *et al.*, 2018).

This study shed the light on comparative genomes of *E. coli* among 25 strains from the various origin and sources. The 16S ribosomal RNA (rRNA) gene sequence is used to build for a phylogenetic tree for a variety of reasons, (i) the existence of the

16S rRNA gene in practically all organisms (prokaryotes), (ii) the function of the 16S rRNA gene through time does not alter (evolution), and (iii) the 16S rRNA gene (1500 bp) is large enough for informatics analysis (Abuelhassan *et al.*, 2016). Subsequently, we used 16S rRNA for the structure of the phylogenetic diagram to determine whether local strains closer to the pathogenic or non-pathogenic of *E. coli* strains. Previously, Desvaux *et al.,* (2020) stated phylogenetically *E. coli* have eight separate groups (A, B1, B2, C, D, E, F, and G), nevertheless Clark & Maresso, (2021) indicated nine groups, and classified furtherly into four major (A, B1, B2 and D) and five minors (C, E, F, G, and cryptic clade I). Non-pathogenic commensal *E. coli* strains are mostly found in groups A and B1, while extraintestinal pathogenic *E. coli* strains are found in groups B2 and D (Książczyk *et al*., 2021, Clark & Maresso, 2021). B1 phylogroup strains may also be more common in domesticated animal isolates. Phylogroup E, on the other hand, comprises the majority of known EHEC isolates, including the well-known O157: H7 serotype. Phylogroup F is quite like phylogroup B2, although they lack many of the extraintestinal pathogenic *E. coli* (ExPEC) associated virulence factors, with the exception of a few sequence types (Clark & Maresso, 2021).

Mobile genetic elements (MGE) are kinds of virulence factors that facilitate many functions of the organism's pathogenicity for example antibiotic resistance. As a result of their ability to leap from one cell to another while also allowing for the mobility of deoxyribonucleic acid (DNA) inside the cell, MGEs are essential for gene transfer (Partridge *et al.*, 2018a). The capacity for modification and adaptation in bacteria is provided by bacterial MGEs, which may take the form of insertion sequences, conjugative transposons, or prophages (Durrant *et al.*, 2020). In addition to phages and plasmids, genomic islands (GIs), a huge section of the genome that may

be transferred from one bacterium to another, have been found. These types of GIs carrying virulence-associated genes have been identified in UPEC in the early 1980s and are referred to as pathogenicity islands (PAIs) (Desvaux *et al.*, 2020). Pathogenicity islands, also known as PAIs, are pieces of DNA that have been horizontally transferred and then fused with the other bacterial chromosome. PAIs are always found combined with the 3'-end of transfer ribonucleic acid (tRNA) and transfer-messenger ribonucleic acid (tmRNA) genes, and they can be identified by the amount of G+C that is present in their DNA (Piña-Iturbe *et al.*, 2018).

As conclusion, identification of pathogenicity for each strain would improve the understanding of the mechanism *E. coli* survival from host defence mechanism. As a result, the purpose of this study was to determine if the isolated local strains were more pathogen or non-pathogen by comparing them to previously categorised pathogen or non-pathogen *E. coli* strains from a public database.

1.2 Rational of the study

Escherichia coli is the most common cause of bacterial bloodstream infections (BSIs) and is associated with a considerable morbidity and mortality burden (MacKinnon *et al.*, 2021). According to publications, in-hospital mortality due to *E. coli* BSI ranges from 5% to over 30% (de Lastours *et al.*, 2020). Comparative genome analysis of *E. coli* in sight on its pathogenicity determines the differences of the pathogenicity based on detection of mobile gene elements and phylogenetic analysis. This study was focused on identifying the pathogenicity from 3 local (Malaysia) strains that were isolated from human blood which are not known whether they are true pathogenic or not compared to other 22 existing strains from available database which provide an

insight of pathogenicity in relation to development anti-microbial specifically target for local strains.

1.3 Problem statement

Escherichia coli (*E. coli*) is a gram-negative rod-shaped bacterium that is considered as a worldwide problem (García & Fox, 2021), since some serotypes can cause serious infections that are known to be virulent. Virulent strains of *E. coli* can cause gastroenteritis, urinary tract infections and newborn meningitis. Besides that, virulent strains can also be responsible for haemolytic uremic syndrome, peritonitis, mastitis, septicaemia, and pneumonia in rare cases. In addition to that, virulent strains cause outbreaks in the different countries on the world like atypical enteropathogenic *E. coli* (aEPEC) diarrheal outbreaks in United States, Japan, China, Brazil, and Finland (Jarocki *et al.*, 2020) and Enteroaggregative *E. coli* (EAEC) strains which have been linked to a number of diarrheal illness outbreaks across the world (Allocati *et al.*, 2013). Furthermore, *E. coli* is the leading antibiotic resistance organisms such as *Escherichia coli* sequence type 131 (ST131), an example of a successful multidrug-resistant pandemic clone (Pajand *et al.*, 2021). Therefore, this comparative study contributed to understanding the pathogenicity, virulence factors and mobile genomic elements of the *E. coli* strain.

1.4 Objective of the study

The general objective of the study is to compare the genomic sequence of 25 strains of *E. coli* and insight into its pathogenicity.

The specific objectives include:

1. To identify mobile gene and genomic islands (GIs) related to the pathogenicity of all *E. coli* strains.

2. To study evolutionary relationship of the 25 strains of *E. coli*

1.5 Overview of the Study

The methodology of the study is as shown below:

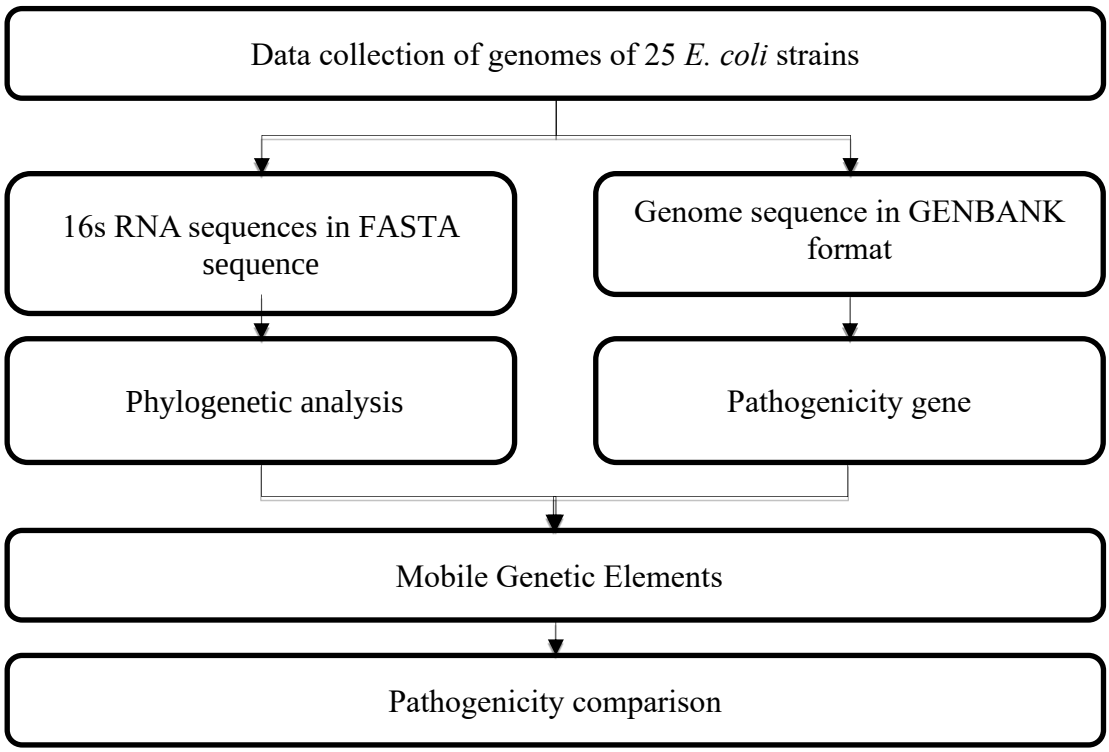

Figure 1.1 Study flowchart

CHAPTER TWO

LITERATURE REVIEW

2.1 *Escherichia coli*

Escherichia coli (E. coli) is an adaptable gram-negative bacillus facultative anaerobic bacterium which belongs to the family of Enterobacteriaceae. It is well known to be a commensal of the human and animal gastrointestinal tract and environments unrelated to the host, such as water, soil, manure, and food (Foster-Nyarko & Pallen, 2022). However, *E. coli* is a major source of human illness and mortality across the world (Clark & Maresso, 2021). According to García & Fox (2021) in Asia and Africa, infectious *E. coli* could probably be linked to mortality in children with moderate-to-severe diarrhoea enterotoxigenic *E. coli* (ETEC) and enteropathogenic *E. coli* (EPEC). In addition to that, *E. coli* is the main cause of bacteraemia in high-income nations, outnumbering other pathogens that cause bacteraemias, such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, and is also a prominent cause of neonatal meningitis (Bonten *et al*., 2021).

Pathotypes or pathovars are referred to the pathogenic *E. coli* which has the potential to cause illness. The pathotypes are arranged according to the various protocols such as the site of infection (e.g. uropathogenic *E. coli* (UPEC), which causes infections in the urinary tract, extraintestinal pathogenic *E. coli* (ExPEC), which causes infections in organs outside the gut, host (e.g. avian pathogenic *E. coli* (APEC), which causes infections in avian species), site, and host (e.g. neonatal meningitis *E. coli* (NMEC), which infects the cerebrospinal fluid in new-borns and pathogenesis (e.g Shiga toxigenic *E.coli* (STEC) (Foster-Nyarko & Pallen, 2022).

Desvaux *et al.,* (2020) stated phylogenetically *E. coli* have eight separate groups (A, B1, B2, C, D, E, F, and G), nevertheless Clark & Maresso., (2021) indicated nine groups, and classified furtherly into four major (A, B1, B2 and D) and five minors (C, E, F, G and cryptic clade I). Non-pathogenic commensal *E. coli* strains are mostly found in groups A and B1, while extraintestinal pathogenic *E. coli* strains are found in groups B2 and D (Książczyk *et al.*, 2021, Clark & Maress., 2021). B1 phylogroup strains may also be more common in domesticated animal isolates. Phylogroup E, on the other hand, comprises the majority of known EHEC isolates, including the well-known O157: H7 serotype. Phylogroup F is quite like phylogroup B2, although they lack many of the ExPEC-associated virulence factors, with the exception of a few sequence types (Clark & Maresso, 2021).

Pathogens cause a disease because of having virulence factors in their genes. Virulence factors are specialised particles synthesised and released by the organism, typically proteins. In bacterial pathogens, these components are encoded by genes on the chromosome or mobile genetic elements (plasmids or transposons) (Pakbin, Brück & Rossen, 2021). According to Pakbin, Brück & Rossen, (2021) also, the virulence factors may be able to differentiate between four major groups of *E. coli* pathotypes: colonisation, fitness, toxins, and effectors. Each of these groups would include a variety of different virulence factors, each of which would have a specific role and activity.

E. coli is the leading of antibiotic cause resistance and some studies have described multidrug-resistant *E. coli* strains found in the environment because of

frequent encounters with antibiotics since they usually exist in the intestines of warm-blooded animals (Jang *et al.*, 2017). The acquisition of resistance or virulent traits and the increase of antimicrobial resistance among the *E. coli* strains constitute a major obstacle to treatment and contribute to increasing numbers of morbidity and deaths and increasing healthcare costs associated with *E. coli* infections. Many antimicrobial genes are injected into conjugative plasmids that may also transport virulence factor determinants and might be designated by antibiotic selective pressure (Čurová *et al.*, 2020).

2.2 Genome of *E. coli* strains

Circular chromosome and plasmids are the components that make up the *E. coli* genome. The genome of *E. coli* strains may be anywhere from 4.2 to 6.0 Mbp in size, which translates to 3,900 to 5,800 genes, respectively (Denamur *et al.*, 2021). Studies of the *E. coli* pangenome found that accessory genes, plasmids, phages, and pathogenicity islands have undergone differential gains and losses due to gene horizontal transfer (HGT), which accounts for this variance (Hall *et al.*, 2021). The whole collection of genes that are present in a species is referred to as its pangenome. This set includes the core genes that are present in every member of the species as well as the accessory genes that are present in just some of the strains (Domingo-Sananes & McInerney, 2021). All *E. coli* strains contribute about 2,000 genes (core genomes), with the pan-genome being used to balance out the genetic make-up of a strain (Denamur *et al.*, 2021) which could be probably cover around 40.9% of the total sequence (Messerer, Fischer & Schubert, 2017), accessory genes would be approximately 51%, while persistence genes take 3% and singletons (species-specific genes) genes takes less than 1% in the average genome (Touchon *et al.*, 2020). In addition to that *E. coli* has approximately 87% protein-coding sequences and 50.6% GC content (Touchon *et al.*, 2020).

Genes are constantly being added to and removed from the *E. coli* genome, which may result in a shift in the overall pattern of the genome. It was predicted that when additional genomes from the same species were examined, the number of core genes would decrease and the number of accessory genes would increase (Domingo-Sananes & McInerney, 2021). Differences in general genome sequences have the

possibility to influence which other genes can or cannot be found in a genome. A gene might be beneficial to one strain of species, but harmful to other strains. So, a study of an *E. coli* pangenome consisting of 400 genomes from 20 distinct sequency types (STs) was constructed for the purpose of identifying the gene-gene relationships. The study found that of all the gene clusters in the gene presence-absence matrix, 45.0% form at least one co-occurrence pair, and 17.9% form at least one avoidance. When it comes to the accessory genes that Coinfinder makes use of, 77.8% of them form at least one co-occurring or avoidant pair (Hall *et al.*, 2021). The authors went on to say that co-occurrence hub genes are linked to MGE and virulence (**Figure 2.1**). They do one of two things for many other genes that are present in any given genome: they either enable or promote their presence. For instance, the functions that Shiga toxin and CRISPR system components play in both defence and offence are the ones that are identified the most often. During this time, the avoidance hub genes are being enhanced for the goals of the secretion system. One example of this is the gene spiA, which encodes a type III secretion system (T3SS) outer membrane secretin. SpiA is also known as escC.

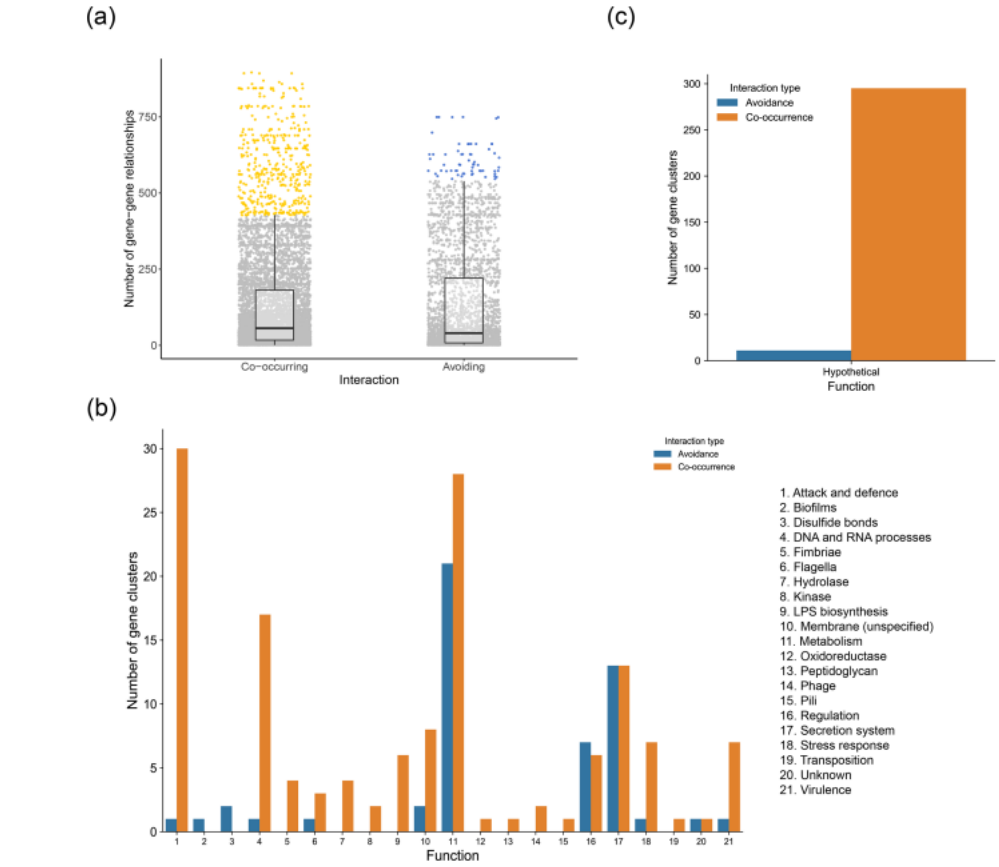

Figure 2.1 Hub genes have high connectivity in the accessory genome.

(A) The quantity of gene-gene connections made by specific genes. Hub genes are coloured orange (co-occurrence) and blue (1.5 times the upper IQR) (avoidance). (b) Hub genes that co-occur (orange) and avoid (blue), organised by loose biological function. Subunits of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system, toxin-antitoxin systems, toxin synthesis, and detoxification are all part of attack and defence. Replication, repair, recombination, and plasmid segregation are among the activities involving DNA and RNA. (c) The quantity of hub genes that code for fictitious proteins (Hall *et al.*, 2021).

According to the host association, Touchon *et al.*, (2020) observed that strains from chicken meat had the biggest mean genomes, followed by human ExPEC strains having the next greatest mean genomes. The shortest genomes were found in strains that originated from fresh water and the faeces of wild birds.

2.3 Pathogenicity of *Escherichia coli*

Although most strains of *E. coli* are commensal that don't cause a disease unless immunocompromised patients, there are some pathological strains that cause more serious intestinal and extraintestinal diseases. Nine different pathotypes of *E. coli* strains that have been isolated from humans have been documented in a few investigations. Seven of them are intestinal pathogens, including Enterotoxigenic *E. coli* (ETEC), Enteroinvasive *E. coli* (EIEC), Enteropathogenic *E. coli* (EPEC), Enterohaemorrhagic *E. coli* (EHEC), Enteroaggregative *E. coli* (EAEC), and Diffusely adherent *E. coli* (DAEC), and one identified in recent years is Adherent-Invasive *E. coli* (AIEC) (Pakbin, Brück & Rossen, 2021)which causes mostly inflammatory bowel disease (IBD) more than the diarrhoea (Desvaux *et al*., 2020).

Uropathogenic *E. coli* (UPEC) and neonatal meningitis *E. coli* (NMEC) are considered the most known ExPECs (Clark & Maresso, 2021). One study suggested that UPEC is the main origin of septicaemias and human sepsis-associated *E. coli* (SEPEC). They have found that 90% of the *E. coli* strains isolated from blood and urine are almost related to each other (Desvaux *et al.,* 2020). In poultry, avian-pathogenic *E. coli* (APEC) is an ExPEC that produces a disease known as avian colibacillosis, which is the primary source of infection in the poultry industry in the world and may serve as an external reservoir for human infection (Jørgensen *et al.*, 2019).

In the following paragraphs, more information regarding the pathogenicity of each of the E. *coli* strains is explored.

2.3.1 Enteropathogenic *Escherichia coli* (EPEC)

The first pathotype of diarrheagenic *E. coli* to be recognized was EPEC, which was responsible for a few epidemics of infantile diarrhoea between 1940 and 1950 (Mare *et al.,* 2021). EPEC infection is well known for causing acute diarrhoea and vomiting in infants and is well characterised in its attaching and effacement (A/E) histopathology (Desvaux *et al.,* 2020) for having locus of enterocyte effacement (LEE) as a pathogenicity island on its chromosome, these genes allow them to adhere to intestinal epithelial cells and cause A/E diarrhoea (Jarocki *et al.*, 2020). This is attended by intimate attachment of bacterial cells articulating cell-surface intimin (Eae) with the related translocated intimin receptor (Tir) on the surface of the host cell resulting inoculation by subtype a, type III secretion system (T3aSS) (Desvaux *et al.*, 2020). In addition to the LEE, EPEC has several non-LEE (nle)-encoded effector genes, they may increase bacterial virulence, and disrupt phagocytosis and apoptosis by interfering with host signalling pathways (Jarocki et al., 2020) and preventing human or animal inflammatory responses (Mare *et al.*, 2021).

EPEC is further classified into typical EPEC (tEPEC) and atypical EPEC (aEPEC) in the presence and absence of the EPEC adherence factor plasmid (pEAF), which encodes bundle-forming pili. Another distinction is that, in contrast to aEPEC, which has over 400 serogroups and includes tEPEC that has lost the pEAF and EHEC that has lost the stx genes, tEPEC belongs to just twelve classical EPEC serogroups and humans are their main reservoir (Jarocki *et al.*, 2020). Because enterohemolysin (ehxA) and heat-stable enterotoxin are aEPEC's virulence factors, they may have an impact on both people and animals (Jarocki *et al.*, 2020). All the EPEC strains are eae positive (+), however aEPEC strains are considered those eae positive + (the gene that

encoded the intimin protein which is essential for the cell attachment) and bfpA (bundlin pilin protein of bundle-forming pilus) − (negative) (the genes that encodes the type IV bundle-forming pilus) while tEPEC are described as eae positive (+) and bfpA positive (+) (Mare *et al.*, 2021).

More about *E. coli* strains, disease presentation and associated virulence factors are demonstrated in the (**Table 2.1**).

Table 2.1 ExPEC and InPEC *E. coli* pathotypes, abbreviation, disease presentations, and associated virulence factors (VFs) (Clark & Maresso, 2021).

Pathotype	**Acronym**	**Disease presentation(s)**	**Associated VFs**
Uropathogenic *E. coli*	UPEC	Urinary tract infections	*im, hlyA, csg, pap, sfa, afa, cdtAB, iha, iutA, iroN, fyuA, sitA, chuA, hma, kpsMT, agn43/flu, pic, sat, vat, usp*
Neonatal Meningitis *E. coli*	NMEC	Bacterial meningitis	*im, sfa, mat, ibeA, irp, iroN, kpsMT, K1*
Avian Pathogenic *E. coli*	APEC	Multiple ExPEC diseases in avian species	*olicin, ibeA, iutA, iroN, sitA, tsh, fim, fyuA, pap, vat*
Adherent-Invasive *E. coli*	AIEC	Associated with intestinal inflammation	*imH, lpf, pap, sfa, afa, vat, hlyA, cnf I, cdtAB, ibeA*
Enterohemorrhagic *E. coli*	EHEC	Bloody diarrhea, hemorrhagic colitis, hemolytic-uremic syndrome	*EE pathogenicity island; stx, espP, lpf, efa, toxB, eibG, ehaA, ompA, iha, paa*
Enteroaggregative *E. coli*	EAEC	Acute and chronic watery diarrhea	*et, agg, aaf, agg3, astA, pet, sat, aap, aagR, shf, pic, irp2, hly, tia*
Enteroaggregative Hemorrhagic *E. coli*	EAHEC	Similar to EHEC, with increased adherence and antibiotic resistance	*iha, pic, pet, stx*
Enterotoxigenic *E. coli*	ETEC	Mild to severe watery diarrhea	*CFA fimbriae, astA, eltAB, estIa, clyA, eatA*
Enteropathogenic *E. coli*	EPEC	Severe acute watery diarrhea	*LEE pathogenicity island, set, paa, lpf, iha, ehaA*

Mare *et al.*, (2021) described four main mechanisms that EPEC interact with the host cells (**Figure 2.2**), those were previously mentioned by Nataro and Kaper, in 1998. The first mechanism, EPEC starts to express Bfp and EspA (short filaments service associated) to attribute the host cells. Then in the second step, EPEC uses its type III secretion system to create a pore and inject Tir into the host cell. Then the adherence of EPEC to intestinal cells will occur, this process facilitates bacterial colonization, and impacts the cytoskeletal system causing loss of microvilli. The third step bacterial strengthens the adhesion by binding the modified Tir, and the cytoskeletal and actin are accumulated the adhesion area, finally and the fourth step the formation of the pedestal structure, which is a characteristic structure of the EPEC appears and causes host cell death (Mare *et al.*, 2021).

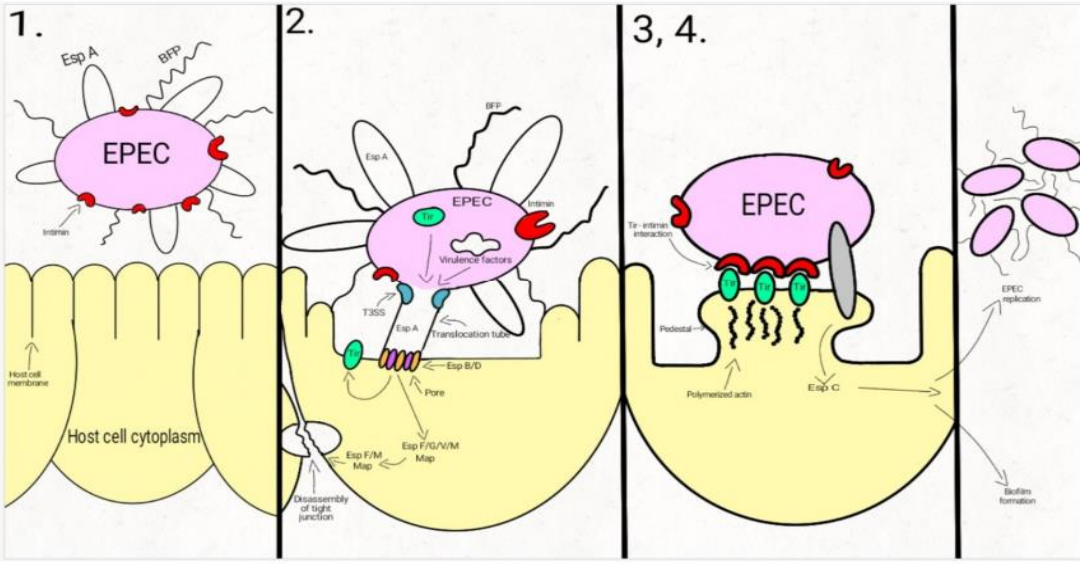

Figure 2.2 Description of four mechanism that EPEC interact with host cell.

1.EPEC make Bfp and EspA. 2. After attaching to the enterocyte through Bfp, EPEC use a T3SS to infect many effectors in the cell. 3. The bacterial intimin binds to the changed Tir, attaching the bacteria to the host cell. Actin and parts of the cytoskeleton build up near where the bacteria stick to the cell. EspC gets into the cell through a system called T5SS, which is an autotransporter. 4. When cytoskeletal parts gather near the place where the cell is attached, they form the pedestal structure that is unique to EPEC (Mare *et al.*, 2021).

2.3.2 Enterohemorrhagic *Escherichia coli* (EHEC)

EHEC is widely recognised for its ability to cause watery paediatric diarrhoea as well as haemorrhagic colitis (HC), both of which may further develop to thrombotic microangiopathies, haemolytic and uremic syndrome (HUS) and thrombotic thrombocytopenic purpura (TTP) (Desvaux *et al*., 2020). Cattle are the main reservoir of EHEC (Pakbin, Brück & Rossen, 2021). This organism was first isolated from a patient suffering from bloody diarrhoea and GIT disorder in 1982 and has developed into a worldwide pandemic (Pakbin, Brück & Rossen, 2021).When compared to the infectious doses required by other intestinal infections, the infectious dosage required by EHEC is exceptionally low, since only 10 to 100 organisms are needed to produce sickness (Cameron *et al.*, 2018).

The main virulence factors of EHEC are a Shiga-like toxin (SLT), also called verotoxin, it is encoded by stx genes. This toxin is a very powerful toxin and is responsible for several clinical manifestations leading to specific disease symptoms such as renal failure and HUS (Pakbin, Brück & Rossen, 2021) and a type III secretion system (T3SS) which is a device that looks like a syringe and is used to transfer bacterial effectors into the host cell. These bacterial effectors either mimic or take over the function of the host cell. The LEE pathogenicity island contains the genetic coding for the EHEC T3SS. Proteins EspA, EspB, and EspD, which are encoded by the LEE gene family, combine to make up the translocon of the EHEC T3SS. A sheath is formed by EspA around the needle, and a hole is formed in the membrane of the host cell by EspB and EspD (Cameron *et al.*, 2018) . Shiga toxin (SLT) is further divided into Stx1 and Stx2, as well as subtypes like Stx2a, Stx2c, and Stx2d. These subtypes of Shiga-toxin are strongly linked to HUS and HC (Pakbin, Brück & Rossen, 2021).

E. coli collected from other sources such as environment, animals, food, and the agri-food chain can be genotypically considered as possessing the *stx* gene and can be rearranged as shigatoxin-encoding *E. coli* and with more than 400 distinct serotypes discovered in STEC, however, only a minority of EHEC serotypes has been clinically associated with epidemic outbreaks, although they can all be potentially involved in human infections (Desvaux *et al.*, 2020).

2.3.3 Enterotoxigenic *Escherichia coli* (ETEC)

ETEC is the major cause of travellers' and children's diarrhoea in low-income countries, especially in children under two years of age, which causes the highest mortality rates. According to the WHO reports, ETEC affects more than 157,000 people annually, which leads to death (Pakbin, Brück & Rossen, 2021). The pathogenicity of ETEC mostly relies on heat-labile (LT) and/or heat-stable (ST) enterotoxins that lead to a net discharge of intestinal fluid. The diarrhoea is usually not bloody but watery as cholera and is sometimes associated with mild or severe vomiting (Desvaux *et al.*, 2020). According to data from 2013, ETEC-caused diarrhea caused an average of 42,000 fatalities in children under the age of five and more than 89,000 deaths in persons over the age of five throughout Africa and South Asia (Mirhoseini *et al*., 2018).

Mostly, ETEC uses colonization factors and enterotoxins as virulence factors (Pakbin, Brück & Rossen, 2021). Surface colonization of ETEC is responsible by pili known as colonization factor antigens (CFAs) which are capable to bind different receptors of the host cells (Desvaux *et al.*, 2020). The infectious dose of ETEC is about 106 to 1010 numbers of the organism, bacteria colonize and attach to the small

intestine once it reached there, after the attachment the organism immediately produces the enterotoxins and affected the intestinal epithelial cells (Mirhoseini *et al*., 2018). The heat-labile enterotoxins are divided into two types: heat-labile type I (LT-I) and heat-labile type II (LT-II) which are distinguished by their immunological, genetic, and chemical properties. Type I (LT-I) enterotoxin facilitates the entry of the organism into the host cell by binding different ganglioside receptors, type I (LT-I) also deactivates the GTPase action of the G protein complex by ADP-ribosylation, this in sequence, causes to the continued activation of the adenylate cyclase action and raised intracellular levels of cAMP, resulting loss of water and electrolyte (**Figure 2.3**) (Desvaux *et al*., 2020, Mirhoseini *et al*., 2018). Type I (LT-I) is a member of the cholera toxin family and is produced by a Type II secretion system (T2SS) with the LeoA (labile enterotoxin output) involvement for its effective secretion (Desvaux *et al*., 2020). In contrast LT-II toxins neither do relate to the cholera toxin nor bind to the ganglioside receptors and almost the pathogenicity of this toxin is not well described (Mirhoseini *et al*., 2018).

On the other hand, the other type of enterotoxins are STs which can be furtherly classified into STa (STI) and STb (STII). In addition, STp and STh are two STa variants (García & Fox, 2021).They bind different receptors, however, they seem to have the same function, ST-I activates a guanylate cyclase C receptor and the following increase in intracellular cGMP levels activates the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel, while ST-II binds the acidic glycosphingolipid sulphatide, and activates a pertussis toxin-sensitive GTP-binding regulatory protein that eventually activates CFTR (Desvaux *et al.*, 2020). Activation of the CFTR channel prevents ion exchange and sodium reabsorption,

leading to the release of water and salt into the intestinal lumen and thus net fluid loss, causing watery diarrhea (Pakbin, Brück & Rossen, 2021)

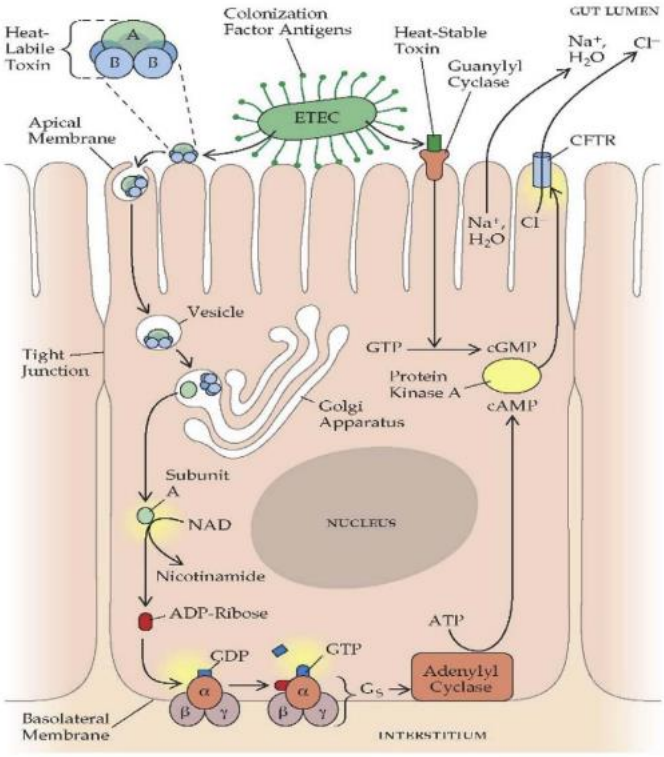

Figure 2.3 The pathophysiology of ETEC (Mirhoseini *et al*., 2018).

The pathogenesis of ETEC involves two pathways, the first bacteria colonize the intestine by its adhesion factors, secondly the organism secrets enterotoxins that activate of adenylyl and guanylate cyclase leading to formation of cAMP and cGMP, these stimulate water and electrolyte loss by intestinal endothelial cells.

2.3.4 Enteroinvasive *Escherichia coli* (EIEC)

EIEC and different species of *Shigella* cause severe mucosal and bloody diarrhea with abdominal pain and fever as aggressive foodborne pathogens and originally recognized as *Bacillus dysenteriae* or *Shigellosis* (Pakbin, Brück & Rossen, 2021). Invasion of bacteria into colonic and rectal mucosa explains well to these symptoms leading to a powerful inflammatory response that elicits the destruction of the colonic epithelium. Hypoglycemia, bacteremia, septicemia, and haemolytic uremic syndrome leading to acute renal failure and perforation are life-threatening problems that may also happen

as a complication (Belotserkovsky & Sansonetti, 2018). Although some studies have found *Shigella* to be more virulent than EIEC, they are virtually indistinguishable due to their shared species, pathotype, and intestinal illness (Desvaux *et al.*, 2020).

Multi locus sequence typing and genome sequency study of some EIEC strains (M4163 from a cheese-related outbreak in 1971 and 4608-58) described that the *Shigella* genome is less than to *E. coli* genome. Relating to lactose utilization also the study determined that the EIEC lactose utilization is variable while *shigella* doesn't utilize lactose (García & Fox, 2021). The virulence factors of EIEC/*Shigella* are the invasion plasmid pINV, which encodes T3aSS, and pathogenicity islands (PIs) acquired by horizontal gene transfer, Pic mucinase (formerly called ShMu for hemagglutinin and *Shigella* mucinase), which represent the initial stage of intestinal colonization mediate, and SigA (*Shigella* immunoglobulin A protease), which plays a role after the incubation phase of infection (prodromal phase), are the other virulence factors of EIEC/*Shigella*. In addition, Shiga enterotoxin 1 (ShET1) and Shiga enterotoxin 2 (ShET2) are the most important virulence factors. (Belotserkovsky & Sansonetti, 2018, Desvaux *et al*., 2020) . ShET1 is detected on the chromosome and a pathogenicity island, whereas ShET2 is located on the plasmid. ShET2 activates inflammation in the intestinal epithelial cells, and both involve intestinal secretory activity (Pakbin, Brück & Rossen, 2021).

Shigella/EIEC uses two pathways to invade the host cell, the first pathway is through the microfold (M) cell which is an intestinal epithelial specialized cell (IECs) that transports antigen from the gut lumen to the underlying mucosal lymphoid tissue, and once the organism passed through the M cells, the organism invades intestinal

epithelial cells from their basolateral side with high effectiveness. The second pathway is through filopodia (finger-like projections found on the apical side of intestinal epithelial cells), the organism attaches to the tip of the filopodia leading to its retraction towards the cell body and finally, the invasion occurs (Belotserkovsky & Sansonetti, 2018).

Specialized M cells allow *Shigella* to cross the intestinal epithelium. After being phagocytosed by macrophages, which are frequently present near M cells, the bacteria are then discharged to the lamina propria after escaping the resultant phagosome. *Shigella* effectively invades these cells and disseminates throughout the epithelium by exploiting actin-based motility as it has access to the baso-lateral side of the IECs. Neutrophils are drawn to the infection site because of the production of pro-inflammatory cytokines (like IL-8) and "danger-associated" molecules (like ATP) by IECs and pyroptotic macrophages. Although neutrophils act as antimicrobials, *Shigella/*EIEC causes massive destruction and apoptosis of B cells, T cells, and dendritic cells (DC) (Belotserkovsky & Sansonetti, 2018). (**Figure 2.4**).

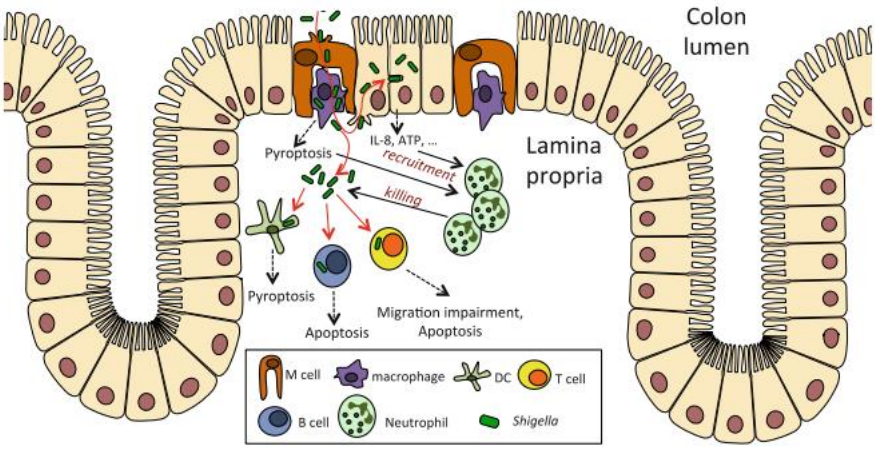

Figure 2.4 Interaction of *Shigella*/EIEC during invasion with immune cells and the epithelium (Belotserkovsky & Sansonetti, 2018).

2.3.5 Enteroaggregative *Escherichia coli* (EAEC)

EAEC, formerly called as "EAggEC," causes acute or chronic watery diarrhoea with or without mucus, most commonly in children from low-income nations. As a result, in high-income nations, EAEC is linked to traveller's diarrhoea in both children and adults (Jenkins, 2018). Furthermore, EAEC is the second most important reason for acute and persistent traveller's diarrhoea after ETEC in high- and low-income countries and one of the main causes of GIT infections in patients with HIA/AIDS (Pakbin, Brück & Rossen, 2021).

The diarrhoea of EAEC is non-bloody without inflammation and vomiting. EAEC forms untied biofilm at the mucosal surface relating to pili and especially aggregative adhesion fimbriae, encoded by a large plasmid (pAA) through adhesion onto intestinal epithelial cells (Desvaux *et al.*, 2020). Aggregative adherence fimbriae (AAF) are the putative virulence factors that are fixed on the pAA virulence plasmid and are always responsible for the adherence, activation of mucus secretion, biofilm

formation, and other effects. In the absence of the AAF, other virulence factors like *E. coli* common pilus (ECP) may mediate aggregative adherence (Ellis *et al.*, 2020). The hallmark virulence regulator of EAEC is AggR, one of the AraC/XylS family of bacterial transcriptional regulators and the distinguishing feature of typical EAEC strains. aggR is found on the pAA plasmid and regulates several genes encoding putative virulence factors located on the pAA and others on the chromosome that suggests most virulence factors in EAEC are regulated by AggR (Jenkins, 2018).

The pathogenesis of EAEC can be summarised by adherence to the intestinal mucosa via aggregative fimbriae, elevated mucus secretion leading to wide biofilm formation on the surface of the cells, and toxin secretion producing an inflammatory response (Jenkins, 2018). EAEC toxins are several types for example plasmid-encoded toxin, *shigella* extracellular enterotoxin, a secreted autotransporter toxin, enteroaggregative heat-stable toxins, and protein-involved enterocyte colonization (Pakbin, Brück & Rossen, 2021). Serine Protease Autotransporters of Enterobacteriaceae (SPATEs) are extracellular proteases that may have proteolytic ability and are produced via the type V secretion system. SPATEs are concerned with immune evasion, mucosal damage, and colonization. The most frequently found SPATEs in EAEC contain plasmid-encoded toxin (Pet), a protein involved in intestinal colonisation (Pic), secreted autotransporter toxin (Sat), *Shigella* IgA-like protease homology (SigA), and *E. coli*-secreted protein (EspP) (Jenkins, 2018).

2.3.6 Diffusely Adherent *Escherichia coli* (DAEC)

DAEC is related to diarrhoea in children under 5 years old in developing and developed countries. It is concerned with UTIs and pregnancy complications as an ExPEC and can be found in asymptomatic children and adults. DAEC produces watery