Intestinal pathogenic *Escherichia coli*: identification and characterisation of their virulence determinants

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Jennifer Caitlin Hallam BSc (Hons)

Submitted March 2022

Institute of Infection, Immunity and Inflammation

College of Medical, Veterinary and Life Sciences

University of Glasgow

Acknowledgments

I would first like to thank my supervisor, Professor Andrew Roe, for his continued support and guidance throughout my PhD. I will always be so thankful for this opportunity, and grateful for all your advice over the years. I would also like to thank my second supervisor, Professor Daniel Walker. Unfortunately, I did not get to work with you as much as I hoped, but I am grateful for your support. I am also thankful to Professor Konstantinos Gerasimidis and Dr Gillian Douce, who encouraged me to come out of my shell and helped me push through the challenging points of my PhD.

I feel extremely lucky to have worked in the Roe lab, and my experience would not have been the same without the constant support from members of the group, both past and present. A special thank you goes to Dr Nicky O'Boyle, whose continued patience and guidance over the years has been particularly instrumental. I'd also like to extend my gratitude to the entire bacteriology department for all their support, with a special mention of thanks to Dr Michael Ormsby and Dr Khedidja Mosbahi for their help and advice over the years. I will forever treasure the memories I've made here. All the laughs, tears, and copious cups of coffee – it has been a rollercoaster that I am so fortunate to have shared with such incredible individuals. I know I have made lifelong friends among many of you, but especially Dr Natasha Turner, whose friendship has been invaluable.

I am so grateful to have had the constant support of parents, Lesley and Tomi. You have always believed that I could achieve my dreams and I appreciate all that you do to support me. Thank you also to David, my uncle, for sharing your wisdom and guidance. Lastly, Connor, you have been my rock. Your continued love and support (and late-night taxi services) have helped me to complete my PhD still smiling.

Table of contents

List of Tables

List of Figures

Abbreviations

Δ	-	deletion
AA	-	aggregative-adherence
A/E	-	attaching and effacing
AEIC	-	adherent-invasive *Escherichia coli*
AMR	-	antimicrobial resistance
ARP2/3	-	actin-related protein 2/3
BFP	-	bundle forming pilus
BSA	-	bovine serum albumin
CDEC	-	cell-detaching *Escherichia coli*
CFU	-	colony forming units
ChIP	-	chromatin immunoprecipitation
ChIP-Seq	-	chromatin immunoprecipitation coupled with next generation sequencing
CRC	-	colorectal cancer
DA	-	diffuse adherence
DAEC	-	diffusely adherent *Escherichia coli*
DEC	-	diarrheagenic *Escherichia coli*
DEG	-	differentially expressed gene
DGE	-	differential gene expression
DMEM	-	Dulbecco's Minimal Eagle Medium
DSBs	-	double stand breaks
dsDNA	-	double stranded DNA
EAEC	-	enteroaggregative *Escherichia coli*
EAF	-	EPEC adherence factor
EHEC	-	enterohaemorrhagic *Escherichia coli*

EIEC	-	enteroinvasive *Escherichia coli*
EMSA	-	electrophoretic mobility shift assay
EPEC	-	enteropathogenic *Escherichia coli*
ENA	-	European Nucleotide Archive
aEPEC	-	atypical enteropathogenic *Escherichia coli*
tEPEC	-	typical enteropathogenic *Escherichia coli*
ETEC	-	enterotoxigenic *Escherichia coli*
ExPEC	-	extraintestinal pathogenic *Escherichia coli*
FAS	-	fluorescent-actin staining
FCS	-	foetal calf serum
FMT	-	faecal microbiota transplant
GFP	-	green fluorescent protein
GI	-	gastrointestinal
HGT	-	horizontal gene transfer
HUS	-	haemolytic uremic syndrome
IBS	-	inflammatory bowel syndrome
IBD	-	inflammatory bowel disease
ICLs	-	interstrand cross-links
InPEC	-	intestinal pathogenic *Escherichia coli*
LA	-	localised adherence
LAL	-	localised-like adherence
LB	-	Luria-Bertani broth
LEE	-	locus of enterocyte effacement
Ler	-	LEE-encoded regulator
LIC	-	low-income country
LPS	-	lipopolysaccharide

LTTR	-	LysR-type transcriptional regulator
MEM-HEPES	-	Minimal Essential Medium, HEPES modification
MGE	-	mobile genetic elements
MOI	-	multiplicity of infection
N-WASP	-	neural Wiskott-Aldrich syndrome proteins
nfH_2O	-	nuclease free water
OMP	-	outer membrane protein
PAI	-	pathogenicity-associated island
PBS	-	phosphate buffered saline
PBST	-	phosphate buffered saline Tween
PCR	-	polymerase chain reaction
PFA	-	paraformaldehyde
rCDI	-	recurrent *Clostridium difficile* infection
RFP	-	red fluorescent protein
RFU	-	relative fluorescence units
RNA-Seq	-	RNA sequencing
RT-qPCR	-	quantitative real time PCR
SBS	-	single base substitutions
SDS	-	sodium dodecyl sulphate
SDS-PAGE	-	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM	-	standard error of the mean
TF	-	transcription factor
Tir	-	translocated intimin receptor
T3SS	-	type three secretion system
UPEC	-	uropathogenic *Escherichia coli*
UTI	-	urinary tract infection

WT	-	wild-type
VNTR	-	variable number of tandem repeats

Abstract

The human microbiota consists of over 100 trillion microbial cells, primarily, bacteria that reside in the gastrointestinal tract (Turnbaugh *et al.*, 2007). The relationship between these bacteria and the host is typically symbiotic, as many microbiota-derived species play an important role in host physiology. Thus, maintaining the balance of the intestinal flora, and controlling the overgrowth of potentially pathogenic species, is important in preventing disease. Indeed, specific changes to the bacterial composition of the gut can initiate chronic inflammation of the intestine and result in the onset of gastrointestinal disorders such as inflammatory bowel disease (IBD) (Baldelli *et al.*, 2021). The rapid advancement of sequencing methods and analytical techniques has enabled researchers to understand the microbiome and to ask the question, how do we define the microbiome and its inhabitants? There are many factors that can influence one's microbiome, including the host genotype, lifestyle and environment (Turnbaugh *et al.*, 2007). Therefore, it is not surprising that the gut microbiome can be up to 80-90% different between individuals, a stark contrast to the human genome, where individual humans are approximately 99.9% identical (Ursell *et al.*, 2012).

The composition of the gut microbiota is highly variable between geographical locations, particularly between humans living in markedly different socio-economic settings (Yatsunenko *et al.*, 2012). Moreover, in certain low-income countries (LICs), many enteric pathogens are endemic throughout these regions and are frequently isolated from stool of healthy individuals, suggesting that organisms can be carried asymptomatically. Enteropathogenic *E. coli* (EPEC) is typically associated with diarrhoeal disease in infants, however, EPEC is frequently isolated from healthy individuals as well as from those with gastroenteritis. Indeed, it remains unknown how these 'pathogens' colonise individuals

without any symptoms, and begs the question whether these can truly be defined as pathogens at all? Therefore, the primary aim of this thesis was to isolate and characterise the virulence determinants of asymptomatic EPEC isolates. And secondly, to identify a novel strategy for clearing these strains without compromising the composition of the host microbiota. Unfortunately, due to ceased contact with my industrial collaborator, no asymptomatic EPEC strains were received and thus these investigations did not produce any conclusive findings.

In consequence, the focus of this project was redirected towards investigating virulence in a different subset of *E. coli,* specifically strains belonging to the B2 phylogroup which harbour the *pks* island, a genomic island encoding the biosynthesis genes for a genotoxin termed colibactin. In eukaryotic cells, colibactin induces DNA damage resulting in chromosomal instability and cell cycle arrest. Strikingly, *E. coli* strains harbouring the *pks* island are often overrepresented in biopsies from IBD patients and furthermore, the production of colibactin has been implicated in the promotion and development of colorectal cancer in these individuals (Arthur *et al.*, 2012). Thus, as the composition of the human diet has been identified as a key factor in governing intestinal homeostasis, the effect of dietary amino acids on the expression of *pks* encoded genes was explored using RT-qPCR. Several D-amino acids exhibited the ability to inhibit expression of *clbB* with D-Serine exerting the strongest repressing activity in two *pks*$^+$ *E. coli* strains CFT073 and Nissle 1917. The effect of D-Serine on the colibactin induced cellular response was also observed during infection of HeLa cells with live *pks*$^+$ strains. Levels of γ-H2AX (a marker of DNA double strand breaks) were significantly reduced in cells infected with D-Serine treatment. Moreover, exposure of *pks*$^+$ *E. coli* to D-Serine during infection caused a reduction in cellular senescence that was observable at 72 hours post infection. These findings have revealed the potential of D-amino acids in reducing colibactin expression in distinct *pks*$^+$ *E. coli* strains. Furthermore, D-Serine, and other D-amino acids are key

components of our diet and therefore present opportunities for dietary supplementation that might yield important health benefits by modulating gene expression of *E. coli pks*$^{+}$ strains present in the microbiota.

Author's declaration

I hereby declare that this thesis is the result of my own work and has been composed for the degree of PhD at the University of Glasgow. This work has not been submitted for any other degree at this or any other institute. All work presented was performed by myself unless otherwise stated. All sources of information and contributions to the work have been specifically acknowledged in the text.

Jennifer Caitlin Hallam

March 2022

1 Introduction

1.1 *Escherichia coli* physiology

Escherichia coli (*E. coli*) is a gram-negative bacillus that colonises the gastrointestinal tract of warm-blooded animals and humans. As the primary facultative anaerobe of the resident microflora, and one of the first colonisers of the human intestine after birth, *E. coli* has an important role in maintaining intestinal physiology (Salomon and Farias, 1992; Palmer *et al.*, 2007; Secher, Brehin and Oswald, 2016). Since its first discovery in the late 19th century by Dr Theodor Escherich, this bacterium has helped to dramatically transform the field of molecular biology. Indeed, *E. coli* has often been referred to as the "laboratory workhorse", typically owing to its widespread use as a model organism and as an effective cloning host for the study of biological phenomena.

In the 1940s, a defined serotyping system was proposed by Friz Kauffmann, the O:H system, that classified strains of *E. coli* based on their major surface antigens, O antigen: part of lipopolysaccharide layer; H: flagellin; K: capsule, for example O157:H7 (Friz Kauffmann, 1947; Orskov *et al.*, 1977). Serotyping became important for studying the epidemiology and pathogenesis of *E. coli* strains, making it possible to compare strains with more accuracy and allowing strains to be subdivided into distinct groups, based on their association with a specific disease outcome. The first subgroup to be defined was enteropathogenic *E. coli* (EPEC) and was proposed by John Bray in 1945, after strains isolated from infants with diarrhoea were found to be antigenically homogenous (Bray, 1945). Traditionally, diarrheagenic *E. coli* (DEC) have been divided into six pathotypes, including enterohaemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC) and diffusely adherent *E. coli* (DAEC) (Kaper, Nataro and Mobley, 2004). Recently, two more pathotypes have been designated such adherent-invasive *E. coli*

(AIEC) and cell-detaching *E. coli* (CDEC) (Pawłowska and Sobieszczańska, 2017). Enteric *E. coli* involved in diarrheal diseases are frequently transmitted via the faecal-oral route, often through consumption of contaminated food sources or through direct person-person spread from insufficient hygiene (Berger *et al.*, 2010). However, the DEC pathotypes each have a unique pathogenicity profile and differ in terms of their preferred site of colonisation, virulence mechanisms and the manifestation of their clinical symptoms and outcomes.

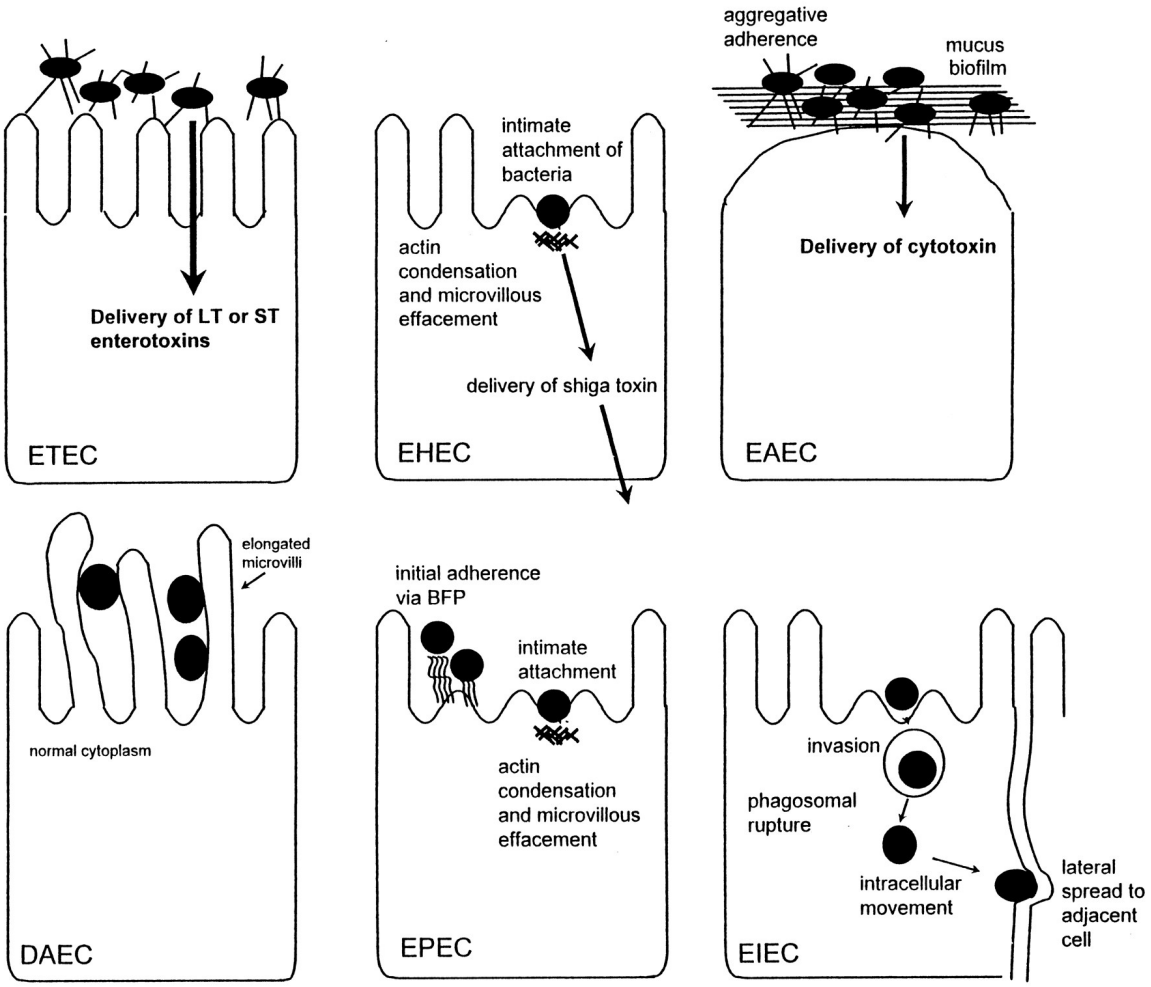


Fig 1-1 A schematic of the six recognised pathotypes of DEC. The illustration by Nataro and Kaper demonstrates the unique interaction of each diarrhoeagenic pathotype with eukaryotic cells (Nataro and Kaper, 1998).

After traversing through the gastrointestinal tract, like most mucosal pathogens, *E. coli* follow a four-step strategy to establish infection, (i) colonisation of a mucosal site; (ii) evasion of host defences; (iii) replication; (iv) host damage. To fulfil the first step, DEC use fimbrial appendages with a high affinity for adhering to epithelial cells in the small bowel. Upon adhering to this site, DEC implement distinct pathogenic strategies to establish infection, illustrated here in Fig 1-1 (Nataro and Kaper, 1998). Entertoxigenic *E. coli* (ETEC) are a major cause of Travellers' Diarrhoea, as well as an important cause of childhood diarrhoea in low-income countries (LICs). They are also a common cause of acute diarrheal disease in farm animals, resulting in severe economic losses in farming industries worldwide (Wang *et al.*, 2019). ETEC adheres to receptors in the apical membrane of the small intestine and delivers the heat-stable (ST) and/or the heat-labile enterotoxin (LT), causing fluid loss and inducing secretory diarrhoea. Enterohemorrhagic *E. coli* (EHEC) adhere intimately to the colonic mucosa, forming attaching and effacing (A/E) lesions and produce the potent cytotoxin, Shiga toxin, that results in local and systemic effects in the host, including non-bloody diarrhoea, bloody diarrhoea and haemolytic uremic syndrome (HUS). Enteroaggregative *E. coli* (EAEC) adheres and forms a thick biofilm on the mucosal surface and causes intestinal secretion and damage. This pathotype is increasingly recognised as a cause of persistent diarrhoea in children and adults throughout the world. Diffusely adherent *E. coli* (DAEC) have been shown to elicit elongation of microvilli *in vitro,* although this has not been confirmed *in vivo.* Enteropathogenic *E. coli* (EPEC) are primarily associated with infant diarrhoea in (LICs), they induce A/E lesions in the small bowel, resulting in intestinal secretion and watery diarrhoea. Enteroinvasive *E. coli* (EIEC) are essentially *Shigella,* and can cause diarrhoea and dysentery by invading the epithelial cells and spreading into adjacent cells, (Nataro and Kaper, 1998; Kaper, Nataro and Mobley, 2004).

Pathogenic *E. coli* can also cause infections at extraintestinal sites. The pathotype known as uropathogenic *E. coli* (UPEC) is the major cause of community-acquired urinary tract infections (UTIs) (Terlizzi, Gribaudo and Maffei, 2017). In addition to the common type 1 fimbriae, some UPEC strains possess a unique type of pili that is encoded by the *pap* genes, known as P fimbriae. These fimbriae are indispensable for UPEC pathogenesis and have an important role in bacterial colonisation through fimbriae-mediated adhesion to uroepithelial cells (Lane and Mobley, 2007). In addition, other adhesins and toxins have also been described as important virulence factors for the pathobiology of UPEC. Another extraintestinal pathotype, the meningitis/sepsis-associated *E. coli* (NMEC) is the predominant Gram-negative etiologic agent associated with meningitis in infants (Wijetunge *et al.*, 2015). NMEC can survive in the bloodstream and invade the meninges by traversing the blood brain barrier, leading to the onset of meningitis in infants. *E. coli* associated meningitis accounts for one of the highest mortality and morbidity rates in neonates (10-30%) and has severe neurological implications in survivors (Wijetunge *et al.*, 2015). However, unlike other pathogenic *E. coli*, distinct virulence traits have not been identified to define NMEC (Logue *et al.*, 2012).

The myriad of virulence factors that distinguish pathogenic from non-pathogenic *E. coli* are often encoded on mobile genetic elements such as plasmids, transposons, bacteriophage and pathogenicity islands. Thus, there is the potential for countless variations of virulence factor combinations in different *E. coli.* Intriguingly, the remarkable genomic plasticity of *E. coli* has been demonstrated in recent years with an increase in reports of strains displaying characteristics of two different DEC. The terms “hybrid-”, or “hetero-pathogens”, have been coined to describe *E. coli* pathotypes that exhibit new combinations of virulence factors, which were previously thought to be unique to each pathotype (Santos *et al.*, 2020). For example, the *E. coli* O104:H4 strain involved in a severe outbreak of HUS in 2011, was characterised as an EAEC/EHEC hybrid after it was

found to be harbouring virulence genes associated with both pathotypes (Bielaszewska et al., 2011; Muniesa et al., 2012). Since the outbreak in Germany, it has become apparent that this was not an isolated event, and in fact, the frequency of hybrid strains is higher than previously thought (Santos *et al.*, 2020). Indeed, due to the dynamic nature of the *E. coli* genome, it is highly likely that these hybrid strains will continue to emerge as significant pathogens that will constitute a significant risk to human health.

1.2 Enteropathogenic *Escherichia coli* pathogenesis

Since the initial discovery by John Bray in 1945, EPEC has emerged as a major cause of diarrhoeal disease in infants and has demonstrated an increased risk of mortality in children under 12 months (Kotloff *et al.*, 2013). EPEC belongs to an over-arching group of diarrhoeagenic pathogens designated the attaching and effacing (A/E) pathogens; referred to as such, because they induce the formation of histopathological lesions, termed A/E lesions on the surface of host enterocytes. This group also includes the human pathogen EHEC, as well as several animal pathogens, such as the mouse-specific pathogen *Citrobacter rodentium* and rabbit-specific enteropathogenic *E. coli* (REPEC) (Robins-Browne et al., 1994; Nataro and Kaper, 1998).

Elucidating the pathogenetic mechanism of EPEC infection, has been a topic of research spanning from the late 1960s. The first insight into EPEC pathogenicity was revealed by Staley *et al.*, 1969, when it was noted that these organisms tightly adhered to cells lining the intestinal mucosa of infant pigs, resulting in the formation of a distinctive ultrastructural histopathologic lesion; an observation that had not been witnessed with any other DEC pathotype (Staley, Jones and Corley, 1969). Subsequent studies confirmed the presence of these lesions in intestinal biopsies of infants infected with EPEC (Ulshen and

Rollo, 1980; Rothbaum *et al.*, 1982), and by the mid-1980s the term "attaching and effacing" had been introduced to describe this EPEC phenotype (Moon *et al.*, 1983; Deborah Chen and Frankel, 2005). In affected epithelial cells, the A/E histopathology is characterised by intimate attachment of bacteria, destruction of the brush border microvilli and proliferation of filamentous actin beneath adherent bacteria to form a cup-like projection, often designated a "pedestal" or "pseudopod" (Moon *et al.*, 1983; Knutton *et al.*, 1989). In light of the realisation that the pedestal was composed of condensed actin, Knutton *et al*.,1989, devised the fluorescent-actin staining FAS test, in which fluorescein isothiocyanate (FITC)-labelled phalloidin binds specifically to filamentous actin in cultured epithelial cells, directly underneath adherent bacteria (Knutton *et al.*, 1989). The development of this assay allowed for the specific diagnostic of EPEC strains and other bacteria capable of forming this histopathology. In addition, the FAS test enabled the screening of clones and mutants which led to researchers elucidating the other genetic factors involved in producing these hallmark lesions.

The A/E phenotype is conferred to EPEC by genes encoded on a 35.6 kb chromosomal pathogenicity-associated island (PAI) designated the locus of enterocyte effacement (LEE) (McDaniel *et al.*, 1995). It encodes all the essential genes for A/E lesion formation, including a type III secretion system (T3SS), seven secreted effector proteins, as well as chaperones and the master regulators of the system (Wong *et al.*, 2011). Indeed, McDaniel *et al*., 1997, demonstrated that the PAI conferred virulence by cloning the entire locus from EPEC strain E2348/69 into the avirulent *E. coli* strain K12; the resulting A/E histopathology confirmed that the LEE was both necessary and sufficient for EPEC pathogenicity (McDaniel and Kaper, 1997). This study also highlighted that the acquisition of a single genetic element was enough to convert a non-pathogenic *E. coli* into a pathogenic strain. The LEE is hypothesised to be transferred horizontally between *E. coli* with distinct chromosomal backgrounds. In support of this notion, sequence analysis of the

locus has revealed that the GC content of the LEE (38%) is distinctively lower compared to the 50.8% that makes up the native *E. coli* chromosome, thus providing some evidence that the locus was acquired through horizontal gene transfer (HGT) from a foreign genome (McDaniel *et al.*, 1995). Furthermore, in EPEC the LEE is located adjacent to the *selC* tRNA locus, which is the same region as EHEC and is also the insertion site of a large PAI in UPEC, suggesting this is a hotspot for the acquisition of mobile genetic elements (Blum *et al.*, 1994). However, alternative insertion sites have been identified suggesting that the LEE PAI has been acquired multiple times throughout the evolution of these A/E pathogens (Frankel *et al.*, 1998).

The LEE PAI is composed of 41 core genes that can be organised into five operons: LEE1-LEE5 (Elliott *et al.*, 1998). Operons *LEE1*, *LEE2* and *LEE3* encode the T3SS-associated inner and outer membrane proteins including the porin EscC and the ATPase EscN (Elliott *et al.*, 1998). *LEE4* encodes the structural proteins required for the T3SS "needle-like" complex, including the filament structure protein EspA and the translocator proteins EspD and EspB (Elliott *et al.*, 1998). *LEE5* encodes intimin and the translocated intimin receptor (Tir), as well as the Tir chaperone CesT (Abe *et al.*, 1999). *LEE1* encodes the master regulator of the system, designated the LEE-encoded regulator (Ler) (Mellies *et al.*, 1999). A regulatory feedback loop on *ler* expression is created by the second master regulator GrlRA, encoded between *LEE1* and *LEE2* (Deng et al 2004). Upon contact with the apical surface, effector molecules are deployed from the bacteria into the host cell via the LEE-encoded T3SS. The EPEC protype strain E2348/69 has seven LEE-encoded translocated effectors (Tir, Map, EspF, EspG, EspH, SepZ and EspB) and fourteen non-LEE encoded (*nle*) effectors encoded on other mobile genetic elements (Iguchi *et al.*, 2009). Translocation of these effectors into host cells enable EPEC to intimately adhere and colonise the mucosal membrane, causing effacement of the brush border microvilli and triggering the formation of A/E lesions (Rosenshine *et al.*, 1992; Platenkamp and Mellies,

2018). In addition, these effectors allow EPEC to survive and proliferate in the host by subverting the immune system and avoiding cell death (Santos and Finlay, 2015).

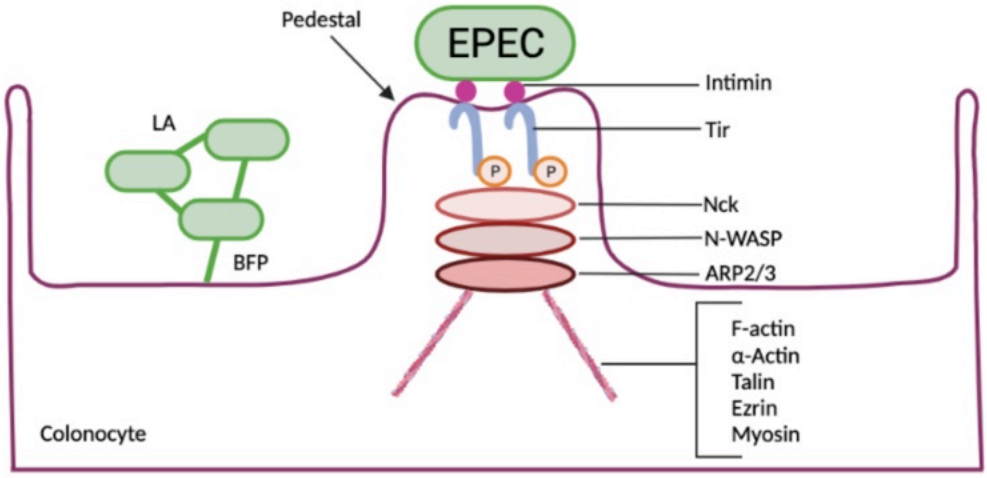


Fig 1-2 The pedestal formation mechanism for enteropathogenic *E. coli.* EPEC attaches to the colon through bundle-forming pilus (BFP), forming localised adhesions (LA) on the cell surface. Intimate attachment is mediated between intimin and the translocated intimin receptor (Tir). Tir is phosphorylated (P) by host tyrosine kinases, and phosphorylated-Tir recruits Nck, which activates neural Wiskott-Aldrich syndrome proteins (N-WASP) and the actin-related protein 2/3 (ARP2/3) complex to mediate actin rearrangements and pedestal formation. Schematic adapted from (Croxen and Finlay, 2009).

A three-stage model has been proposed for the pathogenesis of EPEC, consisting of (i) localised adherence, (ii) signal transduction and (iii) intimate adherence (Donnenberg and Kaper, 1992). In the first stage, bacteria adhere to the microvilli and to each other through the bundle forming pilus (BFP), a type IV fimbria that are encoded on the EPEC adherence factor (EAF) plasmid (Nataro *et al.*, 1987). In addition to these rope-like appendages, it has been suggested that common *E. Coli* pilus may also play a role and work in concert with BFP to stabilise interactions between bacteria and enterocytes (Saldaña *et al.*, 2009). The

second stage is characterised by signal transduction which is initiated by the T3SS system. This apparatus is a key virulence determinant utilised by many enteric pathogens, that functions as a macromolecular syringe by injecting secretory proteins directly from the bacterium into the host cell (Jarvis *et al.*, 1995; Hueck, 1998). Several proteins are translocated by the EPEC secretion system, however, most notably are the proteins EspA and EspB, which are essential for activating signal transduction in epithelial cells (Foubister *et al.*, 1994; Kenny and Finlay, 1995; Kenny *et al.*, 1996). These signals include inositol triphosphate and calcium fluxes, protein-tyrosine phosphorylation and rearrangement of the host cytoskeletal components beneath adherent bacteria (Baldwin *et al.*, 1991; Rosenshine *et al.*, 1992; Foubister, Rosenshine and Finlay, 1994). The third stage of EPEC pathogenesis is characterised by intimate attachment of the bacterium to the host cell. This process is mediated by the interactions between the bacterial outer-membrane protein intimin and Tir, illustrated in Fig 1-2. Briefly, Tir is rapidly translocated into the host cell cytoplasm via the T3SS and displayed on the surface of the plasma membrane, where it acts as a receptor for intimin (Kenny *et al.*, 1997). The interaction with intimin stimulates the clustering of Tir, which is then phosphorylated by multiple host tyrosine kinases (Swimm *et al.*, 2004). The phosphorylation of Tir triggers the recruitment of Nck to the site of attachment and activates the host-cell factors N-WASP and the ARP 2/3 complex, initiating the assembly of highly organised cytoskeletal structures immediately under the adherent bacteria and preceding the formation of pedestal structures (Baldwin *et al.*, 1991; Finlay *et al.*, 1992; Kalman *et al.*, 1999; Gruenheid *et al.*, 2001).

The complete sequence of the LEE from the EPEC strain E2348/69, was revealed in 1998 and therefore many studies adopted this prototype to study EPEC pathogenesis (Elliott *et al.*, 1998). However, advancements in molecular techniques have provided researchers with a better understanding of the genome and virulence determinants among EPEC strains and has led to the sub-classification of EPEC into typical EPEC (tEPEC) and atypical

EPEC (aEPEC) (Trabulsi, Keller and Gomes, 2002). These two classes have been formed on the basis that tEPEC strains possess a large virulence plasmid known as the EAF plasmid, whereas aEPEC do not possess this plasmid (Nataro and Kaper, 1998). Typical EPEC strains adhere to epithelial cells in a distinctive pattern of three-dimensional microcolonies, referred to as a localised adherence (LA) pattern (Scaletsky, Silva and Trabulsi, 1984) and was described by Cravioto *et al*., 1979, who first observed this adherence property in Hep-2 cells, (Cravioto *et al.*, 1979; Baldini *et al.*, 1983). The LA phenotype is dependent on the presence of the EAF plasmid, which encodes a 14 gene operon required for the biogenesis of the bundle forming pilus (Stone *et al.*, 1996). EPEC strains lacking the pEAF have typically been designated aEPEC, which adhere to the surface of cells in either a diffuse-adherence (DA) or aggregative-adherence (AA) pattern (Nataro and Kaper, 1998). A third adherence pattern has been described for some clinical isolates, which have demonstrated a BFP-independent localised-adherence pattern like the LA of tEPEC, called localised-adherent-like (LAL). Later, Scaletsky *et al*., 2005, reported that the presence of a novel adhesin, encoded on a region designated the *locus* for *diffuse adherence* (*lda*), was responsible for the LAL phenotype (Scaletsky *et al.*, 2005).

1.3 Epidemiology of Enteropathogenic *Escherichia coli*

One of the most notable epidemiological features of disease by EPEC, is the astonishing age distribution observed in individuals infected with this pathogen. Indeed, although EPEC is primarily a disease of infants under the age of two years, diarrhoea can be induced in adult volunteers if prescribed at a high inoculum (10^8 to 10^{10}) (Levine *et al.*, 1978). Additionally, in a review by Levine and Edelman, isolation of EPEC was more strongly correlated with the diarrhoeal disease in infants aged less than six months compared to healthy infants, whereas in children over two years old, EPEC was isolated from both

healthy and diarrhoeic individuals (Levine and Edelman, 1984). The explanation for the resistance to EPEC observed in adults is unknown, however, it has been thought that it could perhaps be due to the loss of specific receptors with age or a physiological basis (Nataro and Kaper, 1998). The refined discrimination between distinct EPEC subclasses could also offer an explanation. Typical EPEC is transmitted through the fecal-oral route, often from contaminated surfaces, weaning fluids and human carriers. (Levine and Edelman, 1984). Furthermore, the recently completed Global Enteric Multicentre Study (GEMS) of diarrheal disease in infants, found tEPEC to be more significantly associated with moderate-severe diarrhoea in children under two years in Africa, compared to aEPEC (Kotloff *et al.*, 2013), whereas aEPEC has been found in diarrheic patients of all ages and adults with HIV-AIDS.

The proportion of aEPEC has increased, and the epidemiology of EPEC infection has shifted from predominately tEPEC infections to aEPEC (Hu and Torres, 2015). However, the population structure and virulence profile of this emerging global diarrhoeal pathogen remains poorly understood. Recently, the whole genome sequences of ~200 aEPEC strains collected during the GEMS study revealed that strains could be divided into ten, widely distributed, distinct clonal groups within the *E. coli* core genome phylogeny (Ingle *et al.*, 2016). The phylogenomic analysis performed by Ingle *et al*., 216, also demonstrated that the different aEPEC clones had distinct LEE variants which were integrated at different chromosomal sites, confirming that aEPEC lineages have evolved multiple times via distinct LEE acquisition events (Ingle *et al.*, 2016). Indeed, the genomic framework of aEPEC is complex. Therefore, it is clear that there is a need to investigate the epidemiology and pathogenicity of the atypical EPEC pathotype further.

1.4 Original project aims

Gastrointestinal (GI) disorders are the most common cause of hospital admission in the UK, and they will continue to have major implications on future health care needs as the incidence continues to increase (Williams *et al.*, 2007). Therefore, there is an urgent need to identify cost effective therapeutics to lessen the burden of GI disorders, and one such therapy which has grown in popularity is Faecal Microbiota Transplant (FMT). FMT describes the delivery of stool from a healthy donor, to a recipient with a disease believed to be related to dysbiosis of the gut microbiome. The origins of this therapy can be traced back to ancient Chinese medicine, however, in recent decades it has remerged as a lifesaving intervention for recurrent *Clostridium difficile* infections (rCDI). Furthermore, research has emerged to suggest that FMT can remedy other GI disorders including Inflammatory Bowel Syndrome (IBS) and Inflammatory Bowel Disease (IBD), as well as non-GI associated disorders, including autoimmune and metabolic conditions (Choi and Cho, 2016). In response to the increasing demand for FMT, a growing number of biotech companies are developing biobanks to collect and store healthy stool. Civet Biosciences, who operate in Hong Kong, is one such company that is developing microbiome-based therapeutics for the treatment of GI disorders. They are hoping to become the largest biobank in Asia, however, they contacted our group to propose a collaboration to try and address a contamination issue. Civet Biosciences found that upon screening healthy donors, 80% of the individuals tested positive for EPEC. EPEC as described above, is a pathogenic strain of *E. coli*, which causes severe diarrhoea and has a high mortality rate in children under 5 years old. Thus, the discovery of EPEC in the stool of healthy donors poses a significant problem and renders stool from these individuals unsafe. This set the foundation for this project and the following primary aims were developed:

1. Develop and validate methods to isolate and characterise EPEC from stool specimens
2. Perform whole genome analysis on clinical EPEC strains to identify sequence variations between strains isolated from symptomatic, compared with asymptomatic individuals
3. Identify a novel strategy to facilitate the clearance of contaminating EPEC from healthy stool specimens

The first year of this project was spent optimising isolation methods, developing assays for strain detection and obtaining key isolates. Throughout this period, Civet Biosciences continued to advise that they would send the EPEC strains. However, despite our best efforts to assist with the shipment, we did not receive the clinical EPEC strains and contact with Civet Biosciences ceased. This meant that both my funding and key reagents were not available and thus a new project was devised and performed. Fortunately, we were kindly gifted a small number of EPEC strains by Professor Stephen Baker which were used to best fulfil the aims in the time that was remaining.

1.5 A new threat: the carcinogenic potential of colibactin producing *Escherichia coli*

The intestinal microbiota is crucial to ensure the safeguarding of human health. Indeed, the resident microbial communities provide important benefits to the host, including metabolic activities, development of the host immune system and prevention of colonisation and infection from invading pathogens (O'Hara and Shanahan, 2006; Garrett, Gordon and Glimcher, 2010; Krishnan, Alden and Lee, 2015). Non-pathogenic *E. coli* strains are commonly found as part of the normal human microflora. Commensal strains, such as *E.*

coli Nissle 1917 (Nissle), are particularly beneficial to the host, due to their ability to outcompete and inhibit colonisation of gut pathogens. This strain was named after Professor Alfred Nissle, who isolated the strain in 1917 from the faeces of a soldier who did not suffer from diarrhoea as his comrades did, during the First World War. Since its discovery, Nissle has proven to be an excellent coloniser of the human gut, and has demonstrated the ability to positively affect gut homeostasis and restore balance within the microbiota (Hancock, Dahl and Klemm, 2010). Nissle increases the expression of both antimicrobial and tight junction proteins, thereby improving epithelial barrier function and preventing against pathogen invasion (Ukena *et al.*, 2007; Hering *et al.*, 2014; Sassone-Corsi *et al.*, 2016). The antagonistic activity of Nissle, has led to the therapeutic application of this strain as a probiotic (Mutaflor®) in the treatment of intestinal disorders, such as ulcerative colitis (Kruis *et al.*, 2004).

Recently, comparative genomics of this probiotic strain revealed that Nissle is closely related to the uropathogenic *E. coli* strain CFT073, albeit Nissle lacks genes encoding major virulence factors, including haemolysin and P-fimbriae (Vejborg *et al.*, 2010). Furthermore, both strains have similar genomic profiles and belong to the same phylogroup, B2 group which is comprised of both non-pathogenic and pathogenic *E. coli* strains, (Hancock, Vejborg and Klemm, 2010). Until recently, Nissle was considered as "harmless", and was taken routinely as a probiotic because it was not believed to encode specific virulence determinants. However, it was shown in 2006 that Nissle harbours a 54 kb genomic island, termed the *pks* island that encodes for nonribosomal and polyketide synthases (NRPS and PKS, respectively) required for the biosynthesis of a hybrid peptide-polyketide metabolite called colibactin (Nougayrede *et al.*, 2006) The peptide synthase pathway described by Nougayrède and colleagues was implicated with inflicting a genotoxic insult on eukaryotic cells infected with *pks*$^{+}$ *E. coli,* resulting in DNA damage and chromosomal instability in infected cells (Nougayrede *et al.*, 2006; Cuevas-Ramos *et*

al., 2010; Bossuet-Greif *et al.*, 2018). Subsequently, the *pks* island has been identified in other members of the *Enterobacteriaceae* including *Citrobacter koseri*, *Klebsiella pneumoniea* and *Enterobacter aerogenes* (Putze *et al.*, 2009).

1.6 Colibactin biosynthesis

The *pks* island is comprised of 19 genes (*clbA* to *clbS*) which encode the machinery for the biosynthesis and transportation of the peptide-polyketide hybrid compound, colibactin. The functionality and role of each gene is described in Table 1-1. Oswald *et al.,*2006, performed systematic mutagenesis of the *pks* encoded genes and revealed that all the PKS and NRPS proteins, as well as eight accessory and tailoring enzymes, were essential for colibactin associated genotoxicity. Strikingly, only deletion of *clbM* did not alter genotoxic activity, suggesting that the function could be rescued by other transporters encoded elsewhere in the genome (Nougayrede *et al.*, 2006).

Table 1-1 The function and role of *pks* encoded proteins in the synthesis of colibactin. Adapted from the review by **(Faïs *et al.*, 2018)**.

Protein	**Function**	**Role in colibactin Biosynthesis**
ClbA	Accessory protein	Phosphopantetheinyl transferase activates the enzymatic cascade
ClbR	Transcriptional regulator	Activates production of colibactin
ClbB	Synthesis enzyme	Hybrid NRPS-PKS megasynthase
ClbC	Synthesis enzyme	PKS megasynthase
ClbD	Synthesis enzyme	Hydroxyl acyl coA dehygrodenase

ClbE	Synthesis protein	Acyl transporter protein
ClbF	Synthesis enzyme	α β dehydrogenase
ClbG	Synthesis enzyme	Acyl transferase
ClbH	Synthesis enzyme	NRPS megasynthase
ClbI	Synthesis enzyme	PKS megasynthase
ClbJ	Synthesis enzyme	NRPS megasynthase
ClbK	Synthesis enzyme	Hybrid NRPS-PKS megasynthase
ClbL	Synthesis enzyme	Amidase
ClbM	Accessory protein	MATE transporter facilitates transportation of pre-colibactin across the cytoplasmic membrane
ClbN	Synthesis enzyme	NRPS megasynthase
ClbO	Synthesis enzyme	PKS megasynthase
ClbP	Accessory protein	Transports pre-colibactin from the cytoplasm to the periplasm and cleaves pre-colibactin into mature colibactin
ClbQ	Synthesis enzyme	Thioesterase
ClbS	Accessory protein	Immunity protein inactivates colibactin to protect bacterial cell from autotoxicity

The hybrid NRPS-PKS assembly line is a member of a large family of molecules with interesting biological functions, including the synthesis of antibiotics, cancer therapeutics and immunosuppressives (Fischbach and Walsh, 2006; Faïs *et al.*, 2018). The assembly line is typically organised into mega-complexes, where the synthesised molecule is transferred from one enzymatic complex, onto the next. Colibactin is unique, in that it belongs to a subset of hybrid polyketide-non-ribosomal peptides that become activated via a prodrug mechanism. A structural motif, N-myristoyl-D-asparagine, is added to the N-

terminus of the molecule and later cleaved at the final stage of biosynthesis to produce the mature colibactin molecule (Bian *et al.*, 2013; Brotherton and Balskus, 2013).

The metabolic synthesis starts with ClbA, a phosphopantetheinyl transferase, that activates the NRPS and PKS enzymes. Next, ClbN generates the prodrug motif, N-myristoyl-D-Asparagine, that is accepted by ClbB. ClbB facilitates the addition of either alanine or valine and then incorporates a malonyl-CoA into the intermediate molecule, designated pre-colibactin (Brotherton and Balskus, 2013). The synthesis of pre-colibactin continues along the NRPS-PKS assembly line, using as substrates malonyl-CoA and other amino acids including Glycine, Cysteine and L-Methionine-derived cyclopropane-containing amino acid (Zha *et al.*, 2017; Faïs *et al.*, 2018). The enzymes ClbH and ClbI are essential for the formation of cyclopropane (C_3H_3), which has been shown to be responsible for the alkylation of DNA (Vizcaino and Crawford, 2015). ClbD-G are also important enzymes that are required for the synthesis and installation of an aminomalonyl unit, which is essential for genotoxicity (Nougayrede *et al.*, 2006; Zha *et al.*, 2016; Faïs *et al.*, 2018). Thiazole rings, generated by ClbK (C_3H_3NS), are another structure that have been considered to offer an important role in the biological function of colibactin (Trautman *et al.*, 2017; Faïs *et al.*, 2018). The enzymatic assembly line is predicted to end with ClbO, however, there is evidence to suggest that there are multiple pre-colibactin intermediates (Li *et al.*, 2016; Faïs *et al.*, 2018). The mechanism by which compounds are removed from the pathway remains elusive, however, the enzymatic activity of ClbQ has been implicated with this process, and could therefore have an important role controlling the flow of colibactin production (Guntaka *et al.*, 2017; Faïs *et al.*, 2018). The fully synthesised pre-colibactin is transported to ClbM, a MATE transporter that releases the intermediate into the periplasmic space (Mousa *et al.*, 2016).The peptidase, ClbP, cleaves the prodrug motif from pre-colibactin, generating the mature colibactin ready for exportation out of the bacterial cell (Cougnoux *et al.*, 2012). To protect bacteria against the genotoxic molecule,

pks encoding *E. coli* encode an immunity protein that confers colibactin resistance. The enzyme, ClbS, possesses the ability to hydrolyse the cyclopropane "warheads", enabling the conversion into an innocuous compound (Tripathi *et al.*, 2017)

1.7 Mode of genotoxicity

Despite the discovery of colibactin more than a decade ago, it has remained exceptionally challenging to purify colibactin directly from the culture media. Moreover, the method by which colibactin is transported from the bacterial cell to the host nucleus, remains unknown. Therefore, functional characterisation of this genotoxin has mostly been studied *in vitro,* using cultured human epithelial cells. Effects on the cells include megalocytosis, phosphorylated γ-H2AX foci and G2 cell-cycle arrest (Nougayrede *et al.*, 2006). However, due to the unstable nature of colibactin, cytotoxic activity *in vitro* requires direct contact of cells with live *pks*$^+$ bacteria, as separation of cells and bacteria by an impermeable membrane attenuates the colibactin-associated cytotoxic phenotype observed in cultured cells (Nougayrede *et al.*, 2006). Cultured mammalian cells are transiently infected with *pks*$^+$ *E. coli*, resulting in host DNA double-strand breaks (DSBs) that are detectable up to 4 h after DNA damage has occurred (Nougayrede *et al.*, 2006). In response, the cell recruits the DNA damage response (DDR), and the ataxia telangiectasia mutated kinase (ATM) pathway is activated. The phosphorylation of histone H2AX (γ-H2AX) is one of the first signals in response to DSBs, and detection of this protein can be used to measure the extent of genotoxicity in infected cells (Stiff *et al.*, 2004). Nougayrede *et al., 2006,* also revealed that affected host cells became trapped at the G_2/M transition of the cell cycle, leading to senescence and eventually apoptosis (Nougayrede *et al.*, 2006).

The mode of action on DNA in host cells was first explained by Nougayrede and colleagues, who identified that colibactin bound and generated DNA interstrand cross-links (ICLs) (Bossuet-Greif *et al.*, 2018). Furthermore, following the recent identification of the colibactin structure, two electrophilic "warheads" have been implicated in causing ICLs and forming DNA adducts (Brotherton *et al.*, 2015; Xue *et al.*, 2019). The structure of colibactin contains two nearly symmetrical subunits, each containing a cyclopropane ring, a structural feature common among other DNA alkylating products (Wilson *et al.*, 2019).

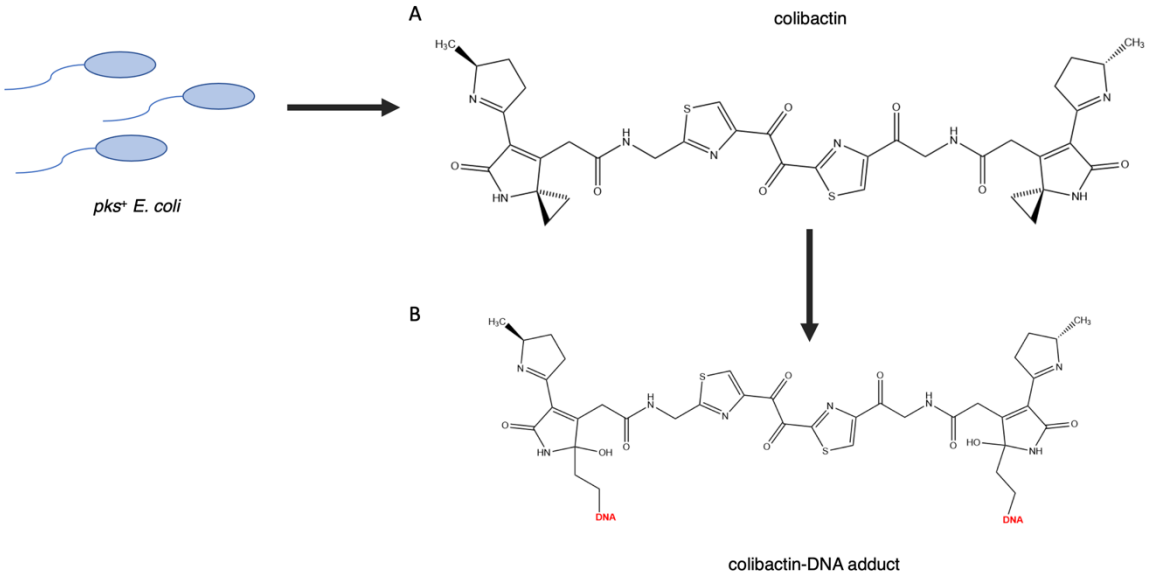


Fig 1-3 The molecular basis for colibactin associated DNA cross-linking. Colibactin produced by *pks*$^{+}$ *E. coli* strains generates DNA interstrand cross-links (ICLs) by binding adenine residues on host DNA. Schematic adapted from **(Wernke *et al.*, 2020)** and colibactin structures provided by Dr Glen Brodie.

The cyclopropane warheads are rendered electrophilic and capable of alkylating DNA after precolibactin is cleaved by ClbP, and the placement of cyclopropane becomes in conjunction with both an imine and amide group (Xue *et al.*, 2019). Further insights into colibactin-DNA interactions have been provided by *in silico* modelling, using structures derived from adductomics which indicate that colibactin has a high affinity for binding

adenine rich motifs (AAAATT) within the DNA minor groove (Dziubańska-Kusibab *et al.*, 2020a). These findings further support the view that colibactin is a potent DNA cross-linking agent.

1.8 Colibactin and colorectal cancer

Concerns regarding the carcinogenic potential of some commensal *E. coli* have recently been sparked, after the *pks* island encoding the genotoxin colibactin was identified in strains belonging to the B2 phylogroup. Colibactin is a potent alkylator of DNA and elicits a cytotoxic response in infected mammalian cells, resulting in DNA damage and gene mutagenesis (Nougayrède *et al.*, 2021). Thus, it is a rational hypothesis to suspect that it may play a role in colorectal cancer (CRC) development.

Colorectal carcinogenesis is a multifactorial process that is influenced by host genetic susceptibility, environmental stimuli and the microbiota (Lopez, Bleich and Arthur, 2021). Dysbiosis of the microbiota is a common observation with CRC patients where there is often an over-representation of certain bacteria including, *Bacteroides/Prevotella*, *Faecalibacterium* and *Fusobacterium* (Sobhani *et al.*, 2011; Kostic *et al.*, 2012; Raisch *et al.*, 2014). Interestingly, studies analysing CRC derived tissues such as colonic adenomas, carcinomas and the mucosa, have revealed that patients are colonised by abnormally higher numbers of adherent *E. coli* (Swidsinski *et al.*, 1998; Martin *et al.*, 2004; Maddocks *et al.*, 2009; Raisch *et al.*, 2014). Furthermore, it has been shown that most mucosa-associated *E. coli* from CRC specimens belong to the B2 phylogroup and harbour the *pks* island (Buc *et al.*, 2013; Raisch *et al.*, 2014). The study by *Riasch et al.,* 2014, also indicated that B2 *pks*$^{+}$ *E. coli* could persist in the tumour microenvironment and promote low levels of inflammation and cell proliferation, pointing to a potential role in the development of

carcinogenesis (Raisch *et al.*, 2014). Indeed, *in vivo* investigations have suggested that colibactin producing *E. coli* significantly promote the development of tumorigenesis in predisposed individuals (Arthur *et al.*, 2012).

Various bacteria derived from the microbiota have been associated with the development of CRC, however, they had not been implicated with having a direct effect on the occurrence of oncogenic mutations. Strikingly, a breakthrough study revealed a distinct mutational fingerprint in tumour biopsies that has linked colibactin-associated genotoxicity to the development of cancer in CRC patients. Intestinal organoids were infected with *pks*$^{+}$ *E. coli* and then analysed by Whole Genome Sequencing (WGS) to identify the resulting mutational signature. Compared to the control, organoids exposed to colibactin producing *E. coli* presented with an increased level of single base substitutions (SBS), with a bias towards T>N substitutions that occurred preferentially at adenine rich residues. In addition, a characteristic small indel signature was characterised by single nucleotide deletions at T-homopolymers. These were like the SBS and occurred in regions that showed enrichment of adenines immediately upstream of the mutation. The mutational signatures were confirmed in CRC patient biopsies and verified that these features were a direct consequence of exposure to *pks* harbouring *E. coli* (Pleguezuelos-Manzano *et al.*, 2020). CRC is one of the most frequently diagnosed malignancies and attributes to over 600,000 deaths per year globally (Raisch *et al.*, 2014). Therefore, colibactin producing *E. coli* present an urgent public health matter that require further investigation. The characterisation of *pks* harbouring *E. coli*, particularly strains that are currently considered as safe such as the commensal Nissle 1917, will be vital for the future of human health.

1.9 The role of L- and D-amino acids

Amino acids are essential building blocks in all forms of life; indeed, all proteins are composed of 20 amino acids. These proteogenic amino acids enable all organisms to synthesise a myriad of products such as enzymes, hormones and antibiotics, each with unique biological functions. All amino acids, except glycine, have a chiral centre and therefore they can exist in two forms called enantiomers, as either the L- or D-enantiomer state (Genchi, 2017). Previously, L-amino acids were considered to dominate throughout nature, however, increasing evidence has demonstrated that D-amino acids are present in high concentrations in many life forms such as humans, mammals, plants and microorganisms (Genchi, 2017).

The extensive role of D-amino acids in bacteria has become an area of intense research. Indeed, D-amino acids have been reported to be utilised as carbon sources to support bacterial growth, incorporated into the bacterial cell wall and to contribute to the regulation of spore germination (Cava et al., 2011). Many bacteria, such as *Vibrio cholerae* can synthesise D-amino acids, by encoding dedicated racemases that convert L-amino acids into their respective D-enantiomers (Lam *et al.*, 2009). Alternatively, bacteria can also utilise host-derived D-amino acids such as D-Serine, which is the most abundant amino acid in the human urine and can be utilised to promote colonisation of certain UPEC strains in the bladder (Anfora *et al.*, 2007). Intriguingly, D-Serine has also been reported to have a role in modulating bacterial gene expression. Recently, a study by Connolly *et al*., 2015, revealed that exposure to D-Serine in EHEC resulted in activation of the SOS response and repression of the T3SS used to adhere to host cells (Connolly *et al.*, 2015). D-Serine has also been reported to affect gene expression in other bacterial species, including *Staphylococcus aureus,* where it was reported that several virulence associated genes

became repressed upon exposure to D-Serine (Iwata *et al.*, 2021). Taken together, these studies highlight the important role of D-amino acids in regulating bacterial gene expression and demonstrate the potential therapeutic application of D-amino acids to modulate virulence in known pathogenic species.

1.10 Revised project aims

Recent work published by the Roe group highlighted that exposure to D-Serine induced distinct transcriptional responses among diverse *E. coli* pathotypes, (Connolly *et al.*, 2021). Analysis of differentially expressed genes (DEGs) revealed that *clb* genes encoding the potent genotoxin colibactin were significantly downregulated in response to D-Serine in the UPEC strain CFT073. Therefore, the following aims were devised to investigate the role of D-Serine on the expression of colibactin:

1. Investigate the effects of L- and D-amino acids on the expression of *clb* genes
2. Examine the effects of D-Serine on the colibactin-associated cytotoxic activity in cultured cells
3. Explore the regulatory mechanism of action of D-Serine induced repression of colibactin

2 Materials and Methods

2.1 Chemicals and growth media

2.1.1 Chemicals and molecular reagents

All chemicals used were purchased from Invitrogen, Thermo Fisher or Sigma Aldrich, unless specified otherwise in the text. GoTaq Green DNA polymerase was purchased from Promega and Multiplex Master Mix, was purchased from Qiagen. 100bp DNA ladder was purchased from Cambridge Reagents. Electrophoretic grade Agarose was purchased from Invitrogen, GelRed (10000X) DNA stain and nuclease free water were purchased from Sigma Aldrich. Oligonucleotides were ordered from Invitrogen. TE Buffer, SDS solution (10%) and Sodium Acetate were all purchased from Ambion. TurboDNase and RNAsae Zap were purchased from Thermo Fisher. LunaScript RT Super Mix and cDNA synthesis kit were purchased from New England Biolabs. Antibodies were purchased from Cell Signalling Technologies.

2.1.2 Growth media and buffers

All growth media and buffers were prepared using either nuclease free water (nfH_2O) or distilled deionised water (ddH_2O) and sterilised by autoclaving or sterilised by filtration (0.2 μM). Heat sensitive agents were filter sterilised and added to media post-autoclaving. For agar to completely dissolve, HiCrome ECC media was heated to 100^oC with stirring. MEM-HEPES and DMEM tissue culture media were purchased from Sigma-Aldrich and stored at 4^oC.

Table 2-1 LB Media (1L; pH7.5)

Component	Quantity
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

*LB components were added to 950 mL ddH_2O and the pH adjusted, before being topped up to 1L and sterilised by autoclaving.

Table 2-2 SOC Media (1L; pH 7.0)

Component	Quantity
Tryptone	20 g

Yeast extract	5 g
NaCl	0.5 g
1M KCl	10 mL
1M $MgSO_4$	2 mL

*SOC components were added to 950 mL ddH_2O and the pH adjusted, before being topped up to 1 L and sterilised by autoclaving. Immediately before use, 5 mL sterile $MgCl_2$ (2M) and 20 mL glucose (1 M) were added to the media.

Table 2-3 M9 minimal media (200 mL)

Component	Quantity
M9 Salts (5X)	40 mL
20% Glucose	5 mL
1M $MgSO_4$	200 µL
1M $CaCl_2$	10 µL

M9 media was made to a final volume of 200 mL with ddH_2O.

Table 2-4 HiCrome ECC Selective Media (500 mL)

Component	Quantity
Agar powder	13.25 g
ddH_2O	500 mL

The mixture was heated to 100°C with stirring, the agar was fully dissolved when the liquid became translucent.

Table 2-5 M63 Media (100 mL)

Component	Quantity
M63 Salts (5X)	20 mL
$MgSO_4$	100 µL
20% Glucose	1 mL
Thiamine	10 µL
Casamino acids	500 µL
ddH_2O	78.39 mL

Table 2-6 Phosphate buffered saline Tween (PBST)

Component	Quantity
10X phosphate buffered	100 mL

saline (PBS)	
Tween-20	0.4 mL

Table 2-7 Tris-acetate-EDTA (TAE)

Component	Quantity
Tris	242 g
Acetic acid (glacial)	57.1 mL
0.5M EDTA (pH 8.0)	100 mL

Table 2-8 Stripping buffer (pH 2.2)

Component	Quantity
Glycine	3.75 g
SDS	0.25 g
Tween-20	2.5 mL

Table 2-9 Coomassie blue

Component	Quantity
Methanol	500 mL
Coomassie blue	2 g
Acetic Acid	100 mL
ddH_20	400 mL

Table 2-10 Coomassie blue de-stain

Component	Quantity
Methanol	100 mL
Acetic Acid	100 mL
ddH_20	800 mL

Table 2-11 4X sample buffer (100 mL)

Component	Quantity
Tris-HCl (pH 6.8)	1 M
SDS	8.52 g
Bromophenol blue	0.43 g
Glycerol	42.6 mL
β-mercaptoethanol	4.97 mL

Table 2-12 4% PFA (500 mL)

Component	Quantity
Paraformaldehyde (PFA)	20 g
10X PBS	50 mL
1M NaOH	5 mL
ddH_2O	445 mL

PFA was dissolved in PBS and 250 mL ddH_2O with heating; the temperature was monitored to ensure it did not exceed 50°C. After the PFA had fully dissolved, the NaOH was added, and the volume topped up to 500 mL with ddH_2O. The liquid was cooled before being filter sterilised (0.2 µM) and aliquoted into sterile 50 mL centrifuge tubes.

2.1.3 Growth media supplements

Antibiotic solutions were prepared in ddH_2O or ethanol to the desired concentration (Table 2-13). Antibiotics prepared in water were filter sterilised (0.2 µM). Stocks were aliquoted and stored at -20°C.

Table 2-13 Antibiotic stock concentrations

Antibiotic	Stock (mg/mL)	Solvent	Final Concentration (µg/mL
Chloramphenicol	25	Ethanol	25
Ampicillin	100	ddH_2O	100
Gentamycin	10	ddH_2O	10
Kanamycin	50	ddH_2O	50

2.2 Bacterial strains and plasmids

For the purpose of this project, bacterial strains were gifted from other researcher institutes or sourced from publicly available culture collections. Clinical isolates were obtained during a placement in Vietnam, the strains were isolated from paediatric patients admitted to the hospital with gastroenteritis during the years 2012 and 2014.

Table 2-14 List of bacterial strains used in these studies

Strain	Description	Source
TUV 93-O	Wild type *E. coli* O157:H7 (Shiga-toxin negative)	Prof John Leong
E110019	Wild type EPEC atypical	Prof Brendan Kenny

E2348/69	Wild type EPEC typical 0127:H6	Prof David Smith
E2348/69 Δ*tir*	Wild type EPEC typical 0127:H6 *tir* deletion mutant	Roe Laboratory Infantry
ICC168	Wild type *C. rodentium*	Prof Bret Findlay
239-2	Wild type ETEC	Scottish Microbiology Reference Laboratories, Edinburgh
NCTC1978	Wild type *S. sonnei*	National Collection of Type Cultures (NCTC), England
01-0132	Clinical Isolate EPEC atypical	Prof Stephen Baker, Vietnam
01-0145	Clinical Isolate EPEC atypical	
01-0259	Clinical Isolate EPEC atypical	
01-0271	Clinical Isolate EPEC atypical	
01-0898	Clinical Isolate EPEC atypical	
01-0017	Clinical Isolate EPEC atypical	
01-0108	Clinical Isolate EPEC atypical	
01-0223	Clinical Isolate EPEC atypical	
01-0112	Clinical Isolate EPEC atypical	
BCB21827	Clinical Isolate EPEC atypical	
BCB21787	Clinical Isolate EPEC atypical	
BCB22675	Clinical Isolate EPEC atypical	
Nissle 1917	Commensal *E. coli* O6:K5:H1	Dr Donal Wall
Nissle 1917 Δ*dsdC*	Nissle *dsdC* deletion mutant	This study
CFT073	Wild type UPEC O6:H1:K2	Roe lab inventory
CFT073 Δ*dsdC*	UPEC *dsdC* deletion mutant	Roe lab inventory

Table 2-15 Bacterial plasmids used in these studies

Plasmid	Description	Source
prpsM:GFP	rpsM promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al*., 2003
prpsM:RFP	rpsM promoter translational fusion of pAJR70 to eRFP (Kan^R)	Roe *et al*., 2003
pclbB:GFP	*clbB* promoter translation fusion of pAJR70 to eGFP (CHl^R)	Roe lab inventory
pKD3	Template plasmid for Lambda Red mutagenesis (Cm^R)	Datsenko and Wanner., 2000
pKD46	Lambda Red recombinase expressing plasmid (Amp^R)	Datsenoke and Wanner., 2000
pACYC184	Multicopy number plasmid (Cm^R)	Roe lab inventory

2.2.1 Bacterial growth and calculation of colony forming units (CFU)

Overnight bacterial cultures were prepared by inoculating a single colony and grown in 5 mL of Luria-Bertani broth (LB) for approximately 16 hours at 37°C, 200 RPM (New Brunswick Scientific controlled environment shaker). The following day, cultures were diluted, at a ratio of 1:100, into fresh media and grown either for a specified number of hours or until they reached a desired $OD_{600\ nm}$.

To confirm appropriate cell density, the CFU was determined from an aliquot of bacterial suspension and diluted serially in PBS. For each sample, 20 µL of each dilution was spotted on to solid agar plates, the plates were allowed to dry and were incubated overnight at 28°C. Single colonies were counted within each spot and multiplied by the relevant dilution factor to calculate the number of CFU. Serial dilutions were performed in triplicate and the CFU was determined as the average number of colonies at a given dilution.

2.2.2 Storage of bacterial cultures

Bacterial stocks were prepared by isolating a single colony from an LB plate and growing overnight (16 hours) at 37°C, 200 RPM (New Brunswick Scientific controlled environment shaker). From this culture, 500 µL was mixed with 1 mL sterile glycerol (40%) and peptone (2%) in a sterile tube and then frozen at -80°C.

2.3 General molecular techniques

2.3.1 Polymerase Chain Reaction (PCR) and primers

The primers for this study were designed using Primer3web. Primers were designed as standard, with a length of around 20 bp, approximately 50% in GC content and a melting temperature between 55-65°C. Primers designed for cloning were designed to have the relevant restriction site flanks necessary for the ligation reactions. Primers were procured from Life Technologies and lyophilised primer samples were reconstituted to a concentration of 100 µM using nfH_2O. From these, working stocks of 10 µM were prepared with nfH_2O and used in standard PCR reactions. Both stocks were stored at -20°C.

Table 2-16 Oligonucleotide sequences used in this study

Primer Name	Description	Sequences (5' – 3')
eae_F	Forward for the intimin gene (*eae)*	TCAATGCAGTTCCGTTATCAGTT
eae_R	Reverse for the intimin gene (*eae*)	GTAAAGTCCGTTACCCCAACCTG
bfp_F	Forward for the bundle forming pilus gene (*bfp*)	GGAAGTCAAATTCATGGGGGTAT
bfp_R	Reverse for the bundle forming pilus gene (*bfp*)	GGAATCAGACGCAGACTGGTAG T
stx_F	Forward for the Shiga toxin genes (*stx1* & *stx2*)	GAGCGAAATAATTTATATGTG
stx_R	Reverse for the Shiga toxin genes (*stx1* & *stx2*)	TGATGATGGCAATTCAGTAT

elt_F	Forward for the heat labile toxin (*elt*)	ACGGCGTTACTATCCTCTC
elt_R	Reverse for the heat labile toxin (*elt*)	TGGTCTCGGTCAGATATGTG
ipaH_F	Forward for the invasion plasmid antigen H gene (*ipaH*)	GTTCCTTGACCGCCTTTCCGATACCGTC
ipaH_R	Reverse for the invasion plasmid antigen H gene (*ipaH*)	GCCGGTCAGCCACCCTCTGAGAGTAC
clbA_F	Forward for *clbA* gene expression using RT-qPCR	TAGCCTCTCCACTCATCACC
clbA_R	Reverse for *clbA* gene expression using RT-qPCR	TCAATTCTGCCCATTTGACGA
clbB_F	Forward for *clbB* gene expression using RT-qPCR	TTGTCTCCGGATGTGTGTCA
clbB_R	Reverse for *clbB* gene expression using RT-qPCR	CACATCGTCAGCATAGCACC
clbR_F	Forward for *clbR* gene expression using RT-qPCR	ACCCGTTATCTCTGCGTGAA
clbR_R	Reverse for *clbR* gene expression using RT-qPCR	TCTCATTCCTGTTAGCAATGTGT
gapA_F	Forward for housekeeping control (*gapA)* gene using RT-qPCR	TTTCCGTGCTGCTCAGAAAC
gapA_R	Reverse for housekeeping control (*gapA)* gene using RT-qPCR	GGCCGTGAGTGGAGTCATAT
mcmA_F	Forward for *mcmA* gene expression using RT-qPCR	ACGGTTAGTCCCTCCATGTC
mcmA_R	Reverse for *mcmA* gene expression using RT-qPCR	TGATGTACCTCTGCCGGATC
mchB_F	Forward for *mchB* gene expression using RT-qPCR	GCTCTCGCCGGAATACCT
mchB_R	Reverse for *mchB* gene expression using RT-qPCR	GCACTACCACTTCCCACGG
VNTR_F	Forward for sequencing the VNTR region in Nissle 1917	TGAATTGATACAGCCACCACA

VNTR_R	Reverse for sequencing the VNTR region in Nissle 1917	TCACGCAGAGATAACGGGTT
yahA_Red_F	Forward for *yahA* lambda red mutagenesis in CFT073	GGAATCCACAGGTAATCACTCA TACTGAACAGCGATAAAAGATA AAAGGTGTGTAGGCTGGAGCTG CTTC
yahA_Red_R	Reverse for *yahA* lambda red mutagenesis in CFT073	GAGAATTAATCGCTGCCAACCT CCCGATGCGCTGTGAATAGCGG TTTACACATATGAATATCCTCCT TAG
yahA_184_F	Check forward for *yahA* lambda red mutagenesis in CFT073	TGAAGTCAGCCCCATACGATTC GATTGATGATTGCCAGGC
yahA_184_R	Check reverse for *yahA* lambda red mutagenesis in CFT073	CAATCCATGCCAACCCGTTCTTT TGCCGGATTACATGGTCC
yahB_Red_F	Forward for *yahB* lambda red mutagenesis in CFT073	TACAGTCATGGCTGGAGAAATT CAGAATCGAAAATGGCGGGAGC GGGTCAGTGTAGGCTGGAGCTG CTTC
yahB_Red_R	Reverse for *yahA* lambda red mutagenesis in CFT073	GGTGGTTGATGTAAACCGCTATT CACAGCGCATCGGGAGGTTGGC AGCGACATATGAATATCCTCCTT AG
yahB_184_F	Check forward for *yahB* lambda red mutagenesis in CFT073	TGAAGTCAGCCCCATACGATTA CAGTCATGGCTGGAGAAATTCA GAATCGAAAATGGCGGGAGCGG GTCA
yahB_184_R	Check reverse for *yahB* lambda	CAATCCATGCCAACCCGTTCGGT

	red mutagenesis in CFT073	GGTTGATGTAAACCGCTATTCACAGCGCATCGGGAGGTTGGCAGCGA
ttk_Red_F	Forward for *ttk* lambda red mutagenesis in CFT073	AATAACGTCATAACATAGCCGCAAACATTTCGTTTGCGGTCATAGCGTGGGTGTAGGCTGGAGCTGCTTC
ttk_Red_R	Reverse for *ttk* lambda red mutagenesis in CFT073	AGTTTGGCGTTTAACGAAACTCGCCGGATGAAAAGTCATCCGGCGTCATCCATATGAATATCCTCCTTAG
ttk_184_F	Check forward for *ttk* lambda red mutagenesis in CFT073	TGAAGTCAGCCCCATACGATAATAACGTCATAACATAGCCGCAAACATTTCGTTTGCGGTCATAGCGTGG
ttk_184_R	Check reverse for *ttk* lambda red mutagenesis in CFT073	CAATCCATGCCAACCCGTTCGATGACGCCGGATGACTTTTCATCCGGCGAGTTTCGTTAAACGCCAAACT
mcmA_Red_F	Forward for *mchA* lambda red mutagenesis in Nissle 1917	TTAAACGTGAATTTTTACTTCTTCACAAATCTTATAGCGAAGGTGTTGAAGTGTAGGCTGGAGCTGCTTC
mcmA_Red_R	Reverse for *mchB* lambda red mutagenesis in Nissle 1917	TCGTCTGCGGGGAGTGGAAGTTAAACCTTATATTGTTAATGAAGCACCTGCATATGAATATCCTCCTT

		AG
mcmA_184_F	Check forward for *mcmA* lambda red mutagenesis in Nissle 1917	TGAAGTCAGCCCCATACGATAC GACCGGCTATTTTCAGGA
mcmA_184_R	Check reverse for *mcmA* lambda red mutagenesis in Nissle 1917	CAATCCATGCCAACCCGTTCAA ACGTTCGGAGGAGCCTAA
mchB_Red_F	Forward for *mchB* lambda red mutagenesis in Nissle 1917	TCTGGTCTGGTCAGGCTGGAAA AACGGAAGTTAAATATGATGGA GTTTATGTGTAGGCTGGAGCTGC TTC
mchB_Red_R	Reverse for *mchB* lambda red mutagenesis in Nissle 1917	TTATATTTTTATTTATTTTACAGG TACTTTTTCGGTAGTACCTGTAA AATCATATGAATATCCTCCTTAG
mchB_184_F	Check forward for *mchB* lambda red mutagenesis in Nissle 1917	TGAAGTCAGCCCCATACGATTCC CGATGTATGACCGTCAG
mchB_184_R	Check reverse for *mchB* lambda red mutagenesis in Nissle 1917	CAATCCATGCCAACCCGTTCTGA GGACTTGATCTGCCTGG

2.3.2 PCR reaction setup and conditions

Polymerase Chain Reaction (PCR) was set up and performed in the standard way, unless otherwise stated in the text. PCR reactions were generally set up in 20 µL and had a final dNTP concentration of 0.2 mM. If using a bacterial colony as template, a colony was isolated from a fresh plate using a pipette tip and mixed into 50 µL nfH_2O before adding 1 µL of the water to the reaction mix. The reaction mixture for end point PCR was prepared with 10 µL GoTaq Green Master Mix (10X), 1 µL primer at 10 µM concentration, 1 µL template DNA and 8 µL nfH_2O. Amplification was performed using a Techne TCV512 thermocycler. The following cycling conditions were set, initial heat start denaturation at 95ºC for 2 minutes; denaturation at 95ºC for 30 seconds; annealing at 55ºC for 30 seconds and elongation at 72ºC for 80 seconds for a total of 30 cycles. A final extension was performed at 72ºC for 10 minutes.

2.3.3 Multiplex PCR amplification

Multiplex PCR reactions were prepared with 25 µL Qiagen Multiplex Master Mix, 5 µL primer mix at 50 µM concentration, 1 µL bacterial DNA and 19 µL nfH_2O. Amplification was performed using a Techne TCV512 thermocycler. The following cycling conditions were set; initial heat activation at 95^oC for 15 minutes, denaturation at 94^oC for 30 seconds, annealing at 57^oC for 90 seconds and elongation at 72^oC for 60 seconds, for a total of 30 cycles. PCR products were run on 2% agarose gel at 100 volts for 80 minutes.

2.3.4 Agarose gel electrophoresis

Frequently, electrophoresis of agarose gels was employed to separate PCR products. Typically, a 2% agarose gel was prepared with electrophoretic agarose and TAE buffer. GelRed Nucleic acid gel stain (Cambridge Bioscience) was incorporated into molten agarose at a ratio of 1:10 per volume of buffer. Agarose gels were set into electrophoresis tanks filled with TAE buffer. Electrophoresis was performed at 100 volts for approximately 80 minutes, or until the dye front had progressed to the bottom of the gel. The DNA was visualised using a UV transilluminator.

2.3.5 Plasmid purification

An overnight culture containing the desired plasmid was centrifuged at 13,000 g for 5 minutes. The QIAprep Spin Miniprep kit (QIAGEN) was then used to the manufactures specifications to purify the plasmid DNA. The plasmid DNA was eluted in 30 µL of nfH_2O and concentrations were measured using the NanoDrop DS-11+ Spectrophotometer (DeNovix).

2.3.6 Transforming plasmid vectors

A single colony or a 1:100 dilution from an overnight culture was inoculated into a culture flask with 10 mL LB media and incubated for 1 hour and 30 minutes or until the $OD_{600\ nm}$ was in a range of 0.3-0.5. To a centrifuge tube, 1 mL of inoculum was transferred and centrifuged at 10,000 g for 5 minutes at 4^oC. The supernatant was discarded, and pellets

were washed with 1 mL ice cold ddH_2O then centrifuged at 10,000 g for 5 minutes at 4°C. The washing process was repeated twice more, to ensure removal of excess salt. After the third wash, the supernatant was discarded and the pellet resuspended in 100 µL ddH_2O. 50 µL of cells was mixed with 1-2 µL of plasmid DNA and dispensed into a pre-chilled electroporation cuvette. Electroporation was performed using an electroporator (Eppendorf Eporator) at 2,500 volts, then 450 µL SOC recovery media was immediately added to cuvette. The transformation mix was transferred to an Eppendorf tube and incubated with shaking at 37°C for 1-2 hours (New Brunswick Scientific controlled environment shaker). 100 µL of transformation mix was spread on an LB plate with appropriate antibiotic selection and incubated at 37°C.

2.3.7 Phenol-chloroform extraction and ethanol precipitation

Prior to phenol-chloroform extraction, samples were made up to a total volume of 400 µL with nfH_20 and then mixed with an equal volume of phenol:chloroform:isoamillic alcohol 25:24:1 (PCIA, Sigma Aldrich). Samples were then centrifuged at maximum speed for 5 minutes, then the top layer was removed and dispensed into a fresh Eppendorf tube. To the top layer, 400 µL chloroform:isolamillic alcohol 24:1 (CIA, Sigma Aldrich) was added and then the tubes were inverted several times before centrifuging at maximum speed for a further 1 minute. After centrifugation, the top layer was removed and dispensed into a new Eppendorf tube containing 40 µL sodium acetate, 1 µL Glycoblue coprecipitate (ThermoFisher Scientific) and 800 µL 100% ethanol. The samples were inverted several times and stored at -20°C overnight or -80°C for 1 hour. The samples were removed from the freezer and centrifuged at maximum speed for 20 minutes at 4°C. The supernatant was carefully removed to avoid disrupting the pellet, then 1 mL of 70% ethanol was added before the samples were centrifuged for a further 5 minutes at maximum speed. After centrifugation, the supernatant was carefully removed, and the pellets were left to airdry.

Once dry, the pellets were resuspended in 100 µL nfH_2O, or TE buffer and concentrations measured using the NanoDrop DS-11+ Spectrophotometer (DeNovix)

2.3.8 Lambda red genetic recombination

This method was performed as described previously (Datsenko and Wanner, 2000). First a linear deletion fragment was prepared by amplifying the chloramphenicol resistance cassette from pKD3 with oligonucleotides bearing 50 bp 5'-end flanking regions bearing homology to the 50 bp regions immediately upstream and downstream of the gene to be deleted. Four 50 µL PCR reactions were performed and checked by gel electrophoresis, before the PCR products were pooled into a single Eppendorf tube. Phenol-chloroform extraction and ethanol precipitation (2.3.7) was then performed on the PCR product before resuspending in a final volume of 10 µL nfH_2O.

Strains were transformed with pKD46 prior to recombination. A single transformed colony was cultured in LB supplemented with 100 µg/mL ampicillin and 100 mM L-Arabinose at 30^oC and 200 RPM until an $OD_{600\ nm}$ of 0.4 was reached (New Brunswick Scientific controlled environment shaker). Cultures were centrifuged at 3,750 g for 5 minutes and the supernatant was removed, then, the cells were washed and resuspended three times with ice-cold ddH_2O. The pellets were resuspended in a final volume of 50 µL of ddH_2O and mixed with 1 µg of PCR product before being added to pre-chilled electroporation cuvettes and electroporated at 2500 volts (Eppendorf Eporator). 500 µL of SOC was immediately added to the cuvettes and cells were recovered at 37^oC, 200 RPM (New Brunswick Scientific controlled environment shaker). 200 µL of recovered cells was plated onto antibiotic selection plates and incubated at 37^oC. Insertional mutants were verified by PCR using check primers detailed in Table 2-16. Resistance cassettes were removed by spreading successful recombinants onto LB plates and incubating at 42^oC, then excision of the resistance cassette was confirmed by PCR.

2.4 Genomic and transcriptomic analysis

2.4.1 RNA extraction and DNase treatment

For screening of amino acids capable of modulating expression of *clbB*, bacteria were cultured in M9 minimal media supplemented with amino acids at 1 mM final concentration for 5 h at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker). The cells were harvested by centrifugation before resuspension in two volumes of RNAprotect Bacteria Reagent (Qiagen). After a 10-minute incubation at room temperature, cells were centrifuged for a further 10 minutes at maximum speed and the supernatant was discarded. The pellet was resuspended in 100 µL of TE buffer with lysozyme (10 mg/mL), then 0.5 µL of 10% SDS solution was added before vortexing to resuspend the pellet. Samples were incubated for 5 minutes at room temperature, then 350 µL of lysis buffer with 1% β-mercaptoethanol was dispensed into each tube and the samples were vortexed to mix. 250 µL of 100% ethanol was added to the samples and then they were vortexed to remove any visible precipitation.

The samples were transferred to a spin column and RNA extractions were carried out using PureLink RNA Mini Kit (Thermofisher Scientific) according to manufacturer's instructions. RNA was eluted in a final volume of 100 µL of nfH_2O and contaminating DNA was then removed from the samples using TurboDNase (ThermoFisher Scientific) treatment. 4 µL of TURBO DNase and 10 µL of 10X TURBO Dnase buffer was added to each sample, then, the samples were centrifuged at maximum speed for 15 seconds before incubating for 60 minutes at 37ºC. During the incubation period, the samples were vortexed and centrifuged at 15-minute intervals. Phenol-chloroform extraction and ethanol precipitation was then performed on the RNA samples, as described in 2.3.7, before final resuspension in 100 µL nfH_2O. The concentration was measured using the NanoDrop DS-

11+ Spectrophotometer (DeNovix) and then samples were normalised to a final concentration of 10 ng/ µL in 100 µL nfH_2O.

2.4.2 Quantitative real time PCR (RT-qPCR)

Ten nanograms of DNA-free total RNA, extracted as described above (2.4.1), was used as a template to prepare 10 µL cDNA using LunaScript®RT SuperMix Kit (New England Biolabs) according to manufacturer's instructions. Briefly, 2 µL of 5X LunaScript RT SuperMix was mixed with 7 µL of nfH_2O and 1 µL of RNA template. The samples were subjected to thermocycling, using the Eppendorf Mastercyler Nexus gradient, under the following conditions (Table 2-17).

Table 2-17 Thermocycling conditions for LunaScript®RT SuperMix cDNA synthesis

Condition	Temperature (°C)	Time (minutes)
Annealing	25	2
cDNA synthesis	55	10
Heat inactivation	95	1

Luna® Universal qPCR Master mix (New England Biolabs) was employed for RT-qPCR according to manufacturer's recommendations. Technical duplicate 20 µL reactions were carried out with 1 µL volumes of cDNA as template and triplicate biological samples of cDNA being analysed. RT-qPCR was performed on the CFX-Connect Real-Time PCR detection system (BIO-RAD) using the following conditions, initial denaturation 95°C, 1 minute; denaturation 95°C, 15 seconds; extension 60°C, 30 seconds; for 39 cycles. Prior to RT-qPCR, the primer efficiency of each primer pair was assessed using the thermocycling conditions described above. A series of concentration standards was prepared using template cDNA (100, 4, 0.8 and 0.16 ng/ µL) and only primers which had an efficiency of

90-110% were selected for use in experiments. Analysis was performed using the CFX-Connect software (BIO-RAD) and expression relative to the untreated control was calculated using $2^{-\Delta\Delta ct}$ method with *gapA* amplification employed as a housekeeping control.

2.4.3 Sanger sequencing

For the sequencing of nucleotide stretches less than 1 kb, primers were designed around the sequence of interest. The sequence was amplified using PCR and purified using a QIAquick PCR purification kit (QIAGEN), then the purified PCR product and primers were sent to Eurofins Genomics for Sanger sequencing. The returned FASTA files were uploaded to Jalview (2.11.1.7 release) and multiple sequence alignment was performed using Clustal Omega.

2.4.4 Whole genome sequencing (WGS)

Samples designated for WGS were prepared in accordance with the instructions provided by MicrobesNG. A single colony of the strain to be sequenced was isolated from a fresh agar plate and mixed in 100 μL in sterile PBS. From the 100 μL culture, the strain was streaked out onto a fresh LB plate, with around 1/3 of the plate a lawn of bacteria and then streaked out to ensure the culture was pure. The plate was incubated overnight at 37°C, until there was an abundant growth observed on the plate. Using a sterile loop all the bacterial growth, including the lawn of bacteria, was taken off the plate and mixed into the barcoded bead tube supplied by MicrobesNG. The tube was inverted 10 times and then the samples were sent at room temperature to MicrobesNG (Birmingham, United Kingdom) for DNA extraction and Illumina sequencing.

2.4.5 Transcriptomic analysis by RNA-Seq

Prior to this PhD commencing, an RNA Sequencing (RNA-Seq) experiment was performed by Dr Nicky O'Boyle and the raw data was made available on the European Nucleotide Archive (ENA) under the accession numbers: ERS4281326-ERS4281334 and ERS4281354-ERS4281356.

Briefly, CFT073 wild type and the corresponding isogenic mutant, Δ*dsdC,* strains were grown in triplicate, in M9 minimal media in the presence or absence of 1 mM D-Serine for 5 hours at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker). The $OD_{600\ nm}$ of the samples was adjusted to 1.0 and 1 mL of culture was harvested by centrifugation before RNA was extracted in accordance with the method described above (2.4.1). RNA samples were quality assessed before they were sent to the University of Glasgow Polyomics facility where the library preparation and sequencing was performed. Sequencing was performed on the Illumina NextSeq 500 platform, with at least 10 million 75 bp single end reads being obtained. The reads were quality assessed (minimum Phred threshold of 20) with FastQC (Babraham Bioinformatics) before they were imported into CLC Genomics Workbench (Qiagen) and mapped to the CFT073 reference genome (NCBI accession number: NC_004431.1) using CLC mapping parameters. Differential expression was assessed using the empirical analysis of differential expression (EdgeR), with genes that displayed an absolute fold change of ≥ 1.5; ≤ -1.5 and a false discovery rate corrected *p*-value of ≤ 0.05 considered. Pairwise comparison between WT and $WT^{D\text{-}Ser}$ was conducted to identify any changes to gene expression induced by the addition of D-Serine. A second pairwise comparison between Δ*dsdC* and $\Delta dsdC^{D\text{-}Ser}$ was conducted to identify whether changes to gene expression in response to D-Serine was caused by the transcription factor DsdC.

2.4.6 ChIP-Seq analysis

Chromatin immunoprecipitation coupled with next generation sequencing (ChIP-Seq) was performed by Dr Natasha Turner and is described elsewhere (Turner, 2021). Briefly, CFT073DsdC-FLAG was grown, in duplicate, in M9 minimal media for 5 hours at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker). Further, CFT073$^{DsdC\text{-}FLAG}$ was also grown, in duplicate, in M9 minimal media for 3 hours at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker), before the media was spiked with 1 mM D-serine and grown for another 2 hours. Non-tagged CFT073 WT samples were also grown in parallel, to be used as a negative control in the ChIP experiment. The ChIP experiment was performed in accordance with methods previously established (Bonocora *et al.*, 2015). After generating ChIP DNA, the samples were sent to the University of Glasgow Polyomics facility, where a ChIP-Seq library was prepared before the samples were sequenced using an Illumina NextSeq 500 platform (75 bp length; single end). The reads were quality assessed (minimum Phred threshold of 20) with FastQC (Babraham Bioinformatics) before they were imported into CLC Genomics Workbench (Qiagen). Raw fastq files were aligned to the CFT073 reference genome (NCBI accession number: NC_004431.1). Sequence data was analysed using the ChIP-Seq analysis tool software. To establish enrichment, the DsdC^{FLAG}-tagged samples were aligned against the mock CHIP WT control samples and the maximum *p*-value for calling enriched peaks was set to ≤ 0.05. Peaks were called from two biological replicates and manually checked to ensure they were the correct bimodal peak shape.

2.5 Biochemical assays

2.5.1 SDS-PAGE

Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) was set up and performed using pre-cast NuPage 4-12% Bis-Tris protein gels (Invitrogen). Proteins were mixed with 1X Sample Buffer and denatured by heating samples to 95ºC for 5 minutes. The volume of sample was predetermined by normalising proteins by western blot using an anti-ß Tubulin polyclonal antibody (Abcam). Band intensity was measured using Image Studio Lite Ver.2.5 software (Li-Cor). The ratio of control band to sample band intensity was calculated, and the loading volume per sample was determined. The ladder, SeeBlue Plus2 protein standard (Invitrogen), and protein samples were loaded into a polyacrylamide gel and placed in NuPage MES (Invitrogen) running buffer. Thereafter, samples were separated by supplying a voltage of 160 volts for 30 minutes. Following SDS-PAGE, proteins were transferred onto a nitrocellulose membrane, 0.45 µM (ThermoFisher) using the standard electroblotting procedure. Electrophoresis was performed in NuPage transfer buffer (Invitrogen), with a constant voltage of 30 volts for 1 hour. Alternatively, gels were stained for 1 hour in Coomassie blue stain (Table 2-9) and de-stained for a further 1 hour with Coomassie blue de-stain (Table 2-10) or overnight in ddH_2O. Coomassie stained SDS-PAGE gels were then images on a transilluminator (BIO-RAD ChemiDoc MP Imaging System).

2.5.2 TCA extraction of T3SS associated proteins

Extraction of secreted proteins from EPEC strains was carried out using a method previously described (Tree *et al.*, 2011). Overnight cultures were inoculated at a ratio of 1:100 into 25 mL of Dulbecco's Minimal Eagles Medium (DMEM) (Thermofisher) and grown at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker) until

an $OD_{600\ nm}$ of ~0.8 was reached. Cultures were transferred to 50 mL centrifuge tubes and harvested at 3,750 g for 10 minutes, the supernatant only was removed. The supernatants were filtered through a 0.2 μM filter (Fisher) into new 50 mL centrifuge tubes before ice-cold Trichloroacetic acid (TCA) was added to a final concentration of 10% (v/v); 5 mL TCA to 45 mL supernatant. 1 μL of lysozyme (2 mg/mL) was added to each sample, as a co-precipitant for maximum protein recovery, then samples were briefly vortexed and stored at 4°C overnight. After overnight precipitation, the secreted proteins were harvested by centrifuging the samples at 4°C for 45 minutes, 3,750 g. The supernatant was carefully removed, and the tubes were turned upside down and left to air dry for 10 minutes before the inside of the tubes were blotted dry with tissue paper, taking care not to disrupt the pellet. The pellet was then allowed to airdry for a further 10 minutes, before being resuspended in 25 μL of 1X sample buffer (Table 2-11) and transferred to an Eppendorf tube. The samples were boiled at 97°C for 10 minutes and analysed by SDS-PAGE as described above (2.5.1).

2.5.3 Western blot

Proteins for analysis by Western blot were transferred from an SDS-PAGE gel to a 0.45μM nitrocellulose membrane (ThermoFisher) using an Xcell SureLock® Transfer system. Electroblotting was performed in NuPage transfer buffer (Invitrogen), with 30 volts applied for 1 hour. Nitrocellulose membranes were blocked with 5% milk made with PBST for 1 hour. Primary antibody was prepared to the specified concentration (Table 2-18) with 5% BSA in PBST and incubated on the membrane for 1 hour at room temperature on an orbital shaker. Membranes were washed three times with 50 mL PBST for 10 minutes each wash, then the membrane was probed with horse-radish peroxidase (HRP) conjugated secondary antibody for 1 hour. Membranes were washed a further four times with 50mL PBST for 10 minutes each, then membranes were developed using either,

Pierce ECL Plus Substrate or SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFisher) for 5 minutes. Membranes were exposed for 12 minutes using the C-Digit by Li-Cor.

Table 2-18 Antibody concentrations used for Western Blotting and Immunofluorescence

Antibody	Source	Primary Concentration	Secondary Concentration
Phospho-H2AX (Cell Signalling technologies)	Rabbit	1:100	1:2500 or 1:400 for Immunofluorescence
H2A.X (Cell Signalling Technologies)	Rabbit	1:100	1:2500
ß-Tubulin (Abcam)	Rabbit	1:5000	1:2500
Phospho-H2AX (BD Biosciences)	Mouse	1:200	1:1000 for FACs

2.5.4 Western blot analysis of H2AX phosphorylation

Post-infection (2.6.5), HeLa cell tissues were washed and lysed directly in the cell culture well by applying 100 µL of 1X SDS sample buffer (Table 2-11) and incubating for 5 minutes. Then wells were scraped to release attached cells and cell lysates were immediately stored on ice. The cell lysates were heated for 10 minutes at 90°C, and aliquots were stored at −20°C. Proteins were separated on 4-12% Bis Tris Gel by SDS-PAGE (Invitrogen) as described in 2.5.1 and transferred to a nitrocellulose membrane (FisherScientific). Blocking was performed using 5% skimmed milk powder for 1 hour in 1X Phosphate-Buffered Saline, 0.1% Tween®20 (PBST). The membrane was then incubated with anti-γ H2AX primary antibody (Cell Signalling Technologies) diluted 1:100 in 5% BSA-PBST and incubated overnight at 4ºC. The membrane was washed three times with PBST for 10 minutes before being incubated for 1 hour with anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (Invitrogen) diluted 1:2500

in PBST. The membrane was again washed three times with PBST for 10 minutes. Bound secondary HRP-labelled antibodies were revealed with SuperSignal™ West Femto maximum sensitivity substrate (ThermoFisher Scientific) and analysed with the C-DiGit® blot scanner (LI-COR). Membranes were stripped with mild stripping buffer and incubated with the primary antibody H2AX (Cell Signalling Technologies) and detected with the secondary antibody as described above. To control for sample loading, membranes were probed for β Tubulin (Abcam). Proteins were quantified with Image Studio Lite (Licor) and normalized in relation to the β Tubulin level.

2.5.5 Immunofluorescence analysis of H2AX phosphorylation

HeLa cells were seeded on 13 mm coverslips and infected as described in section 2.6.5. Following 4 hours infection, cells were washed twice with PBS and fixed in 4% (*w/v*) paraformaldehyde for 20 minutes at room temperature. The cells were permeabilised with 0.1% Triton X-100 and then blocked with PBST + 10% normal goat serum for 1 hour at room temperature. Next cells were incubated with rabbit monoclonal anti γ-H2A.X antibodies diluted 1:100 in blocking solution and incubated for 1 hour at room temperature. The tissues were washed three times with PBS, then a fluorescent secondary antibody, Alexa Fluorophore 555 Goat anti-rabbit IgG (Invitrogen) diluted 1:400 in blocking solution was applied and incubated for 1 hour in the dark at room temperature. Following incubation, tissues were washed three times with PBS, then coverslips were mounted to a glass slide with 4 μL Vectashield with DAPI and sealed with clear nail polish. Nuclear foci were visualized using a Zeiss LSM 880 confocal microscope (Zeiss).

2.5.6 Flow cytometry analysis of H2AX phosphorylation

Twenty-four well tissue culture plates were collagen-coated, seeded with 10^6 cells/well and infected as described in section 2.6.5. The cells were collected by trypsinisation 4 hours

post-infection and washed in PBS (without calcium). The cells were then collected by centrifugation at 200 g and resuspended in 800 μL fixable live viability dye eFluor 780 (Invitrogen) stain diluted at a ratio of 1:1000 in PBS and incubated for 20 minutes on ice. The cells were washed in excess Stain Buffer (BD Biosciences) and FC receptor block (PBS + 10% FCS) was applied before incubating for a further 20 minutes on ice. Next, the cells were fixed with Cytofix (BD Biosciences) for 15 minutes at 37°C and permeabilised with Perm Buffer (BD Biosciences) for 30 minutes on ice, before purified mouse anti γ-H2AX diluted 1:200 in Stain Buffer (BD Biosciences) was applied and cells were incubated for 1 hour at room temperature. A multichromatic-conjugated secondary antibody, goat anti-mouse IgG diluted 1:1000 in Stain Buffer (BD Biosciences) was applied and cells were incubated for 1 hour at room temperature, keeping samples protected from the light. Cells were eventually suspended in Stain Buffer (BD Biosciences) and filtered with a 70 μM filter. Cells were analysed using the BD FACSAria and the data was analysed using FloJo software. Analysis of the stained cell populations was performed by gating on single, live cells.

2.6 *In vitro* HeLa cell culture assays

2.6.1 Maintaining HeLa cells

HeLa cells were routinely cultured in Dulbecco's Minimal Eagles Medium (DMEM) (ThermoFisher Scientific) with 10% (*v/v*) foetal calf serum (FCS) at 37°C, in a 5% CO_2 incubator and were maintained by serial passage. Confluent cells were trypsonised by adding 2 mL of Trypsin-EDTA to the flask and incubating for 5 minutes at 37°C, in a 5% CO_2 incubator. Cells were resuspended in 8 mL pre-warmed DMEM media and either seeded into a fresh flask or onto glass coverslips.

2.6.2 EPEC cell adhesion assay

HeLa cells were cultivated for 48 hours (70-80% confluency) in Dulbecco's minimum essential media (DMEM) supplemented with 10% (*v/v*) FCS. 12-well plates were seeded with $2x10^4$ HeLa cells per well and incubated overnight at 37°C with 5% CO_2. EPEC strains (transformed with pGFP or pRFP) were used for adhesion assays. Bacterial cultures for infection were inoculated at a ratio of 1:100 from overnight cultures and grown in DMEM at 37°C for 4 hours until an $OD_{600\ nm}$ of 0.6 was reached. Seeded cells were washed with fresh media and the tissues were infected with 100 µL bacterial culture (adjusted to an $OD_{600\ nm}$ of 0.1) in 400 µL fresh DMEM. Plates were centrifuges at 200 g for 2 minutes and incubated at 37°C in 5% CO_2 for 1.5 hours. Wells were washed four times with PBS to remove non-adherent bacteria and then cells were lysed with 500 µL of 1% Triton X-100 solution for 15 minutes. Bacterial lysates were serially diluted in PBS and spotted on to LB agar plates. After overnight incubation at 28°C, single colonies were counted and colony forming units (CFU) for each strain was determined. Infections were performed in triplicate and bacterial counts were analysed in GraphPad Prism 8.

2.6.3 Fluorescent acting staining (FAS) assay

For visualising attaching and effacing (A/E) lesions on the surface of infected HeLa cells, the FAS assay was performed**.** Cells were cultivated and infected as described above (2.6.2), with the exception that HeLa cells were seeded onto 13 mm glass coverslips in a multi-well plate. Post-infection, tissues were washed three times with PBS to remove any non-adherent bacteria, before cells were fixed with 250 µL 4% (*w/v*) paraformaldehyde (PFA) for 20 minutes. Wells were washed three times with PBS and permeabilised with 0.1% Triton X-100 solution for 5 minutes. Following three more washes with PBS, cells were incubated in the dark with 1:500 dilution of 0.2 U Phalloidin-AlexaFluorophore555 (Invitrogen) for 1 hour. Cells were washed another three times then coverslips were

mounted onto microscope slides using Vectashield mounting media with DAPI staining and sealed with clear nail polish. Slides were visualised using a Zeiss AxioImager M1 microscope and images were acquired and deconvoluted using Zen 2.3 Pro software.

2.6.4 Examination of cellular senescence in HeLa cells

HeLa cells were cultivated for 48 hours (70-80% confluency) in Dulbecco's minimum essential media (DMEM) supplemented with 10% (*v/v*) FCS. Glass coverslips were pre-coated with 0.5% (*v/v*) collagen then seeded with $4x10^4$ HeLa cells and incubated overnight at 37°C with 5% CO_2. Nissle 1917, DH10B pBAC-*pks* and DC10B *E. coli* strains were used for senescence assays. Bacterial cultures for infection were grown in MEM-HEPES with or without 1 mM D-Serine at 37°C for 4.5 hours, 200 RPM (New Brunswick Scientific controlled environment shaker). Seeded cells were washed with fresh media and then infected with 200 µL bacterial culture (adjusted to an $OD_{600\ nm}$ of 0.1) in 300 µL fresh MEM-HEPES (supplemented with 1% (*v/v*) L-Glutamine) with or without 1 mM D-Serine. Plates were centrifuged at 200 g for 2 minutes and incubated at 37°C in 5% CO_2 for 4 hours. The cells were washed twice with PBS, 4 hours after inoculation, then cells were replenished with DMEM containing 10% (*v/v*) FCS and 50 µg/mL gentamicin and incubated for 72 hours at 37°C, 5% CO_2. The media was removed, and cells were washed with PBS then fixed in 4% (*w/v*) paraformaldehyde for 20 minutes. Cells were permeabilised with 0.1% (*v/v*) Triton X-100 for 5 minutes. After two washes with PBS, each coverslip was stained with 0.2 U Phalloidin-AlexaFluorophore555 and incubated in the dark for 1 hour. The cells were washed twice with PBS before the coverslips were mounted onto microscope slides with 4 µL Vectashield with DAPI and sealed with clear nail polish. Images were acquired using a Zeiss AxioImager M1 and images were processed by deconvolution using Zen 2.3 Pro software (Zeiss). The area of each cell was measured using a pipeline developed on CellProfiler (McQuin *et al.*, 2018). Briefly,

sample images were acquired at 10X magnification to allow for >100 cells to be captured per image. Image files were uploaded to the CellProfiler workspace and analysis was performed for images taken from 3 replicate experiments.

2.7 Isolation of Enteropathogenic *Escherichia coli* (EPEC)

2.7.1 HiCrome ECC selective agar

E. coli-coliform (ECC) ChromoSelect Selective Agar was purchased from Sigma Aldrich (product no. 85927-26.5G-F) and the media was prepared in accordance with the manufacturer's instructions. Briefly, into a sterile bottle, 13.25 g of ECC agar powder was added and the bottle topped up to 500 mL with ddH_2O. The media was heated to 100^oC, with stirring, for 40 minutes until the liquid had turned translucent, indicating the agar was fully dissolved and ready to be poured into circular Petri dishes. Glycerol stocks of EPEC strains were removed from the -80^oC freezer, and a 1 µL loop was used to streak for single colonies. Plates were incubated at 37^oC overnight, then plates were inspected for blue-violet colonies indicating the presence of *E. coli*.

2.7.2 Spiking healthy stool specimens with EPEC

Healthy stool specimens, fresh frozen and stored at -80^oC, were gifted by Professor Konstantinos Gerasimidis at the Royal Infirmary Glasgow. Approximately 100 mg of stool was weighed out into a pre-weighed Eppendorf tube. 800 µL of PBS was added to the Eppendorf and the tube was vortexed before storing at 4^oC for 30 minutes. After softening in PBS, the stool was homogenised into a slurry by vigorous vortexing. Overnight cultures of EPEC strains were serially diluted 1:10 with PBS before 100 µL of bacterial culture was spiked into healthy stool slurry. 100 µL of spiked stool slurry was spread on to ECC plates and incubated overnight at 37^oC. Single colonies were isolated from ECC plates and boiled

in 50 µL nfH_2O for 10 minutes. Samples were centrifuged at maximum speed for 2 minutes and 10 µL of supernatant was removed and diluted 1:10 into nfH_2O before storing genomic DNA at -20ºC. *E. coli* colonies were characterised using 1 µL of genomic DNA in a multiplex PCR (2.3.3) to check for the presence of EPEC associated genes.

2.8 Growth inhibition assays

2.8.1 Colicin killing activity assay

Colicin concentrations were determined by measuring levels of absorbance with a spectrophotometer and dividing this by the coefficient factor for each colicin to give the protein concentrations. Colicins were diluted in Tris Buffer to give a final concentration of 1 mg/mL. Overnight cultures of EPEC strains were inoculated 1:100 into LB media and grown at 37ºC until an $OD_{600\ nm}$ of 0.6 was reached. A top agar solution was prepared by dissolving 0.8 g of agar powder in to 100 mL ddH_2O. Bacterial cultures were inoculated 1:100 into the top agar solution after reaching the desired optical density. Inoculated agar was then poured on to plates and allowed to dry before spotting 5 µL of colicin D, Ia, E9 and S4 on to each strain. The plates were incubated at 37ºC and inspected the following day for zones of inhibition, indicating colicin killing activity.

2.8.2 Competitive growth assay

Single colonies of Nissle 1917 and LF82 were grow in LB broth overnight at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker). 500 µL of each overnight culture was inoculated into 10 mL M63 media alone or supplemented with 1 mM D-Serine and incubated for 2 hours at 37ºC, 200 RPM (New Brunswick Scientific controlled environment shaker). From the 2-hour cultures, the two strains (Nissle 1917 and LF82) were inoculated at 10^6 CFU/mL into 10 mL fresh M6 media alone or supplemented

with 1 mM D-Serine and co-cultured for 24 hours at 37°C, 200 RPM (New Brunswick Scientific controlled environment shaker). The strains were also grown individually for 24 hours under the same conditions as a control. Following the incubation period, CFUs were enumerated by serially diluting the cultures in PBS and spotting the dilutions onto antibiotic (ampicillin) and LB agar plates. The CFU for each strain was calculated as described in section 2.2.1.

2.9 Bioinformatic analysis and statistical analysis

2.9.1 Database tools

Nucleotide sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov) and EBI (https://www.ebi.ac.uk). Alignments were performed using BLAST (hhtp://blast.ncbi.nlm.nih.gov/Blast.cgi) and Clustal Omega available through Jalview software, Version 2.11.1.7 (jalview.org). Nucleotide sequences were reverse complemented using reverse-complement (reverse-complement.com).

2.9.2 Bioinformatic analysis of LEE-encoded genes

Whole genome sequences of the clinical EPEC strains were sequenced by MicrobesNG, Birmingham. Specific LEE encoded T3SS-associated genes were selected for comparison based on previous research (Connolly *et al.*, 2015). The percentage identities of *LEE* genes in the uncharacterised strains to those of the E2348/69 EPEC reference genome (NCBI GenBank FM180568.1) were identified using the BLAST tool of the CLC Genomics Workbench (Qiagen).

2.9.3 Statistical analysis

Statistical analysis was performed using GraphPad Prism (Version 8), unless stated otherwise, an unpaired Student *t*-test was used for comparisons between 2 samples.

3 Phenotypic and genotypic characterisation of Enteropathogenic *E. coli* (EPEC)

3.1 Introduction

Diarrhoeal disease is a global health concern and in low-income countries (LICs) it remains one of the leading causes of morbidity and mortality in infants under five years of age, (Liu *et al*., 2012). The aetiological agents most often associated with these infections are diarrhoeagenic *Escherichia coli* (DEC) (Nataro, 1998), with most outbreaks associated with enteropathogenic *Escherichia coli* (EPEC) (Canizalez-Roman *et al*., 2016).

The complex pathogenesis of EPEC has been well characterised and the importance of certain virulence factors has been well defined. The *eaeA* gene located on the locus of enterocyte effacement (LEE) pathogenicity island and *bfpA* gene positioned on the EPEC adherence factor (EAF) plasmid, both have an integral role in disease and have been used for identifying and subdividing EPEC into typical and atypical strains, (Nataro & Kaper, 1998). Typical EPEC (tEPEC) strains are eae^{+} and $bfpA^{+}$, whereas atypical EPEC (aEPEC) strains are eae^{+} and $bfpA^{-}$. Indeed, most aEPEC lack the EAF plasmid, although it has been reported that some atypical strains carry a truncated version of this plasmid (Teixeira *et al.*, 2015), therefore, differentiating between the two subtypes is not always simple. EPEC belongs to a family of bacteria known as the attaching and effacing (A/E) pathogens, which also includes enterohaemorrhagic *E. coli* (EHEC) and the murine pathogen *Citrobacter rodentium* (*C. rodentium*) (Croxen and Finlay, 2009). The A/E phenotype is characterised by effacement of the microvilli and intimate adherence between the bacterium and the epithelium. Attachment of EPEC to the surface of epithelial cells initiates the translocation of effector proteins into the cell, inducing actin polymerisation and leading to the formation of distinct pedestal like structures, termed A/E lesions and the hallmark of EPEC infection (Moon *et al.*, 1983).

The A/E lesion was first identified by Knutton *et al*., 1989, who reported the accumulation of electron-dense material under adherent bacteria in human intestinal biopsies. Knutton realised that the accumulated material was polymerised actin, and that the A/E phenotype could be confirmed by using a specific actin stain to identify adherent EPEC bacteria on mucosal surfaces (Knutton *et al.*, 1989). Thus, the fluorescent actin staining (FAS) test was the first diagnostic tool to enable the identification of strains that produce A/E lesions. However, in hospital and community settings, conventional methods including microscopy and culture are labour intensive and time consuming, they can also be less sensitive resulting in low detection rates. Therefore, molecular based diagnostics have become the most common method for identification of EPEC and other enteric pathogens. PCR-based assays provide high sensitivity and specificity for the detection of strain specific virulence genes, however, many of these are not suitable for the detection of multiple diarrhoeagenic organisms. Thus, this results chapter will aim to describe suitable screening methods to isolate and distinguish DEC pathotypes from stool specimens collected in Vietnam. In addition, the phenotypic and genotypic characterisation of isolated EPEC strains will be discussed.

3.2 Results

3.2.1 Isolation and characterisation methods for EPEC contaminated stool

A multiplex PCR assay to differentiate between DEC pathogroups was designed and evaluated using prototype laboratory strains. DEC strains were obtained from our own laboratory collection and from external sources including the Natural Culture Collection (NCTC) and the Edinburgh Reference Laboratory; a description of each strain is listed in Table 3-1.

Table 3-1 Bacterial strain and source information.

Diarrhoeagenic Pathogroup	Strain	Description	Source
Enterohaemorrhagic *E. coli*	TUV 93-O	Wild type *E. coli* O157:H7 (Shiga toxin negative)	Roe lab inventory
Enteropathogenic *E. coli* (Atypical)	E110019	Wild type atypical EPEC	Prof Brendan Kenny
Enteropathogenic *E. coli* (Typical)	E2348/69	Wild type typical EPEC	Prof David Smith
Shiga Toxin *E. coli*	ICC168	Wild type *C. rodentium*	Prof Bret Findlay
Enterotoxigenic *E. coli*	239-2	Wild type ETEC	Scottish Microbiology Reference Laboratories, Edinburgh
Enteroinvasive	NCTC12984	Wild type *S. sonnei*	National Collection

E. coli			of Type Cultures (NCTC)

To characterise each DEC pathogroup, the detection of strain specific virulence genes was employed. The gene targets selected for each group were *eae* for EHEC and aEPEC, *eae* and *bfpA* for tEPEC, *stx* for STEC, *elt* for ETEC and *ipaH* for EIEC isolates. The oligonucleotide sequences and amplicon sizes for these genes are listed in Table 3-2.

Table 3-2 PCR oligonucleotide sequences for amplification of diarrhoeagenic *E. coli* virulence genes.

Diarrhoeagenic Category	Target Gene	Gene Product	Oligonucleotide Sequence	PCR Product Size	Reference
EHEC	Intimin	*eae*	TCAATGCAGTTCCGTTATCAGTT GTAAAGTCCGTTACCCCAACCTG	482	Vidal et al, 2004
aEPEC	Intimin	*eae*	TCAATGCAGTTCCGTTATCAGTT GTAAAGTCCGTTACCCCAACCTG	482	Vidal et al, 2004
tEPEC	Intimin	*eae*	TCAATGCAGTTCCGTTATCAGTT GTAAAGTCCGTTACCCCAACCTG	482	Vidal et al, 2004
	Bundle Forming Pilus	*bfp*	GGAAGTCAAATTCATGGGGGTAT GGAATCAGACGCAGACTGGTAGT	300	Vidal et al, 2004
STEC	Shiga Toxin	*stx1* & *stx2*	GAGCGAAATAATTTATATGTG TGATGATGGCAATTCAGTAT	518	Toma et al, 2003
ETEC	Heat Labile Toxin	*elt*	ACGGCGTTACTATCCTCTC TGGTCTCGGTCAGATATGTG	273	Rodas et al, 2009
EIEC	Invasion Plasmid Antigen H	*ipaH*	GTTCCTTGACCGCCTTTCCGATACCGTC GCCGGTCAGCCACCCTCTGAGAGTAC	199	This study

To test the specificity of the oligonucleotide sequences listed in Table 3-2, a conventional singleplex PCR was performed on single colonies for each DEC pathotype. PCR products were run on a 2% agarose gel to confirm the amplification of specific virulence genes.

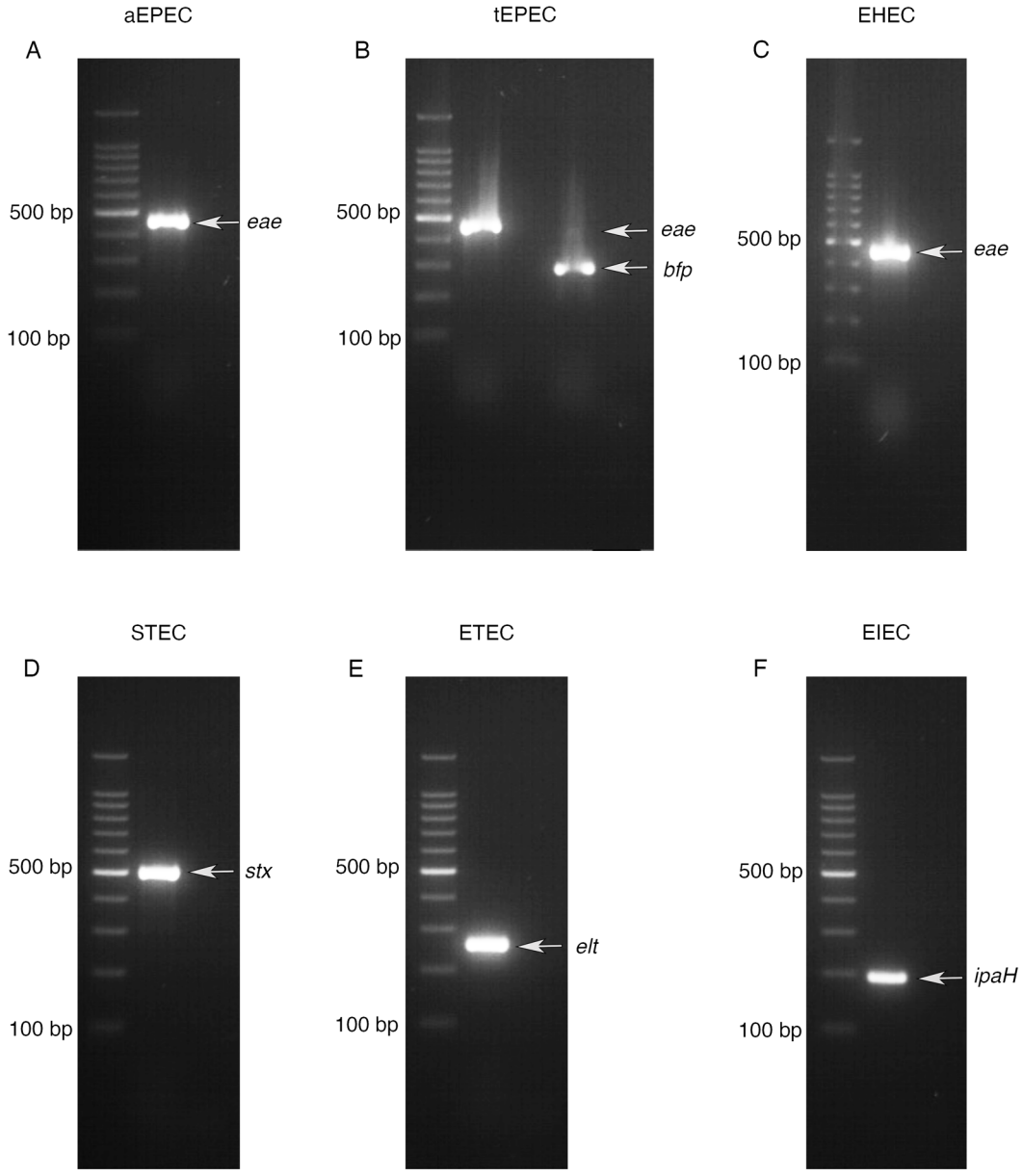


Fig 3-1 Singleplex PCR reactions to validate amplification of virulence genes. PCR products obtained from singleplex PCR with strain-specific primers, analysed by agarose gel electrophoresis showing strain specificity.

Fig 3-1A-C shows the amplification of *eae* the structural gene for intimin in aEPEC, tEPEC and EHEC; Fig 3-1B shows the amplification of *bfpA,* the structural gene for the bundle forming pilus (BFP) of tEPEC; Fig 3-1D illustrates the amplification of *stx* specific for Shiga toxins 1 and 2 of EHEC; Fig 3-1E shows amplification of *elt,* the heat labile enterotoxin of ETEC and Fig 3-1F illustrates the amplification of *ipaH,* the invasion plasmid antigen found in EIEC and *Shigella*. Together, these results demonstrated the specificity of the oligonucleotide sequences and confirmed the amplification of individual DEC associated virulence genes.

A protocol for isolating *E. coli* strains from stool specimens was developed using the prototype DEC strains listed in Table 3-2. The methods were developed and validated using control samples. Healthy stool was obtained from the Glasgow Royal Infirmary and homogenised before being spiked with a single DEC strain and plated on to *E. coli*-coliform (ECC) HiCrome selective agar plates. Plates were incubated and then single colonies were subjected to a multiplex PCR; this protocol is illustrated schematically in Fig 3-2.

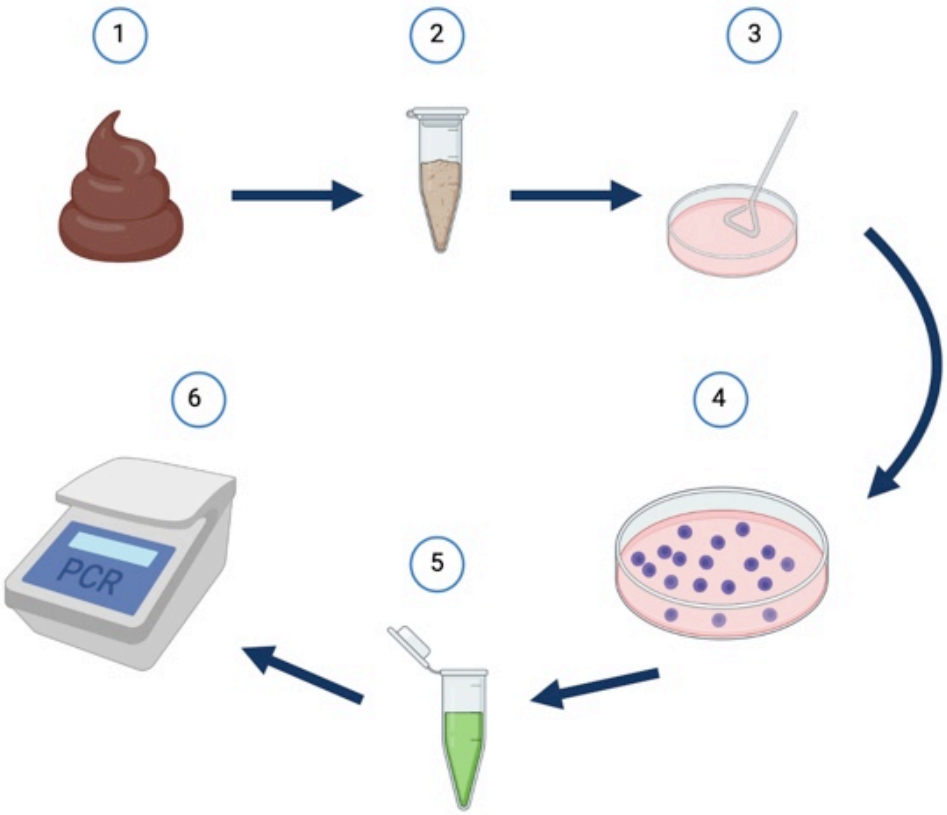


Fig 3-2 Schematic detailing sample processing for the multiplex PCR assay. 1- Stool from individuals presenting with gastroenteritis is collected. **2**- Stool is homogenised with PBS and processed into a slurry. **3**- Slurry is spread on to *ECC* selective agar plates and incubated at 37ºC overnight. **4**- Single *E. coli* colonies present as purply/blue to pink in colour. **5**- A single colony is isolated from the ECC plate and inoculated into an Eppendorf containing the multiplex reaction mix. **6**- The PCR mixture is then amplified by thermal cycling. The final PCR products are then visualised following electrophoresis on a 2% agarose gel.

To enable the simultaneous detection of DEC specific genes, the multiplex reaction mix contained multiple primer pairs which enabled amplification of all the genes listed in Table 3-2, in a single reaction. The primers for each gene were designed so that the PCR products would differ in size, thus enabling DEC pathogroups to be distinguished on this basis. To test the specificity of the multiplex reaction mix, healthy stool slurries were prepared and spiked with a single DEC strain before being plated on to HiCrome ECC plates as

described above. After incubating, single colonies were isolated and subjected to multiplex PCR, then following electrophoresis on a 2% agarose gel, PCR products were visualised. The result for each DEC strain is shown in Fig 3-3.

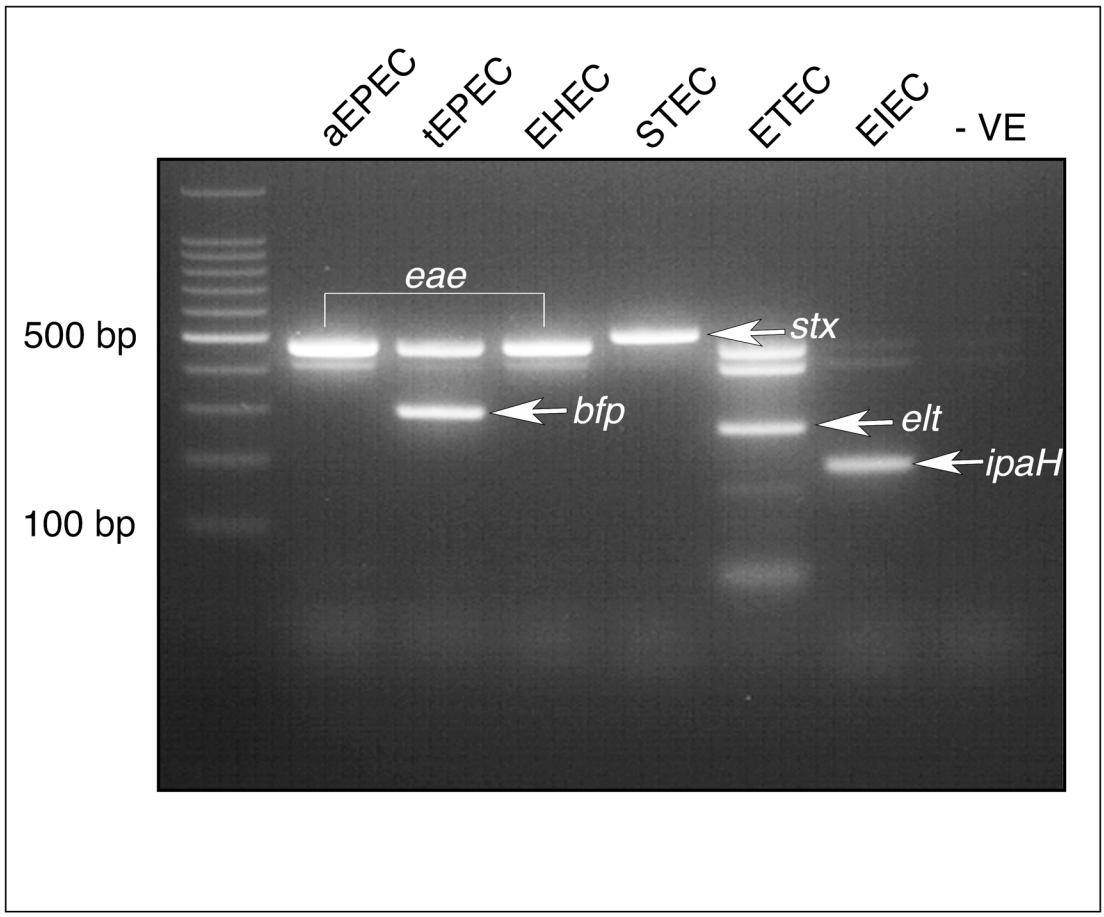


Fig 3-3 Multiplex PCR reactions to validate species specific amplification. PCR products obtained from a multiplex PCR with strain-specific primers, analysed by agarose gel electrophoresis showing primer specificity in each DEC pathogroup.

Fig 3-3 confirms that primer specificity was conserved in the multiplex reaction mix, as amplification of strain-specific virulence genes was evident in each DEC. Furthermore, the efficient amplification of DEC virulence genes by multiplex PCR demonstrates that the conditions of the reaction were optimal for primer-specific amplification. A few non-specific bands were observed in ETEC that indicated a slight degree of mismatching,

however, it remained possible to distinguish amplification of the *elt* virulence gene from these other bands. Therefore, overall, the multiplex PCR demonstrated strong specificity and enabled identification of DEC pathotypes using a single reaction.

To determine the sensitivity of the multiplex reaction, stool slurry was spiked with decreasing CFU/mL of typical and atypical EPEC.

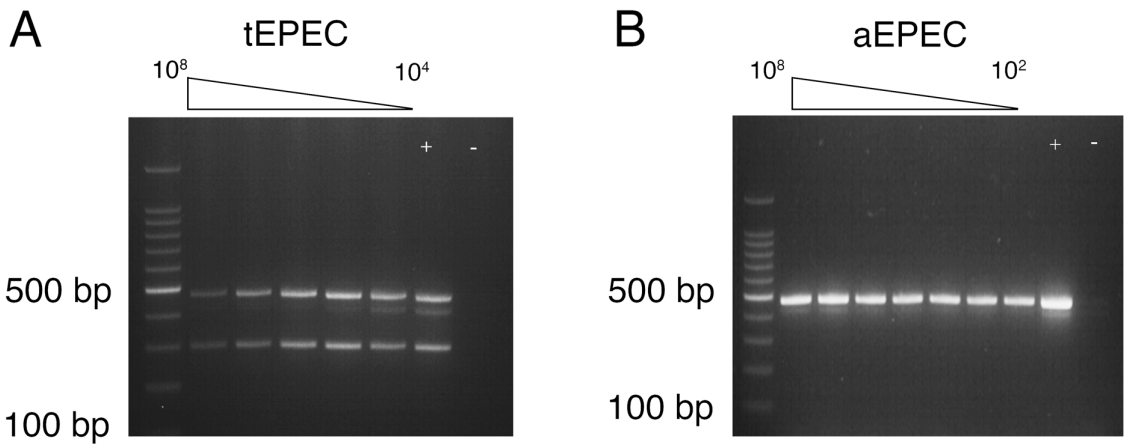


Fig 3-4 Sensitivity of multiplex PCR from a stool sample spiked with typical and atypical EPEC. PCR products obtained from a multiplex PCR on spiked stool samples (A) decreasing dilutions of tEPEC from 10^8 – 10^4 CFU/mL and (B) decreasing dilutions of aEPEC from 10^8 – 10^2 CFU/mL (B) are indicated. Genomic DNA was used as a positive control (+) and a negative control (-) containing no bacteria, was also run.

The limit of detection for the multiplex PCR was determined by spiking healthy stool samples with a PBS suspension of prototypic EPEC strains in serial 10-fold dilutions, to give a range of CFU/mL concentrations. Each serial dilution of spiked stool sample was spread on ECC HiCrome agar plates and incubated at 37^{o}C, then single colonies were tested by multiplex PCR. Fig 3-4 demonstrates the sensitivity of the assay, as detection of strain-specific genes was evident for tEPEC and aEPEC, at the lowest concentrations.

3.2.2 Phenotypic characterisation of prototype EPEC strains

Induction of A/E lesions are typically associated with EPEC infections in eukaryotic cells and can be characterised by performing the fluorescence actin staining (FAS) assay. The FAS assay was first described in 1989 and utilises a fluorescence-labelled phallotoxin to stain the actin cytoskeleton; intense spots of fluorescence correspond to the precise attachment site of single bacterium (Knutton *et al.*, 1989).

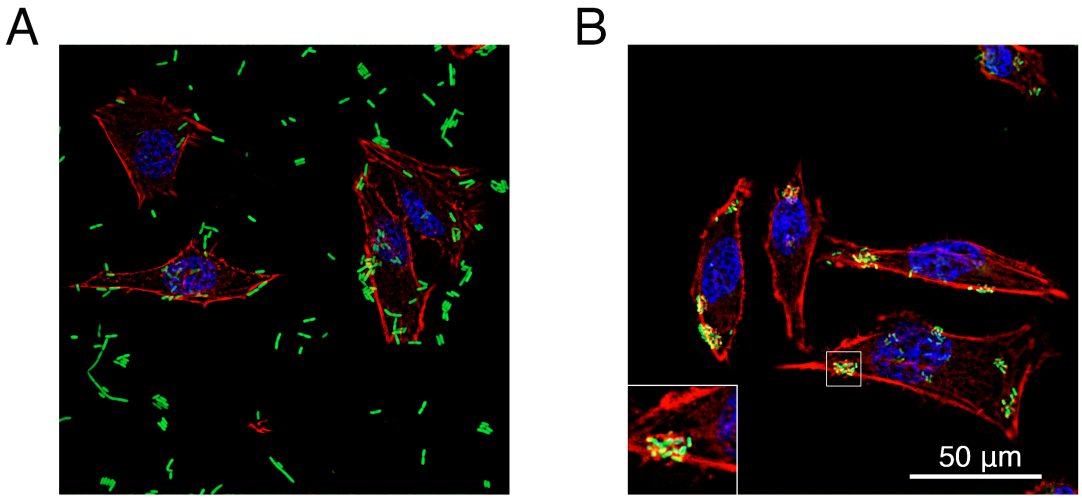


Fig 3-5 Atypical and typical EPEC adherence patterns. Cells were infected with aEPEC strain E110019 (A) and tEPEC strain E2348/69 (B) transformed with p*rpsM-gfp* (green) at an MOI = 10. Actin cytoskeleton was stained with Phalloidin-Alexa Fluor 555 (red) and DNA was counterstained with DAPI (blue). Scale bar = 50 µm. The inset in (B) contains a magnified section of the indicated area and depicts actin condensation at the site of bacterial attachment. Images were taken at 40X magnification and stacks of 14 images were deconvoluted using ZenPro software.

HeLa cells were infected for 1.5 h with EPEC prototype strains, aEPEC and tEPEC. Next, tissues were washed and fixed before coverslips were stained with Phalloidin-Alexa Fluor

555 and mounted on to microscope slides using Vectashield with DAPI. Cells were visualised using fluorescence microscopy and notably different adherence patterns were observed for each strain. Fig 3-5A depicts the colonisation of aEPEC, in which a diffuse adherence (DA) pattern was demonstrated where the bacteria covered the cells uniformly (Scaletsky, Silva and Trabulsi, 1984). Whereas tEPEC (Fig 3-5B) adhered tightly to HeLa cells and demonstrated a localised adherence (LA) pattern where dense microcolonies were formed on the cell surface. Furthermore, in Fig 3-5B A/E lesions were visible in regions where dense spots of red indicated actin condensation underneath the bacterial cell, suggesting that more efficient lesion formation was associated with tEPEC infection.

The LA adherence phenotype of tEPEC has been attributed to the presence of the EAF plasmid, which carries the cluster of genes encoding for the BFP and enables bacteria to tightly interconnect forming microcolonies on the cell surface (Nataro and Kaper, 1998). To determine the colonisation efficiency, adherent bacterial cells were enumerated following infection with atypical and typical strains.

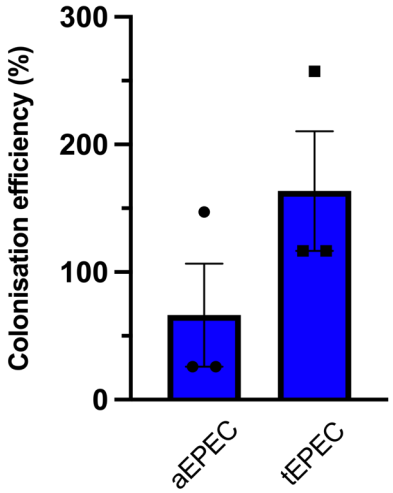


Fig 3-6 Colonisation efficiency of atypical and typical EPEC. Following a 1.5 h infection, HeLa cells were washed to remove nonadherent bacteria and subsequently lysed to release attached bacteria. The CFU for each of the prototypic EPEC lysates was

enumerated and the colonisation efficiency was calculated by expressing the CFU as a percentage of the inoculum.

Fig 3-6 shows the colonisation efficiency of each EPEC strain, determined from three independent experiments. The results indicated that tEPEC displayed a higher ability to adhere to HeLa cells compared with aEPEC and confirmed the adherence phenotypes described in Fig 3-5.

3.2.3 Identifying a potential mode for clearing EPEC carriage

In preparation for receiving clinical EPEC samples, assays were developed to characterise strains genotypically and phenotypically. In addition, to identify a potential mode for clearing EPEC carriage in stool, naturally produced colicins were tested for their killing activity against tEPEC (E2348/69) and aEPEC (E110019) prototype strains. Colicins are toxic proteins produced by and active against *E. coli* and closely related bacteria. There are several mechanisms by which colicins can elicit antibacterial activity, including pore-formation and nuclease activity (Yang *et al.*, 2014).

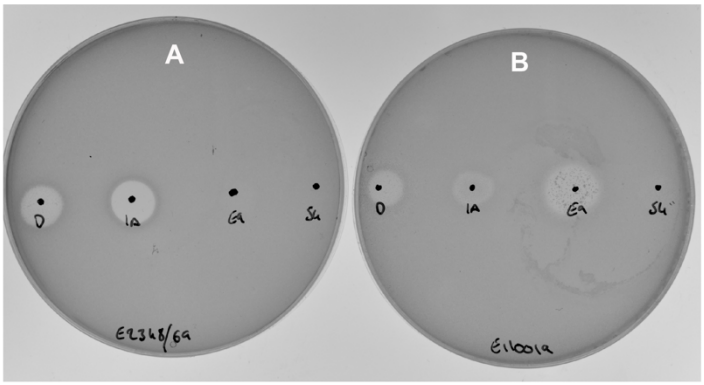


Fig 3-7 Colicin killing activity against typical and atypical EPEC. Solid medium killing assays of (A) tEPEC (E2348/69) and (B) aEPEC (E110019) strains exposed to purified

colicin D, Ia, E9 and S4. The circular zones of inhibition represent the degree of colicin killing activity.

Cultures of typical and atypical strains, E2348/69 and E110019 respectively, were grown to an $OD_{600\ nm}$ of 0.6 and then overlaid onto LB agar plates. Colicin D, Ia, E9 and S4 were then spotted on to the dry agar and incubated at 37^{o}C. Inhibition zones were produced by the activity of colicin D and Ia on plates overlaid with tEPEC (Fig 3-7A). The same colicin inhibition was observed on plates overlaid with aEPEC (Fig 3-7B), however, this plate displayed an additional inhibition zone which was produced by colicin E9. Colicin D and E9 exert nuclease activities by digesting DNA and RNA, respectively whereas colicin Ia forms pores in the inner membrane and leads to bacterial cell death (Yang *et al.*, 2014). Colicin S4, also a pore-forming colicin with a different receptor (Yang *et al.*, 2014), displayed no killing activity against either EPEC subtypes. Therefore, as colicin D and Ia displayed killing activity against both EPEC strains, these colicins were carried forward to be tested on the clinical EPEC strains.

3.2.4 Phenotypic characterisation of clinical EPEC isolates

Stool specimens were collected from children with gastroenteritis admitted to the Hospital of Tropical Diseases in Ho Chi Minh City, Vietnam. Colonies isolated from MacConkey selective agar plates were subjected to in-house diagnostic tests to identify the causative agent/s, and isolates that were identified as EPEC were gifted to the Roe group for further characterisation. For ease, the EPEC strains were named according to the sample ID of the patient that they originated from; a summary of the clinical data is presented in Table 3-3.

Table 3-3 Patient clinical data.

Sample ID	Patient status	Age (months)	Date Received (mm/yy)
0259	gastroenteritis	Unknown	03/14
0898	gastroenteritis	12	01/16
0112	gastroenteritis	10	11/14
0223	gastroenteritis	9	10/14
0271	gastroenteritis	4	06/14
0132	gastroenteritis	2	04/14
0145	gastroenteritis	22	11/14
0017	gastroenteritis	16	05/14
0108	gastroenteritis	18	07/14
BCB-22675	healthy	42	04/16
BCB-21827	healthy	54	04/16
BCB-21787	healthy	54	04/16

First, the clinical EPEC strains were streaked onto plates containing selective agar. HiCrome ECC media was used for this purpose as the composition of the media promotes rapid growth of *E. coli* and coliforms. The inclusion of Tergitol™ inhibits the growth of Gram-positive bacteria as well as other Gram-negative bacteria, other than coliforms. Interestingly, the strains displayed an array of colours which ranged from dark blue to light pink as shown in Fig 3-8.

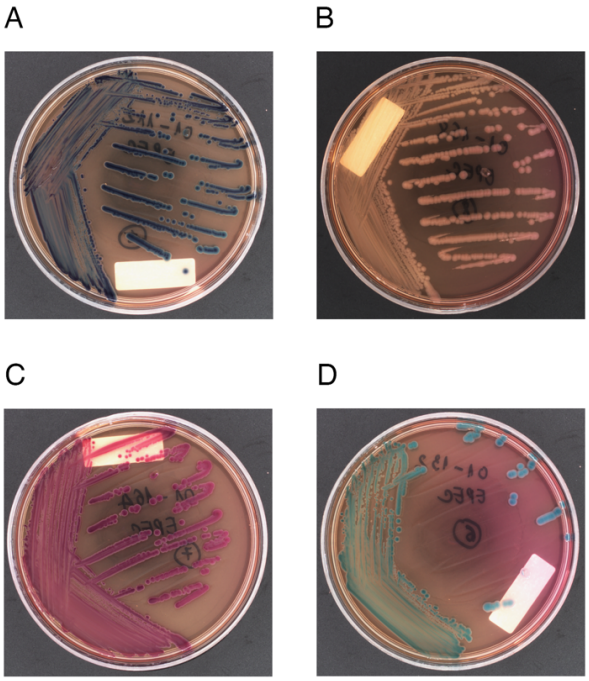


Fig 3-8 ECC HiCrome agar plates. Streak plates for strains isolated from Vietnam (A) 0112, (B) 0168, (C) 0164 and (D) 0132 indicating the range of colony colours.

ECC HiCrome agar can be used for the simultaneous detection of *E. coli* and coliforms from environmental samples. The chromogenic mixture contains two chromogenic substrates, Salmon-GAL and X-glucuronide. The enzyme β-D-galactosidase produced by coliforms cleaves Salmon-GAL, resulting in the appearance of salmon-to-red coloured colonies. The enzyme β-D-glucuronidase produced by *E. coli*, cleaves X-glucuronide. *E. coli* forms dark blue to violet-coloured colonies because of cleaving both Salmon-GAL and X-glucuronide. Indeed, the colony colours represented in Fig 3-8A-D indicated that the isolates were strains of *E. coli*, however, to confirm if they were EPEC, single colonies were subjected to multiplex PCR.

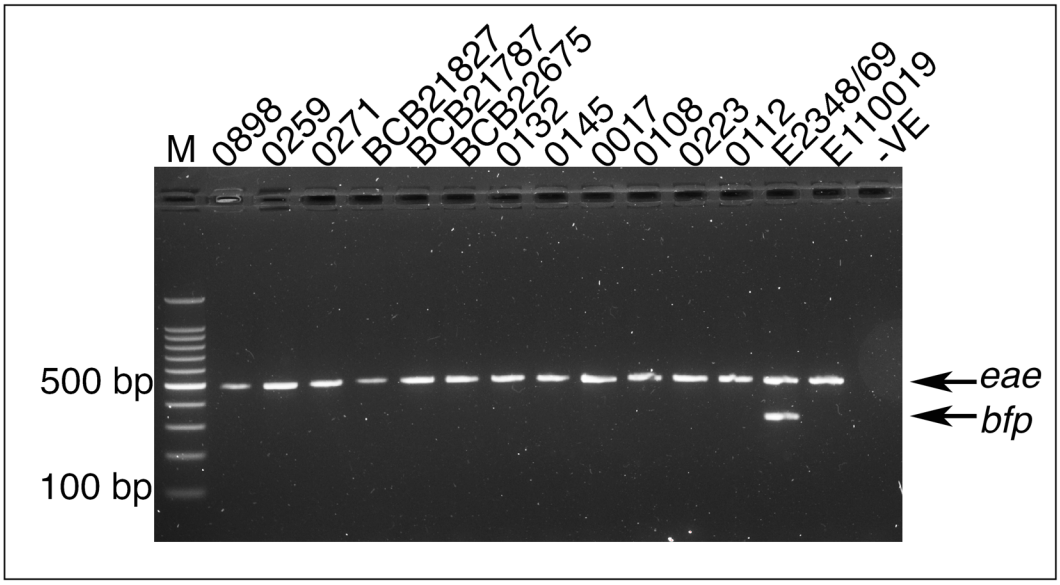


Fig 3-9 Multiplex PCR amplification of clinical EPEC isolates. PCR products obtained from a multiplex PCR on single colonies isolated from ECC HiCrome plates. Lane 1: size marker; Lane 2 -13: clinical EPEC strains; Lane 14 & 15: positive controls using genomic DNA from tEPEC (E2348/69) and aEPEC (E110019) prototype strains; Lane 16: negative control that contained no template.

The PCR products from the multiplex PCR revealed that the single colony isolates were *bfp* negative and only the *eae* gene was amplified in the clinical strains, suggesting that these were aEPEC. The size of the band, approximately 482 bp, indicated that the amplicons were for the intimin gene and not for Shiga toxin, which would have presented as a slightly higher band on the gel. The PCR products were also identical to the product in lane 14, in which genomic DNA from aEPEC strain E110019 had been amplified, further confirming that the clinical isolates were of the atypical EPEC subtype.

To characterise the clinical aEPEC isolates phenotypically, HeLa cells were infected with strains transformed with a plasmid expressing GFP or RFP and their ability to form A/E lesions was assessed. Transient infection of HeLa cells was performed for 1.5 h with aEPEC clinical isolates at an MOI = 10. Cells were fixed and stained as described in

section 3.2.2, before coverslips were mounted onto microscope slides. Cells were visualised using fluorescence microscopy and strains were assessed for their ability to induce actin condensation and form A/E lesions.

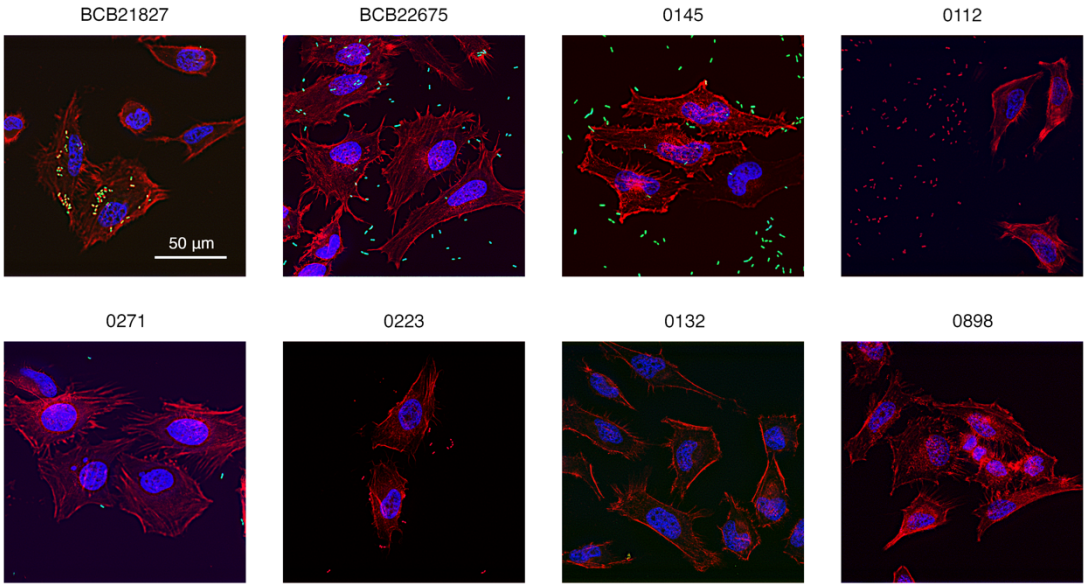


Fig 3-10 Adherence patterns of clinical aEPEC isolates. HeLa cells were infected with aEPEC isolates transformed with p*rpsM-gfp* (green) or p*rpsM-rfp* (red) at an MOI = 10. Actin cytoskeleton was stained with phalloidin-Alexa Fluor 555 (red) and DNA was counterstained with DAPI (blue). Scale bar = 50 µm. Images were taken at 40X magnification and stacks of 14 images were deconvoluted using ZenPro software.

The images depicted in Fig 3-10 were representative of the different adherence patterns observed. Punctate spots of red indicated actin condensation in HeLa cells that were infected with strain BCB21827, however it did not appear that dense microcolonies were formed on the cell surface and fewer adherent bacteria were observed compared to the prototypic tEPEC control (Fig 3-5A). Strains BCB22675 and 0145 displayed the more characteristic diffuse pattern of adherence, commonly associated with aEPEC (Fig 3-5B), though actin condensation was not apparent at the site of attachment. Interestingly, strain 0112 presented with a biofilm like adherence pattern, where the bacteria appeared to

adhere to the surface of the tissue culture plate, as opposed to attaching to the HeLa cells. The remaining strains demonstrated poor adherence to HeLa cells, as very few bacteria remained after cells were washed.

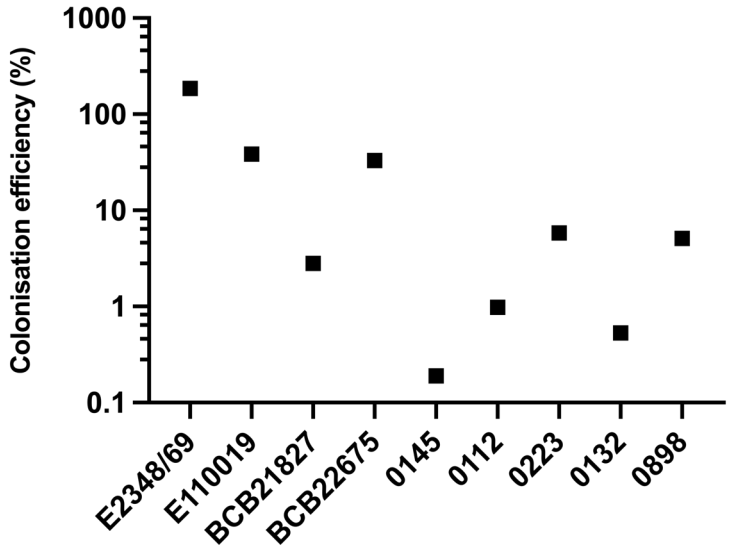


Fig 3-11 Colonisation efficiency of clinical aEPEC isolates. Following a 1.5 h infection, HeLa cells were washed to remove nonadherent bacteria and subsequently lysed to release attached bacteria. The CFU for each of the clinical isolates and prototypic EPEC lysates was enumerated, and the colonisation efficiency was calculated by expressing the CFU as a percentage of the inoculum.

The number of adherent bacteria was enumerated, and the colonisation efficiencies were compared. Fig 3-11 shows that none of the clinical isolates had a colonisation efficiency higher than that of the tEPEC (E2348/69) and aEPEC (E110019) prototype strains. Strikingly, the most efficient coloniser out of the clinical isolates was strain BCB22675, which was isolated from a heathy individual. The remaining clinical isolates appeared to be poorly adherent, as the colonisation efficiencies were determined as less than 10% for each strain in this experiment. Growth curves performed in DMEM media prior to the cell culture assay revealed that growth in this media was variable among the strains, see

Appendix 1, and could be a possible explanation for the poor adherence observed in some of the strains.

Examination of the adherence patterns had revealed that most of the clinical isolates were poorly adherent and did not form characteristic A/E lesions during the infection of HeLa cells. Therefore, the protein section profiles of these strains were analysed to investigate the production of T3SS-associated effector proteins. Cultures were grown in DMEM media until an $OD_{600\ nm}$ of approximately 0.8 was reached. Next, cells were pelleted and the supernatant was removed, then filtered before proteins were extracted by TCA precipitation. Proteins were separated by SDS-PAGE and visualised by staining the gel with Coomassie blue.

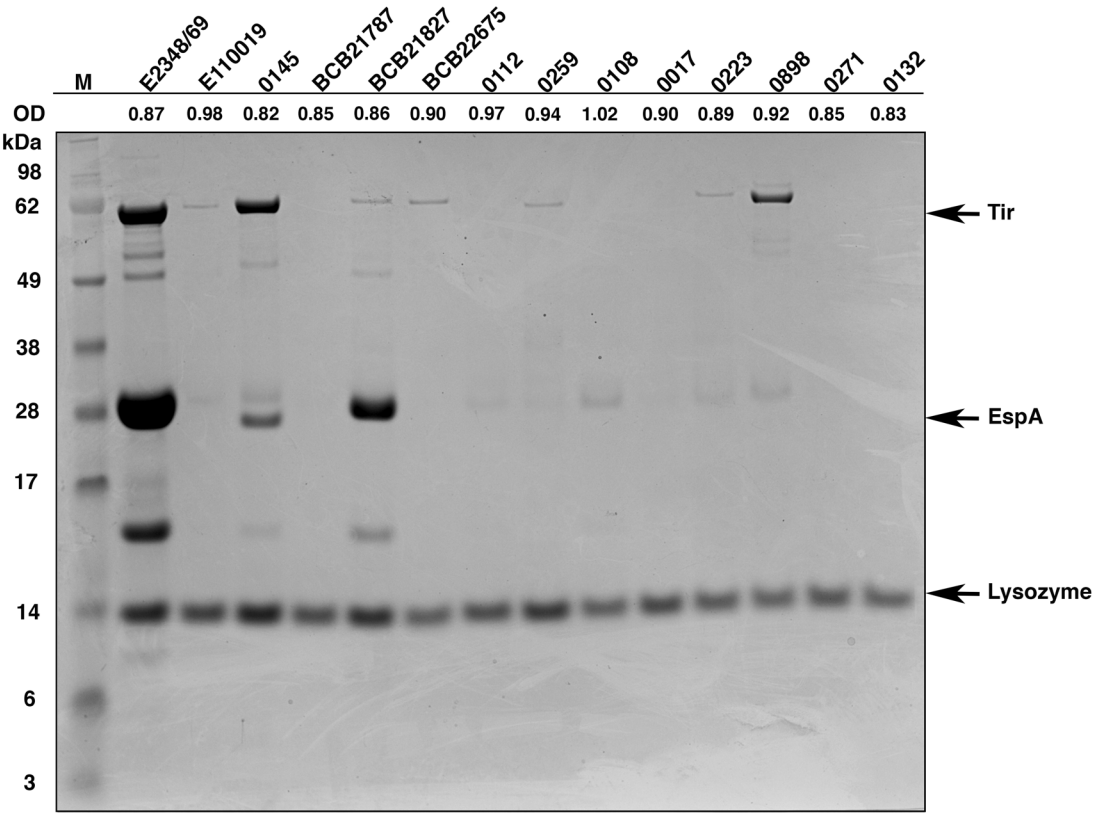


Fig 3-12 Type 3 secretion profile of EPEC isolates. SDS-PAGE analysis of the type 3 secretion profile from tEPEC (E2348/69) and aEPEC (E110019) prototype strains and

aEPEC clinical isolates cultured in DMEM until an $OD_{600\ nm}$ of approximately 0.8 was reached. Protein bands suspected to correspond to Tir and EspA are indicated with black arrows; lysozyme was added to the samples as a protein control.

The protein bands in Fig 3-12 illustrate the diverse secretion profiles that were displayed amongst the clinical aEPEC isolates. Protein bands of approximately 74 kDa in size were predicted to correspond to the translocated intimin receptor protein, Tir (Kenny *et al.*, 1997) and was present in 6 out of 12 clinical aEPEC isolates. The strongest band for Tir was produced by strain 0145 followed by strain 0898, however, neither of these strains produced A/E lesions and were poorly adherent to HeLa cells. Interestingly, the secretion profile of strain BCB21827, which was the only A/E lesion forming isolate, indicated a faint band corresponding to Tir, but a strong band corresponding to EspA, and exhibited a similar band pattern to that of the tEPEC prototype strain E2348/69.

To further investigate the variable adherence phenotypes observed in the clinical isolates, strains were sent for whole genome sequencing (WGS) and then the sequences of LEE encoded genes were compared using the BLAST feature available through the CLC genomics workbench (Qiagen, version 7.5.2). Trimmed FASTA files containing whole genome sequences were uploaded onto the CLC genomics work bench and BLAST searches were performed on genes encoding T3SS-associated effector proteins using the tEPEC strain E2348/69 (NCBI GenBank FM180568.1) as the control. The percentage identity of a gene was calculated as the number of identical base pairs (defined as “hits”) in the nucleotide sequences of the clinical strain, compared with the control. In addition, the strains were assigned a phylogroup using ClermonTyping an *in-silico* phylo-typing method (Clermont *et al.*, 2013). The assigned phylogroup and percentage identities for the T3SS associated genes for each clinical strain, are summarised in the heat map in Fig 3-13.

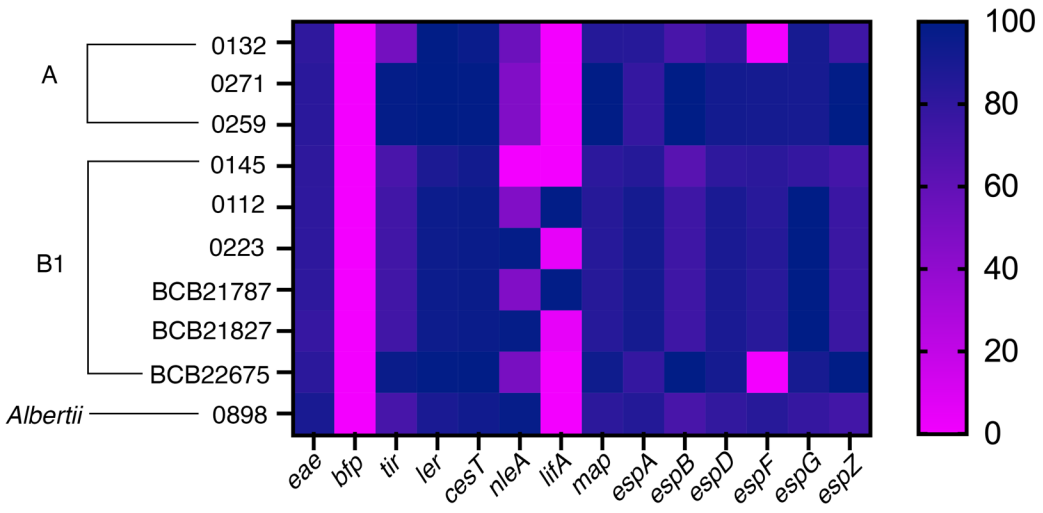


Fig 3-13 Percentage identity of aEPEC clinical isolates using nucleotide coverage. Heat map showing BLAST percentage identity for best reciprocal BLASTn hits to the complete genome of tEPEC strain E2348/69.

The BLAST searches verified that the *bfp* gene was not carried by any of the clinical strains, as no hits were detected, confirming that these were indeed aEPEC strains. The intimin encoding gene, *eae,* showed high sequence similarity in the clinical strains, however, there was a degree of variability in the percentage identities for the *tir* gene among the strains with BCB2675, 0259 and 0271 sharing >90% identity compared to the other strains which shared <75% identity to E2348/69. Interestingly, all but two of the clinical strains displayed either low or no sequence identity to the gene *lifA,* for lymphocyte inhibitory factor A, which has been associated with cell adhesion and the efficiency of A/E lesion formation in EPEC (Badea *et al.*, 2003; Cepeda-Molero *et al.*, 2017). However, despite the proposed role of *lifA*, the only lesion producing strain, BCB21827, displayed a low percentage identity (4.5%) with the control. Another intriguing observation was made when comparing the percentage identities of the gene *espF* which encodes for an effector protein that disrupts tight junctions and interferes with sodium hydrogen exchange in eukaryotic cells (Hodges *et al.*, 2008; Holmes *et al.*, 2010).

The clinical strains BCB22675 and 0132 both shared no percentage identity (0%) to *espF* in the BLAST query, which was in stark contrast to the other strains which displayed >80% identity and suggested that this gene might be lacking in these two strains. Altogether, the BLAST queries performed on a small set of genes revealed that there was a degree of variation amongst the clinical isolates, indicating that the strains were not clonal, even though they were collected from patients admitted to the same hospital.

To investigate if the strains could be inhibited using a common colicin, the colicins identified as having strong killing activity against the prototypic strains (Fig 3-7), were tested against the clinical strains. As described previously, cultures were grown to an $OD_{600\ nm}$ of 0.6 and then overlaid onto LB agar plates. Colicin D and colicin Ia were then spotted on to the dry agar and incubated at 37^{o}C. Plates were inspected for circular zones of inhibition to indicate the killing activity of each colicin; the extent of killing activity in each strain is illustrated in Fig 3-14.

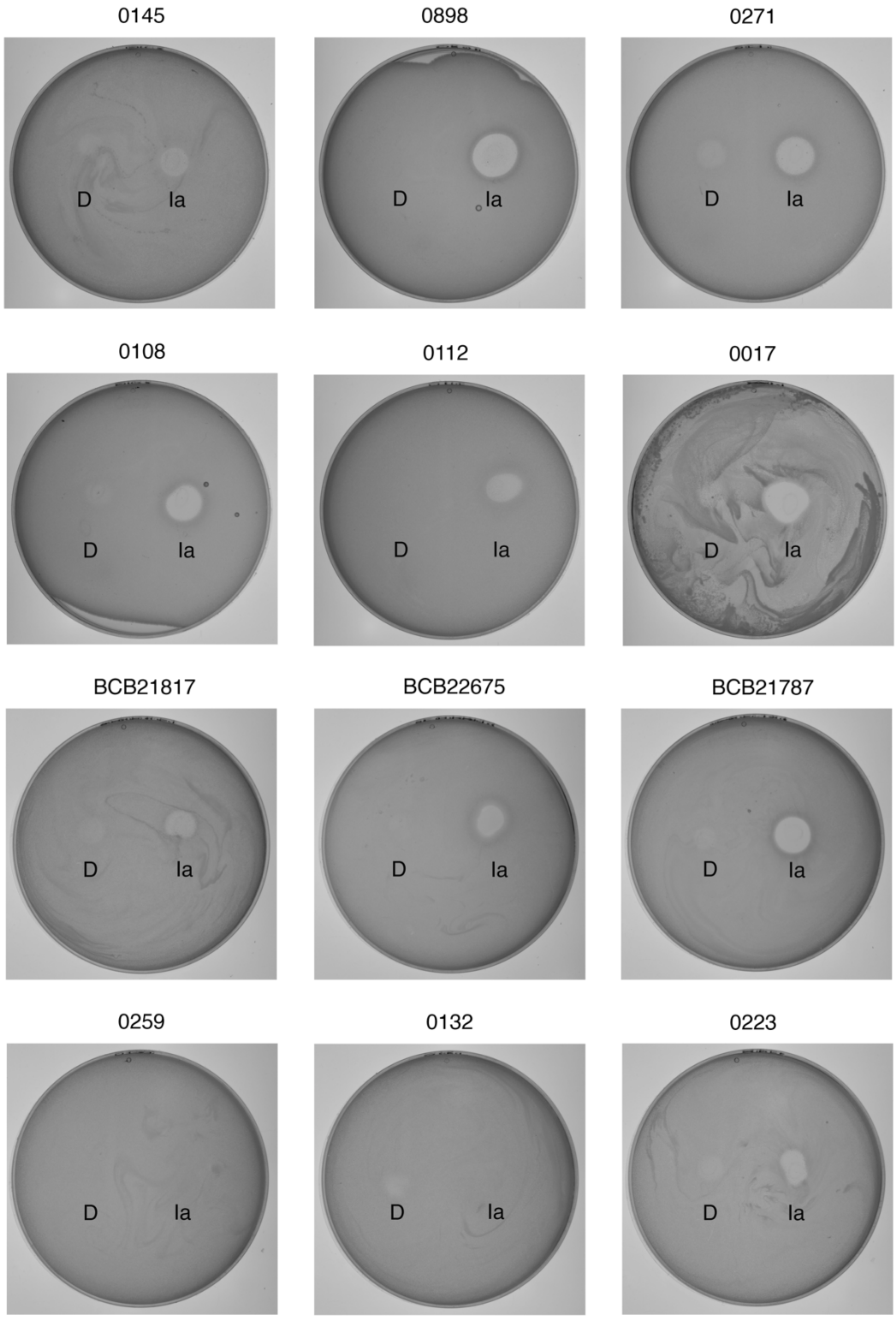


Fig 3-14 Colicin D and colicin Ia killing activity against clinical aEPEC isolates. Solid medium killing assays of clinical strains exposed to purified colicin D and colicin Ia. The circular zones of inhibition represent the degree of colicin killing activity.

Inspection of the plates revealed that colicin Ia caused inhibition in 10 out of the 12 strains tested, with large circular zones of inhibition clearly visible on each plate. Whereas, in comparison, colicin D demonstrated killing activity in 8 out of the 12 strains. However, the circular zones of inhibition were small and faint, suggesting that the killing activity was weaker compared to colicin Ia. Interestingly, this observation contrasted with the killing activity of colicin D and colicin Ia in the atypical prototype strain, E110019, where conversely, stronger killing activity was associated with colicin D versus colicin Ia (Fig 3-7B). There was only one isolate, strain 0259, where no killing activity for colicin D and colicin Ia was observed, interestingly this strain was also resistant to chloramphenicol and kanamycin and could therefore suggest that this strain possesses multiple resistance mechanisms.

Altogether, these results have described a simple model for isolation and characterisation of EPEC using ECC HiCrome selective media and a single-step multiplex PCR, and they have proved to be effective in characterising both prototype and clinical EPEC strains. In addition, the multiplex PCR demonstrated strain specificity and enabled EPEC strains to be characterised and accurately distinguished from other DEC pathotypes. However, phenotypic studies performed with clinical aEPEC isolates revealed the extent of variation amongst these strains, compared with the prototype lab EPEC strains. Thus, future investigations into virulence determinants and the physiopathology of A/E lesions in EPEC should include clinical isolates, to gain a better understanding of pathogenesis in these strains.

3.3 Discussion

3.3.1 Enteropathogenic *E. coli* are an important concern for public health

The World Health Organisation (WHO) has estimated that around 785 million people have access to basic drinking water services which are often contaminated with infectious microorganisms (UNICEF and WHO, 2019). As a consequence of this situation, it is estimated that diarrhoea causes around 526,000 deaths in children under five years of age each year (UNICEF, 2016; Bolukaoto *et al.*, 2021). *E. coli* is frequently identified as the contaminating organism, with EPEC being the most dominant aetiological agent of infant diarrhoea (Zhou *et al.*, 2018). In LICs, this contamination often occurs when untreated sewage containing pathogenic *E. coli* is released into the environment where it can persist for considerable periods of time and disseminate throughout communities. Interestingly, in regions where gastrointestinal pathogens are endemic, it has been proposed that repeated exposure to these organisms results in asymptomatic carriage of enteric pathogens in certain individuals (Kotloff *et al.*, 1995; Hu and Torres, 2015). Therefore, it is important that active surveillance on the prevalence of diarrhoeagenic *E. coli* is maintained, particularly for the emergence of highly virulent strains.

Epidemiological studies investigating persistent diarrhoea in children have reported that EPEC is the major cause (Abba *et al.*, 2009), with atypical EPEC strains associated with outbreaks in LICs and high-income countries (HICs), including Australia and Norway. Interestingly, in both these case studies, the pathophysiological effects such as the duration of diarrhoea was found to be longer, however, the symptoms associated with the diarrhoea were less severe, with no vomiting or abdominal pain (Ochoa *et al.*, 2008; Hu and Torres,

2015). Indeed, the pathogenesis of aEPEC appears to be more complex than that of tEPEC, and it has been proposed that the heterogeneity observed in their virulence traits could be related to the serotypes of aEPEC. There are several examples of studies in which different aEPEC strains have demonstrated extremely variable levels of virulence, including the case of aEPEC O128:H2, in which the strain was administered to healthy volunteers, none of whom developed any illness (Levine *et al.*, 1978). Conversely, another aEPEC strain O127a:K63 was isolated from an outbreak in China where 112 adults developed food poisoning; furthermore the strain was highly resistant to antibiotics (Hao *et al.*, 2012). Interestingly, this chapter has described similar observations which were made in the 12 aEPEC strains isolated from Vietnam. For example, aEPEC BCB22675 and BCB21827 adhered to HeLa cells and displayed characteristic EPEC adherence patterns, BCB22675 adhered diffusely to cells, whereas the formation of A/E lesions was apparent in HeLa cells infected with BCB21827. Strikingly, these strains were isolated from healthy individuals, whereas strains isolated from patients with gastroenteritis were less adherent to HeLa cells. However, the strains isolated from symptomatic individuals indicated higher antibiotic resistance with 4 out of 9 of these strains resistant to chloramphenicol. Alarmingly, aEPEC strain 0259 was resistant to both chloramphenicol and kanamycin, and it was not susceptible to killing activity by either colicin D or colicin Ia, suggesting that resolving an infection with this strain could be very difficult. Indeed, this chapter has discussed only the basic characterisation of these Vietnam isolates, however, it is apparent that there are phenotypic and genotypic variations which warrant further investigations.

3.3.2 Current technologies and future directions for the identification and classification of Enteropathogenic *E. coli*

Detection of DEC strains is complex, as classical microbiology techniques cannot be used to differentiate between pathotypes. Therefore, research has focused on the development of molecular based assays to separate pathovars based on strain specific virulence determinants. Moreover, as many DEC have become endemic in LICs, more sensitive diagnostics are necessary to increase the detectability from both symptomatic and asymptomatic individuals. Recently, Barletta *et al*., 2011, described a highly sensitive RT-qPCR method for the detection and quantification of EPEC, and demonstrated a threshold limit of 5 bacteria per mg of stool using this method. The study also revealed that children with diarrhoea presented with a significantly higher bacterial load compared with healthy individuals, although, there was no correlation between bacterial load and the duration of the diarrhoeal episode (Barletta *et al.*, 2011). However, the identification method described by Barletta *et al*., 2011, was limited because the RT-qPCR was restricted to the detection of a single DEC and did not allow for the identification of other *E. coli* enteric pathovars. Conversely, this chapter describes the design of a simple single-step multiplex PCR for the simultaneous detection of five DEC pathotypes using conventional PCR. Although this method was not suitable for quantifying the bacterial load, Fig 3-4 demonstrated the high specificity and sensitivity of the assay, with strain specific genes detectable at the lowest concentration of *E. coli.* The one-step multiplex PCR method provides a cost-effective diagnostic that requires access to inexpensive laboratory equipment and enables DEC pathotypes to be accurately distinguished within 48 hours. This protocol could be modified to increase the efficiency by directly extracting DNA from stool specimens using a cetyltrimethylammoniumbromide extraction method, as described by Cleary and colleagues. This would avoid the need for culturing techniques and would rapidly increase

the rate of diagnosis and aid infection control efforts (Barletta *et al.*, 2011). Indeed, the use of this assay could be particularly beneficial as a rapid diagnostic test for diarrhoeal illness in LICs.

Co-infections with more than one DEC, or other enteric pathogens are often significantly associated with diarrhoeal illness in LICs (Zhang *et al.*, 2016; Shrivastava *et al.*, 2017). However, a recent study investigating the proportion of diarrhoea cases associated with DEC in Vietnam revealed that co-infection with one or more DEC pathotypes was common in both children with diarrhoea and healthy controls (Duong *et al.*, 2020). Therefore, nucleic acid amplification tests that can detect multiple pathogens are more appropriate in clinical laboratories due to their increased specificity and sensitivity in symptomatic and asymptomatic individuals. Another emerging concern is the recent surge in reports detailing hybrid DEC pathotypes, the first of which was reported in 2011 during an outbreak of diarrhoea in Germany. The isolates recovered displayed phenotypes that were characteristic of EHEC and EAEC, including production of Shiga toxin and displaying an aggregative adherence pattern to epithelial cells. Therefore, the *E. coli* O104:H4 strain was designated as an EAEC/EHEC or EAHEC hybrid (Bielaszewska *et al.*, 2011; Bolukaoto *et al.*, 2021) Indeed, hybrid strains harbour a combination of virulence genes, potentially acquired due to horizontal gene transfer among diarrheagenic groups or through transmission of virulence genes encoded on plasmids (Santos *et al.*, 2020; Yang *et al.*, 2020). The growing occurrence of these hybrid strains demonstrates the potential for enteric pathogens to emerge with blended virulence profiles, and presents a serious public health concern with extreme consequences for those that become infected.

Antimicrobial resistance is also becoming a major public health concern. The number of emerging DEC strains with resistance patterns spreading across penicillins, cephalosporins and aminoglycosides has alarmingly increased throughout the globe (Guerra *et al.*, 2006;

Langendorf *et al.*, 2015; Konaté *et al.*, 2017). Therefore, there is an urgent need to identify novel therapeutics to control EPEC infection. One therapy currently under investigation is the use of bacteriocins, which are antimicrobial peptides produced during times of stress by bacteria such as *E. coli,* to selectively eliminate competing bacterial strains (Reeves, 1972). Bacteriocins released by *E. coli* are termed colicins, and they can be separated into distinct groups, A or B, according to whether they translocate the periplasmic space via the Tol or Ton systems, respectively (Law *et al.*, 2003). The mode of toxic action for most colicins can be divided into three categories, including pore-forming, DNase and Rnase modes (Chang *et al.*, 2018). Colicin D, which is known to cleave tRNA in targeted cells and colicin Ia which exerts its bactericidal activity in cells by forming pores in the cytoplasmic membrane (de Zamaroczy *et al.*, 2001; Jakes and Finkelstein, 2010), were tested for their toxicity against EPEC. Interestingly, colicin Ia was identified as the most toxic, with killing activity apparent in 10 out of the 12 clinical aEPEC strains tested (Fig 3-14). Colicin Ia also inhibited the growth of tEPEC strain E2348/69 (Fig 3-7), suggesting that this colicin could be used to target both EPEC subtypes. Recently, the antimicrobial properties of colicins have made these proteins an attractive alternative therapeutic for human and veterinary medicine. Indeed, Cutler *et al.*, 2007, demonstrated that the dietary inclusion of colicin E1 decreased the incidence and severity of ETEC infections in pigs (Cutler *et al.*, 2007) and similarly, the antagonistic activities of bacteriocins against *E. coli* have been demonstrated in poultry (Ogunbanwo, Sanni and Onilude, 2004). However, although bacteriocins are a promising therapeutic, much research into the safety and efficacy of purified bacteriocins is required before these can be considered as alternatives to conventional antibiotics.

To conclude, this chapter has confirmed the importance of characterising EPEC strains genotypically and phenotypically. Furthermore, the analysis of clinical EPEC strains has highlighted that understanding EPEC pathogenesis should not be limited to observations

made in prototype *E. coli* strains, as the genomic plasticity of this species has led to the emergence of strains with a blended virulence profile. Consistent with other research, this chapter also concludes that nucleic acid amplification methods should detect antibiotic resistance and virulence genes for a range of enteric pathogens, including DEC pathotypes, to better inform clinicians and researchers on resistant strains and trends in co-infecting organisms. Together, this will elucidate the pathogenesis of EPEC infection and enable better surveillance for the emergence of highly virulent strains.

As explained in the introduction, the project set out to design and implement novel methods for the screening and clearance of contaminating EPEC strains in healthy stool specimens. The techniques were developed and optimised using prototype EPEC strains, however, no clinical samples were ever received from the industrial partner and therefore, it was not possible to test the effectiveness of these methods on the intended samples. Subsequently, the industrial partner ceased all contacted and was unable to support this project. Therefore, the research in the following chapters reflects a change in direction that was essential to ensure the completion of my thesis.

4 Investigating the role of D-Serine in the downregulation of colibactin expression

4.1 Introduction

Composition of the human diet has been identified as a key factor in governing intestinal homeostasis, as alterations to this delicate balance have been shown to contribute to the development of chronic illnesses, immune system dysfunction and the development of cancer (Rooks and Garrett, 2016; Yang *et al.*, 2018; Francescangeli, de Angelis and Zeuner, 2019). Notably, a Western-style diet, which consists of higher levels of sugar and fat, has been associated with inducing pathophysiological changes in the gut that can contribute to an increased risk of developing CRC (Francescangeli, de Angelis and Zeuner, 2019). In addition, as well as inducing physiological changes to the intestine, the diet can modulate the composition of the microbiota, causing adverse effects to human health. Indeed, *E. coli* belonging to the B2 phylogroup have been isolated from biopsy specimens of CRC patients and were found to be the dominant colonisers (Raisch *et al.*, 2014). Furthermore, these strains have been implicated in the formation of colon tumours, as they harbour the polyketide synthase (*pks*) locus responsible for the biosynthesis of a potent genotoxic compound referred to as colibactin (Buc *et al.*, 2013).

Colibactin was first described by Nougayrede *et al*, in 2006, where *E. coli* encoding the *pks* island was implicated with inflicting a genotoxic insult on eukaryotic cells (Nougayrede *et al.*, 2006). The *pks* genomic island is a 54 kb hybrid non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) biosynthetic cluster, consisting of 19 genes. These genes encode the machinery for biosynthesis and secretion of the polyketide, colibactin (Putze *et al.*, 2009). The assembly line comprises three NRPS enzymes encoded by *clbH*, *clbJ* and *clbN*; three PKS enzymes encoded by *clbC*, *clbI* and *clbO* and two hybrid NRPS/PKS megasynthases, encoded by *clbB* and *clbK* (Nougayrede *et al.*, 2006). The schematic in Fig 4-1 illustrates the biosynthesis of colibactin using a prodrug resistance mechanism. A

prodrug scaffold (pre-colibactin) is assembled and is translocated into the periplasm by ClbM, a multidrug and toxic compound extrusion (MATE) inner-membrane transporter (Mousa *et al.*, 2016). Once inside the periplasm, the peptidase ClbP cleaves the N-acyl-D-asparagine motif of pre-colibactin to release the active colibactin molecule (Brotherton and Balskus, 2013). Furthermore, the gene product of *clbS* confers resistance to colibactin toxicity in the host bacteria by encoding a hydrolase that converts active colibactin into an innocuous product (Tripathi *et al.*, 2017). Upon secretion into the environment, active colibactin can induce cytotoxicity in eukaryotic cells (Nougayrede *et al.*, 2006; Cuevas-Ramos *et al.*, 2010).

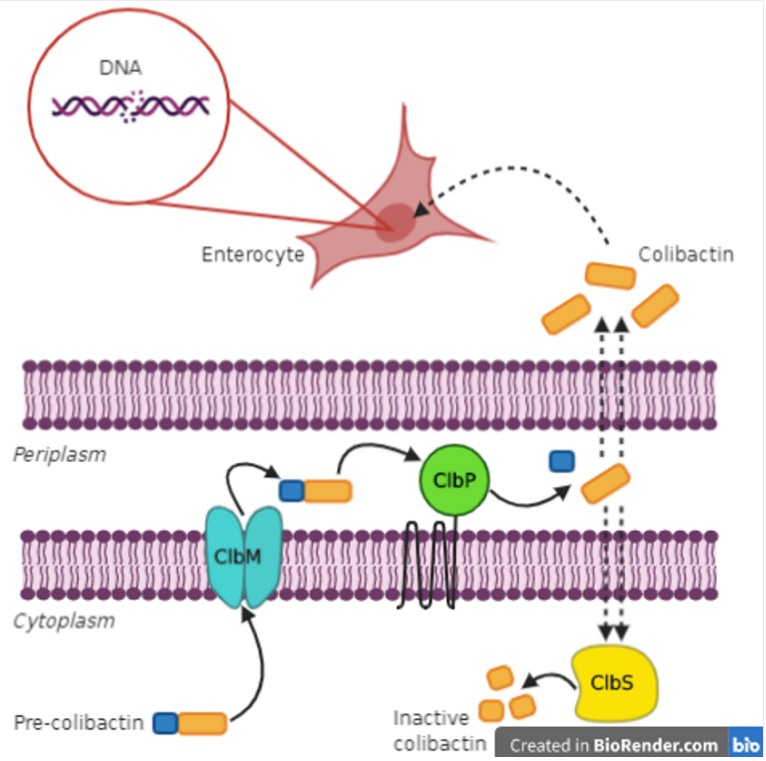


Fig 4-1 Schematic detailing colibactin maturation and transportation in the producing bacteria. Pre-colibactin is assembled in the cytoplasm before it is translocated into the periplasm via the inner membrane transporter, ClbM. Upon entering the periplasm, the prodrug motif is cleaved by ClbP and the active colibactin molecule is released into the environment. Colibactin initiates a genotoxic assault and induces DNA damage in

eukaryotic cells. However, the precise mechanism that facilitates entry of colibactin in to host cells remains unknown. Colibactin producing bacteria encode a hydrolase, ClbS, that converts active colibactin into a harmless derivative and provides the producing cell with immunity. Figure adapted from (Faïs *et al.*, 2018).

Expression analysis of the *pks* gene cluster revealed that the *pks* locus could be divided into at least seven transcriptional units, of which four were found to be transcribed polycistronically (Homburg *et al.*, 2007). Co-transcription of *clbR*/*clbA* indicated a potential regulatory function exerted by the *clbR* encoded LuxR-like regulatory protein on *clbA*. Indeed, *clbA* encodes a phosphopantetheinyl transferase that is responsible for the post-translational activation of the PKS and NRPS proteins of the colibactin biosynthesis pathway. Therefore, tight control over *clbA* expression is crucial for colibactin production (Homburg *et al.*, 2007). Recently, ClbR was identified as the key transcriptional activator of the *pks* locus and expression of the *clbR* gene was found to directly correlate with the function and production of colibactin in *pks*$^{+}$ *E. coli* strain M1/5 (Wallenstein *et al.*, 2020). Furthermore, it has been reported that different carbon sources and cultivation methods can influence transcript levels of *clbR* and *clbA.* Interestingly, Homburg *et al*, revealed that exposure to glucose and glycerol increased transcript levels of *clbR* and *clbA* compared with pyruvate and acetate. In addition, the group reported that overall increased transcription was observed throughout the whole cluster in shaking cultures, compared with growth in static conditions (Homburg *et al.*, 2007). However, the current understanding on the regulation of the colibactin gene locus remains limited. Therefore, investigations to identify potential inducers and repressors of colibactin are needed.

Upon cocultivation of *pks*$^{+}$ *E. coli* with eukaryotic cells, colibactin alkylates host DNA through two distinct cyclopropane warheads, causing DNA double-strand breaks (DSBs) and interstrand cross-links (Bossuet-Greif *et al.*, 2018). Furthermore, exposure to

colibactin has been shown to induce cellular arrest at the G_2 phase of the cell cycle, leading to activation of the DNA repair signalling cascade (Buc *et al.*, 2013). The cytopathic effect exerted by colibactin leads to progressive cell enlargement, a phenomenon referred to as megalocytosis, and eventually invokes cell death (Nougayrede *et al.*, 2006). The effects of colibactin are comparable to those observed with other *E. coli* cyclomodulins; the name ascribed to the family of bacterial toxins and effectors that interfere with the host cell cycle, as reviewed by Oswald *et al*, 2005. Conversely, unlike cytolethal distending toxin (Deng and Hansen, 2003) and cycle-inhibiting factor (Marchès *et al.*, 2003), the delivery mechanism of colibactin into the eukaryotic cell remains elusive. However, the cytopathic activity of colibactin has been described as "contact-dependent", as cytotoxicity is observed only when bacteria harbouring the *pks* island have direct contact with mammalian cells (Nougayrede *et al.*, 2006). Indeed, a recent breakthrough study revealed that exposure to colibactin causes a specific mutational

signature in adenine rich residues. This mutational signature was also identified in CRC-derived tissues, and it was confirmed that the mutation was specific to colibactin exposure (Dziubańska-Kusibab *et al.*, 2020b; Pleguezuelos-Manzano *et al.*, 2020). These data highlight that colibactin-producing *E. coli* represents an urgent public health matter.

This chapter will discuss how the host metabolite, D-Serine, was identified as a potent repressor of the genes encoded on the *pks* genomic island in UPEC and Nissle 1917. Furthermore, the effect of D-Serine on colibactin-associated cytotoxicity in mammalian cells will also be described.

4.2 Results

E. coli harbouring the *dsdCXA* locus possess the ability to metabolise D-Serine, a host metabolite that is present in higher concentrations at distinct sites of the human body, including the brain and bladder (Roesch *et al.*, 2003; Chen *et al.*, 2013). Certain pathogenic pathotypes of *E. coli* can utilise this amino acid as a sole carbon source, for example *dsdCXA*$^{+}$ NMEC and UPEC strains, which can metabolise D-Serine and establish infection at these nutrient deficient sites. As well as being a host metabolite, D-Serine has also been shown to selectively affect the expression of genes in *E. coli* pathotypes that do not possess the *dsdCXA* tolerance locus (Connolly *et al.*, 2015). However, little is known of the global effects on gene expression in strains that are found in D-Serine rich environments. Therefore, an investigation into the implications of D-Serine on gene expression in *E. coli* that encoded the complete *dsdCXA* locus, was prompted. An RNA-Seq experiment was performed by Dr Nicky O'Boyle to compare the effects of D-Serine on transcript levels throughout the genome of the UPEC strain CFT073, the detail of these results has been reported elsewhere (Connolly *et al.*, 2021).

4.2.1 Transcriptional response to D-Serine in the UPEC strain CFT073 revealed downregulation of *pks* encoded genes

Transcriptome analysis of CFT073 treated with D-Serine revealed 77 of 140 differentially expressed genes were downregulated throughout the genome, but of notable interest was the distinct downregulation of genes encoded on the colibactin biosynthesis locus (Connolly *et al.*, 2021). Read peaks are shown in Fig 4-2A from D-Serine treated (red) CFT073 compared with an untreated control (black), indicating that exposure to D-Serine resulted in a downshift in gene expression throughout the locus. The fold changes indicated

by the heat map in Fig 4-2B revealed that more than half of the genes were significantly downregulated. Of these, the genes with the most significant reductions were *clbK*, *I*, *H*, *G*, C and *B*, all of which are synthesis enzymes required to produce pre-colibactin, the precursor to cytotoxic colibactin (Zha *et al.*, 2016).

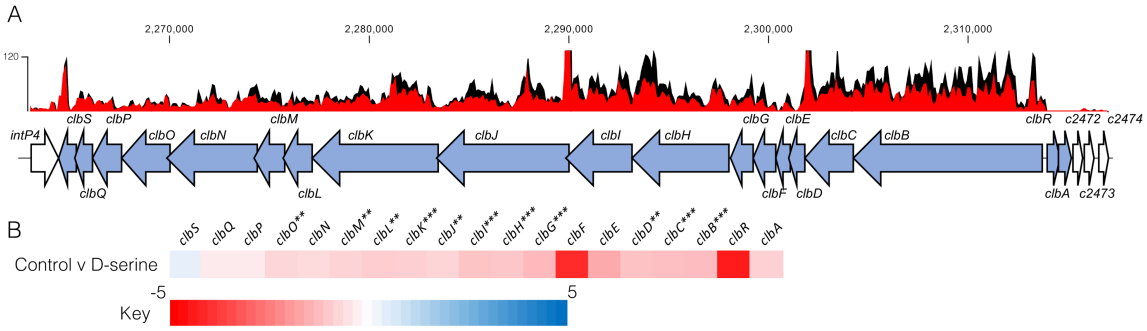


Fig 4-2 Transcriptome analysis of the colibactin biosynthesis operon in response to D-Serine in CFT073. Duplicate populations of CFT073 were cultured for 3 h before spiking one population with 1 mM D-Serine. Samples were taken for RNA extraction at 2 h post-addition of D-Serine and processed for RNA-Seq. (A) Read density in the colibactin biosynthesis locus for representative samples of the untreated control (black) and D-Serine treated (red) CFT073 is indicated. Read tracks were normalized to a maximum of 120 reads, exported from CLC Genomics Workbench and overlayed. Genomic coordinates are displayed above the read tracks and the corresponding genes within the colibactin biosynthesis operon (blue) beneath. (B) Heat map indicating the edgeR calculated $\log_2$ relative fold changes for each gene in the colibactin biosynthesis locus with corresponding colour key beneath. False discovery rate corrected *P* values are indicated with significantly differentially expressed genes with *, ** and *** indicating $P < 0.05$, 0.01 and 0.001, respectively. See Appendix 2 for the complete set of $\log_2$ relative fold changes and *P* values.

To validate the observations made by RNA-Seq, real-time polymerase chain reaction (RT-qPCR) was utilised to determine gene expression of *clbB*, selected as it encodes an NRPS/PKS megasynthase that is involved in the initiation of colibactin biosynthesis (Brotherton and Balskus, 2013). Relative expression of *clbB* was first determined for CFT073 using the same growth conditions as was used during the RNA-Seq experiment. CFT073 was grown in M9 minimal media for 5 h with the addition of D-Serine from the start or spiked in after 3 hours.

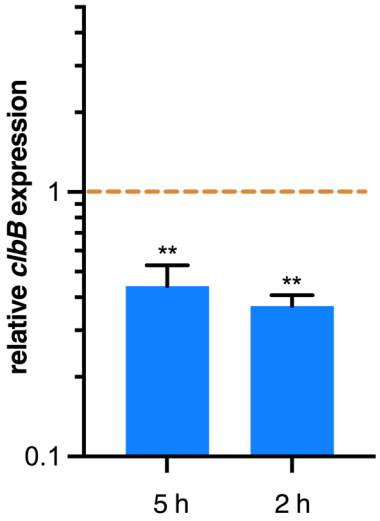


Fig 4-3 CFT073 relative *clbB* expression in response to D-Serine. Relative *clbB* expression was measured by RT-qPCR. CFT073 was grown in M9 minimal media and supplemented with 1 mM D-Serine. Cells were exposed to D-Serine either from the start (5 h) or media was spiked with D-Serine after 3 hours (2 h). The orange dashed line indicates baseline expression, measured from cells grown in the absence of D-Serine, with columns below this line representing downregulation. Statistical significance was determined from three biological replicates using a one-tailed Student's *t*-test, ** indicating significance, $P < 0.01$.

The expression values presented in Fig 4-3 indicate that *clbB* was significantly downregulated by 2.30 and 2.67-fold upon exposure to D-Serine for 5 h and 2 h

respectively. The fold changes observed by RT-qPCR matched those observed in the RNA-Seq, therefore confirming the initial findings that colibactin genes were downregulated in response to D-Serine in CFT073.

4.2.2 L- and D-amino acids can modulate expression of colibactin

To investigate whether repression of the colibactin locus was unique to D-Serine, a comprehensive selection of proteogenic L- and D-amino acids were tested in M9 minimal media and expression of *clbB* was measured by RT-qPCR.

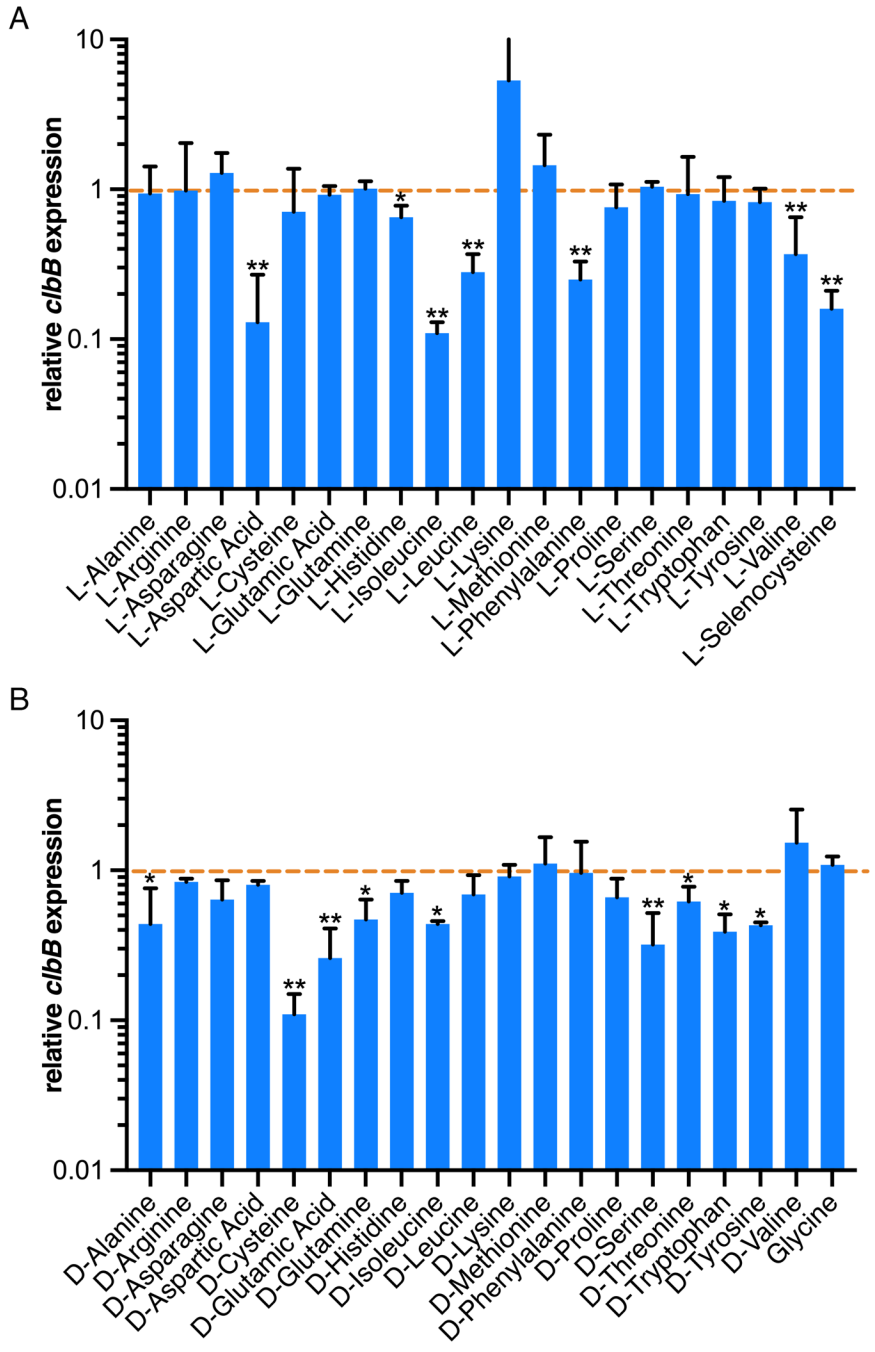


Fig 4-4 Expression of colibactin is modulated by L- and D-amino acids. Relative *clbB* expression was measured by RT-qPCR. CFT073 was grown in M9 minimal media supplemented with amino acids to a final concentration of 1mM for 5 h. (A) Shows relative expression in the presence of L-amino acids and (B) indicates expression in the presence of D-amino acids. The orange dashed line indicates baseline expression with bars above and below this line representing up and downregulation, respectively. Statistical significance was determined from three biological replicates using a one-tailed Student's *t*-test, * and

** indicating significance, $P < 0.05$ and 0.01, respectively. See Appendix 3 for full details on expression values, fold changes and *P* values.

The addition of L- and D-amino acids to the growth media revealed several interesting findings. Exposing CFT073 to L-amino acids revealed 7 of 20 amino acids tested induced repression of *clbB.* As illustrated in Fig 4-4A, L-Isoleucine demonstrated the most significant effect on gene expression with *clbB* reduced 8.84-fold ($P = 0.0071$). Amongst the D-amino acids tested, 9 of 20 significantly modulated expression of *clbB* (Fig 4-4B). The most consistent fold changes were observed for D-Cysteine and D-Serine, where gene expression was decreased 9.11 and 3.12-fold respectively ($P \leq 0.01$). However, growth was severely impeded in the presence of D-Cysteine and therefore the large fold change could have resulted from stress within the cells. Thus, D-Serine remained one of the most consistently potent repressors of colibactin expression in CFT073.

Next, the activity of D-Serine on *clbB* expression was tested in a different *pks* carrying *E. coli* strain to determine whether D-Serine associated inhibition of colibactin was restricted to CFT073. The pks^+ commensal isolate Nissle 1917, hereafter referred to as “Nissle”, was selected based on its alleged non-pathogenic nature and clinical significance as a prescribed probiotic (Olier *et al.*, 2012). Nissle is also a closely related strain to CFT073 and shares 99.9% homology in the *pks* gene cluster, therefore it was predicted that a similar effect on *clbB* expression would be induced upon exposure of Nissle to D-Serine.

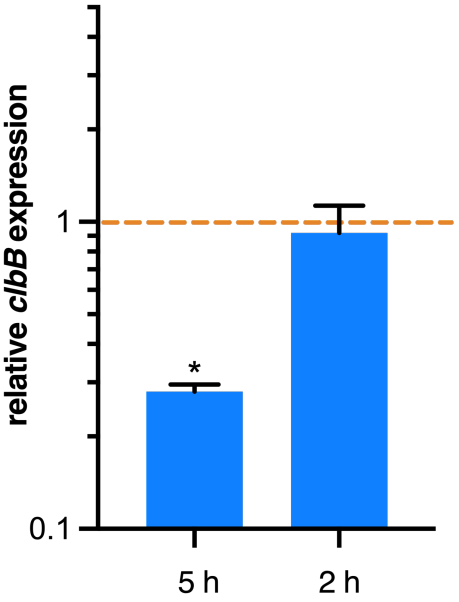


Fig 4-5 Exposure to D-Serine induces downregulation of *clbB* in Nissle. Relative *clbB* expression was measured by RT-qPCR. Nissle was grown in M9 minimal media and supplemented with 1 mM D-Serine. Cells were exposed to D-Serine either from the start (5 h) or media was spiked with D-Serine after 3 hours (2 h). The orange dashed line indicates baseline expression with columns below this line representing downregulation. Statistical significance was determined from three biological replicates using a one-tailed Student's *t*-test, * indicating significance, $P < 0.05$.

In a similar response to UPEC, Fig 4-5 illustrated that *clbB* was significantly repressed in Nissle treated with 1 mM D-Serine. Exposure to D-Serine for 5 hours induced a 3.62-fold reduction of *clbB*. However, *clbB* expression was not affected in cells that were exposed to D-Serine for only 3 hours. Interestingly, expression of *clbB* was lower in Nissle (0.28, $P = 0.035$) than expression in CFT073 (0.44, $P = 0.0038$) after both strains were exposed to D-Serine for 5 hours. Exposure to D-Serine had a strong modulatory effect on the expression of *clbB* in Nissle, indicating that D-Serine could regulate gene expression in both *pks*$^+$ *E. coli* strains. Next, a comprehensive selection of the amino acids tested in CFT073 were

retested in Nissle in order to investigate, if like D-Serine, the modulatory effect they exerted was the same in Nissle.

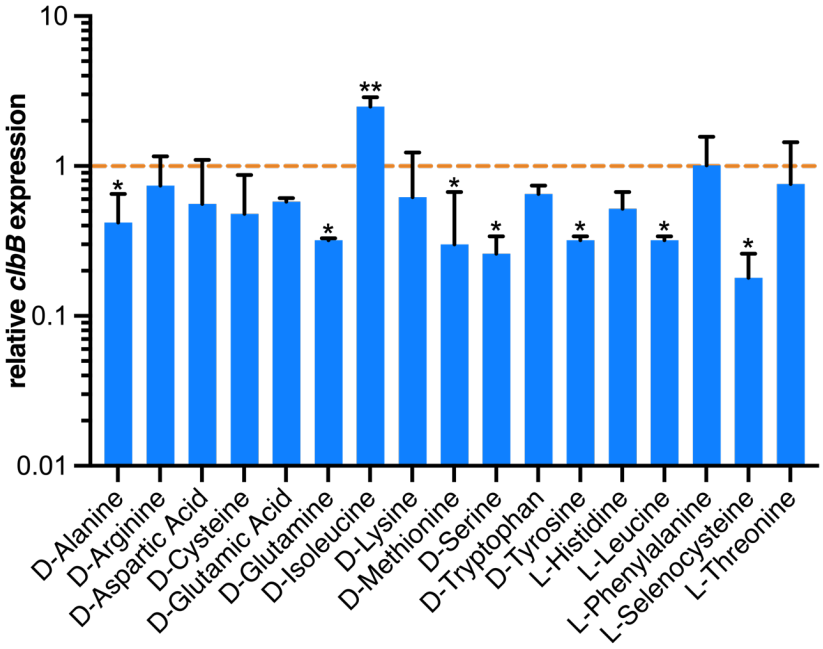


Fig 4-6 D- and L-amino acids induce modulation of colibactin expression in Nissle. Relative expression of *clbB* in the presence of L- and D-amino acids was measured by RT-qPCR. Nissle was grown in M9 minimal media supplemented with amino acids to a final concentration of 1 mM for 5 h. The orange dashed line indicates baseline expression with bars above and below this line representing up and downregulation, respectively. Statistical significance was determined from three biological replicates using a one-tailed Student's *t*-test, * and ** indicating significance, $P < 0.05$ and 0.01, respectively. See Appendix 4 for full details on expression values, fold changes and *P* values.

In response to the L- and D-amino acids tested, repression of the *clbB* gene in Nissle was observed for 7 of 17 amino acids which are highlighted in Fig 4-5. L-Selenocysteine elicited the strongest effect on *clbB* expression. However, like CFT073, growth was inhibited in the presence of L-Selenocysteine and the 5.64-fold decrease in gene expression could have been in response to the unfavourable growth conditions. The second most

significant decrease in *clbB* expression was observed for D-Serine where a 3.81-fold change was observed ($P = 0.037$). These data confirmed the important role both L- and D-amino acids have in modulating expression of colibactin, but notably D-Serine displayed one of the most profound effects in both *pks*$^+$ *E. coli* strains CFT073 and Nissle. It is also interesting to note that D-Serine has important physiological significance, as both these *pks*$^+$ *E. coli* will be exposed to fluctuating concentration of D-Serine in the mammalian host. UPEC strains can colonise both the gut and the bladder, however, in the bladder D-Serine concentrations are 1000-fold higher (Anfora *et al.*, 2007). Although strikingly, Oswald and colleagues recently revealed that UPEC strains harbouring the *pks* island were implicated with causing DNA damage in the bladder (Chagneau *et al.*, 2021).

4.2.3 HeLa cells infected with Nissle exposed to D-Serine do not elicit a cytopathic phenotype

Colibactin has been associated with causing cellular senescence in eukaryotic cells, whereby DNA damage leads to an irreversible state of cell-cycle arrest in cultured cells. This phenotype is characterised by progressive enlargement of the cell body and nucleus and the abolishment of mitosis; a phenomenon that has been termed megalocytosis (Nougayrede *et al.*, 2006). Therefore, to investigate the effects of D-Serine on megalocytosis, a cell culture assay was designed using *pks*$^+$ strains Nissle and DH10B pBAC-*pks.*

First, establishing the correct multiplicity of infection to elicit this effect was required for the desired test strains Nissle and DH10B pBAC-*pks*. These strains were chosen because Nissle has successfully been used in characterisation of the colibactin associated phenotype observed in HeLa cells (Massip *et al.*, 2019). Conversely, CFT073 was precluded from

such analysis as the cytopathic effects of colibactin are often confounded by haemolysin activity (Wiles and Mulvey, 2013), therefore, CFT073 was not a suitable strain for infections with eukaryotic cells. The *E. coli* strain DH10B pBAC-*pks* was gifted by Eric Oswald and has routinely been used as positive control for genotoxic activity.

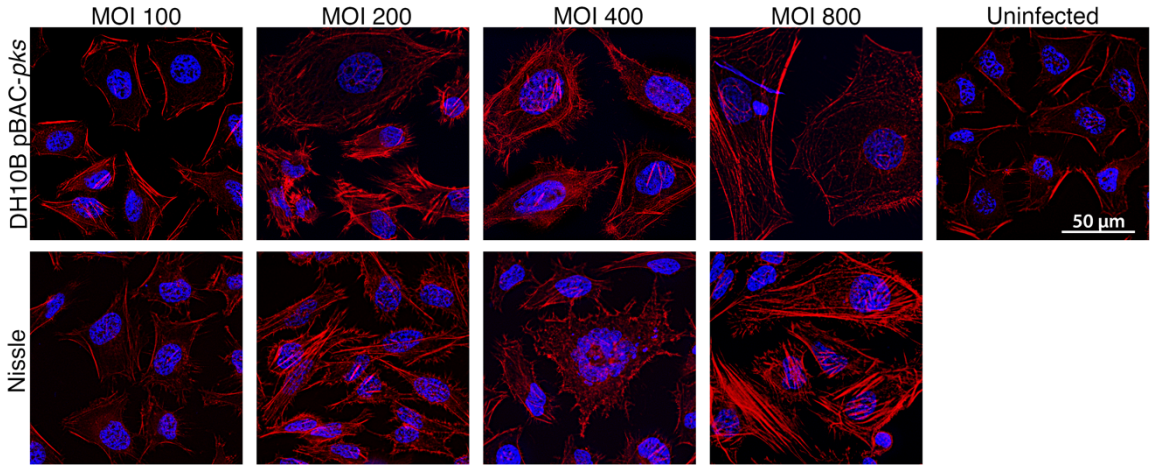


Fig 4-7 The cell enlargement phenotype is dependent on increased MOI. HeLa cells were infected with *pks*$^{+}$ *E. coli* strains at varying MOI concentrations for 4 h. Cells were treated with gentamicin and incubated for a further 4 h. Cells were further washed and incubated for 72 h to allow for the large cell phenotype to develop. HeLa cell morphology was observed by wide field fluorescence. Actin cytoskeleton was stained with Phalloidin in red and DNA was counterstained with DAPI in blue. Scale bar = 50 µm.

Cell morphology was visualised at 72 h post-infection using fluorescent microscopy. The images revealed an apparent dose dependent effect on the formation of the large cell phenotype (Fig 4-7). HeLa cells infected with an MOI = 100 displayed similar morphology to that of uninfected cells, whereas the large cell phenotype became more exaggerated as the dose increased to an MOI = 800. The dose dependent effect was true for infections performed with both Nissle and DH10B pBAC-*pks*. Infections using an MOI = 800 were too severe, and although megalocytosis was apparent, a large proportion of the cell

population succumbed to apoptosis. Therefore, the optimal dose was selected as an MOI = 400. At this concentration, cell death was avoided and the large cell phenotype was noticeably apparent during infections with both *pks*$^{+}$ strains.

4.2.4 D-Serine reduces the colibactin-associated cytopathic effect in eukaryotic cells infected with Nissle

To investigate the effects of D-serine on colibactin-associated megalocytosis during infection of HeLa cells with *pks*$^{+}$ *E. coli*, cells were exposed to 1 mM D-Serine. This concentration was selected on the basis of its physiological relevance within the host.

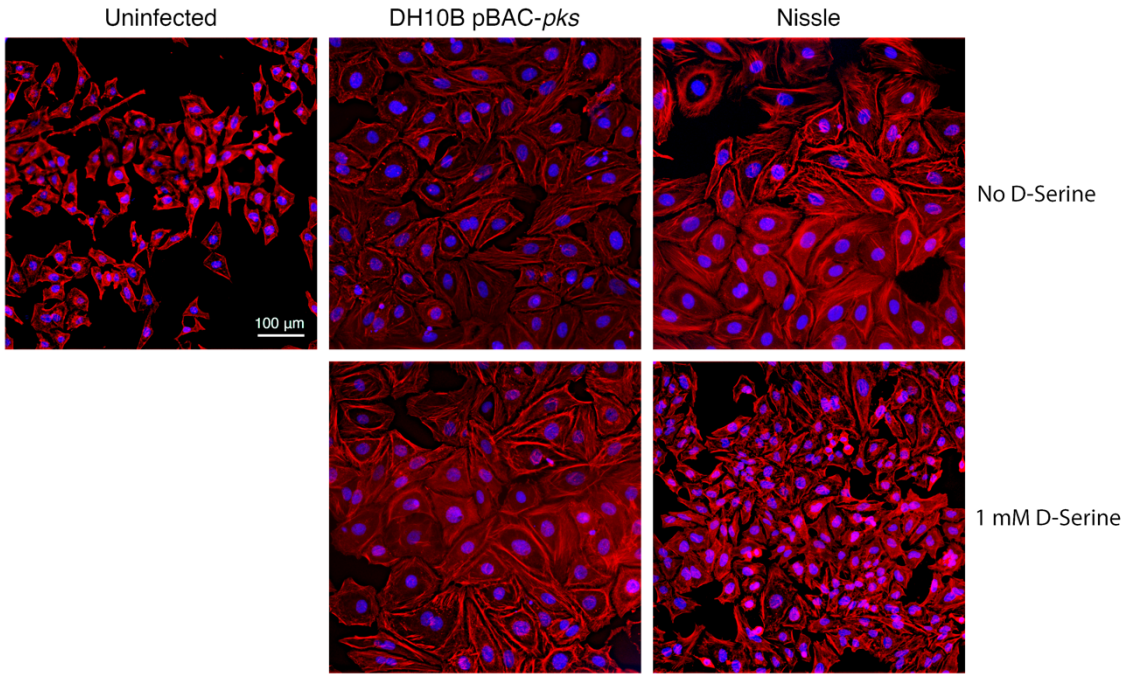


Fig 4-8 Exposure to D-Serine reduced colibactin-associated cellular senescence. HeLa cells were infected for 4 h with *E. coli* Nissle or DH10B hosting pBAC-*pks* (MOI = 400). Infections were performed with and without the addition of 1 mM D-Serine to the growth media. At 8 h after infection, cells were washed and incubated for 72 h to allow for the megacell phenotype to develop. HeLa cell morphology was observed by wide field

fluorescence. Actin cytoskeleton was stained with Phalloidin in red and DNA was counterstained with DAPI in blue at 72 h post infection.

HeLa cells infected with *pks*$^{+}$ *E. coli* displayed the characteristic colibactin-associated megacell phenotype at 72 h post-infection. However, upon exposure to 1 mM D-Serine, cells infected with Nissle displayed markedly reduced cellular senescence, and cell morphology was markedly like that of the uninfected control. Strikingly, as shown in Fig 4-8, D-Serine treatment did not result in decreased senescence-associated morphological alterations to cells infected with DH10B pBAC-*pks*; therefore, indicating that D-Serine associated repression of colibactin requires native expression of the *pks* locus.

Progressive enlargement of the cell body and nucleus are the defining features of colibactin induced megalocytosis. Therefore, to quantify the cytopathic effect observed by infecting HeLa cells with Nissle, 10X images were acquired by fluorescent microscopy 72 h post-infection. Cell area was calculated from single channel image files using the image processing software, CellProfiler (McQuin *et al.*, 2018).

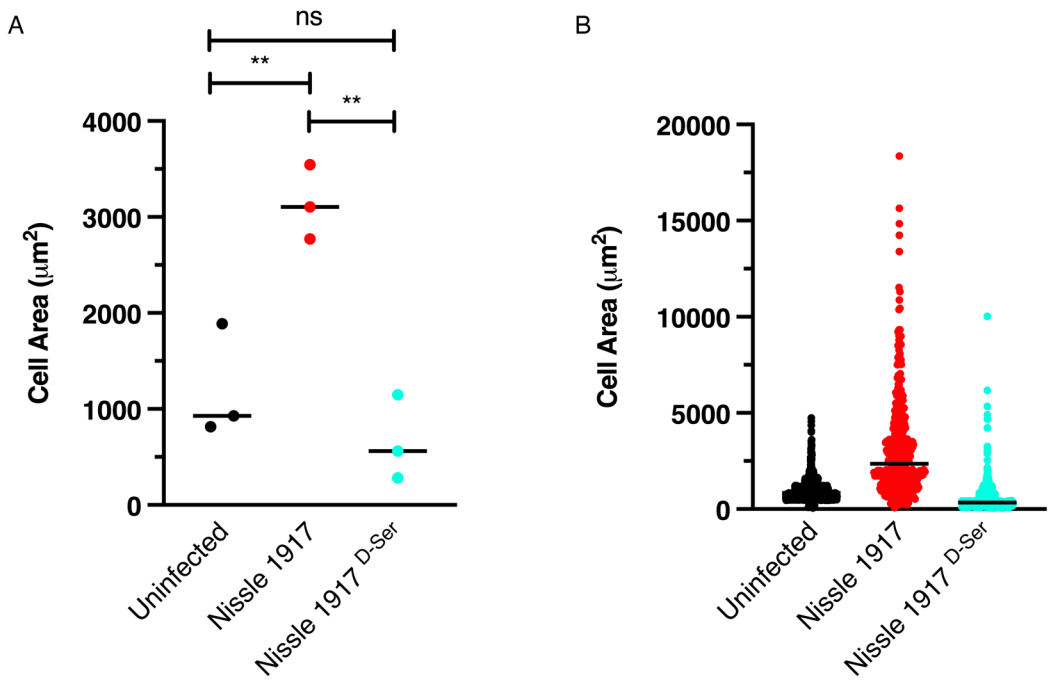


Fig 4-9 Cell area is reduced in HeLa cells infected with Nissle bacteria treated with D-Serine. The area of HeLa cells at 72 h post-infection was determined by acquiring images at 10X magnification and using CellProfiler software to measure 100 cells per image. (A) Columns represent the mean cell area measured with individual experimental observations indicated by data points for each infection condition. Measurements were acquired from images taken from three independent experiments and statistical significance was assessed by unpaired Student's *t*-test with, ** indicating $P < 0.01$. (B) Individual cell area measurements were recorded across triplicate experiments. Black lines indicate the mean.

Fig 4-9A shows the mean cell area calculated for three independent experiments. In consequence of infecting HeLa cells with Nissle, cell enlargement increased 2.60-fold. However, upon the addition of D-Serine, the mean cell area was significantly decreased ($P < 0.01$) in cell populations that were infected with treated Nissle. Indeed, there was no significant difference between cells infected with D-Serine treated Nissle compared to the uninfected control. Fig 4-9B represents individual cell area measurements and highlights the variable extent of cell enlargement observed across the three replicates, with one of the

largest cell areas recorded as 15000 μm^2 approximately 10 times larger than the area of an average HeLa cell. Furthermore, the results in Fig 4-8 and 4-9 show that exposing Nissle to D-Serine provides infected cells with long-term protection against genotoxic assault as there is little evidence of cell enlargement and the average cell area recorded at 72 h post infection was approximately 1500 μm^2. Overall, these data confirm that treatment with D-Serine reduces the megacell phenotype in HeLa cells infected with Nissle and provides cells with protection against colibactin associated cellular senescence.

4.2.5 Colibactin expression is downregulated during infection with host cells

To assess the effects of D-Serine on the expression of colibactin genes during transient infection with HeLa cells, RNA was extracted first from the inoculum, then from infected tissues at 2 h and 4 h post-infection. Bacterial RNA was purified and transcript levels of *clbB* were measured using RT-qPCR.

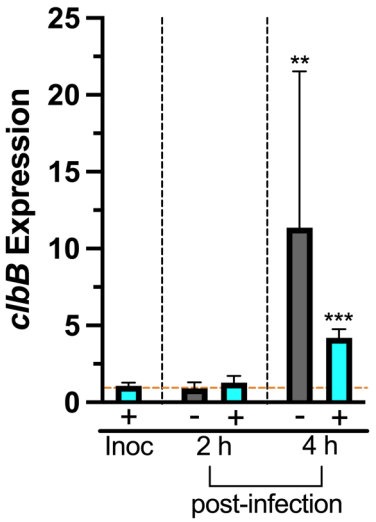


Fig 4-10 D-Serine induces modulation of colibactin expression during transient infection with HeLa cells. Relative expression of *clbB* in the presence or absence of D-

Serine was measured by RT-qPCR. HeLa cells were infected with Nissle in MEM-HEPES media alone (-) or supplemented with 1 mM D-Serine (+) for 4 h. RNA was extracted at 2 h and 4 h to compare transcript levels during infection. The orange dashed line indicates baseline expression with bars above and below this line representing up and downregulation respectively. Statistical significance was determined from three biological replicates using a one-way ANOVA, ** and *** indicating significance, $P < 0.01$ and 0.001, respectively.

Fig 4-10 shows *clbB* expression values for Nissle extracted from transient infections with HeLa cells over a 4 h infection period. Before infecting, Nissle was cultured in MEM-HEPES alone or in the presence of 1 mM D-Serine for 4.5 h, RNA was extracted at this time point and deemed the inoculum (Inoc). To establish an MOI = 400, cultures were standardised to an $OD_{600\ nm} = 0.1$ and HeLa cells were infected for a 4 h period in MEM-HEPES media alone (-) or media supplemented with 1 mM D-Serine (+). RNA was extracted from the midpoint (2 h) and the end point (4 h) of the infection period. Relative expression of *clbB* indicated that gene expression was significantly upregulated in untreated Nissle at 4 h post-infection when compared to the untreated inoculum; indeed, *clbB* expression increased a striking 11.37-fold ($P = 0.0039$). However, there was no significant difference between the inoculum and untreated Nissle extracted at 2 h. Upon exposure to D-Serine, there remained no significant difference in *clbB* expression in transcripts extracted at 2 h post infection. However, expression of *clbB* was significantly increased in D-Serine treated Nissle extracted at the 4 h time point compared to the inoculum ($P = 0.000021$). Notably, when comparing expression levels of treated and untreated Nissle at the 4 h time point, exposure to D-Serine reduced transcript levels by 2.71-fold. These data suggest that a minimum 4 h infection period may be necessary to ensure the upregulation of *clb* genes and the production of colibactin. Further, exposure to D-Serine induces downregulation of *clbB* expression, albeit transcript levels remained

higher than the inoculum, suggesting that colibactin production is not completely repressed under these conditions.

The results in Fig 4-10 indicated that *clbB* expression was significantly upregulated midway through the infection period, suggesting that colibactin production increased over time. Therefore, to investigate if expression of colibactin could be growth phase dependent, Nissle was cultured in the infection media and the effects of D-Serine on *clbB* transcript levels were measured throughout the growth period.

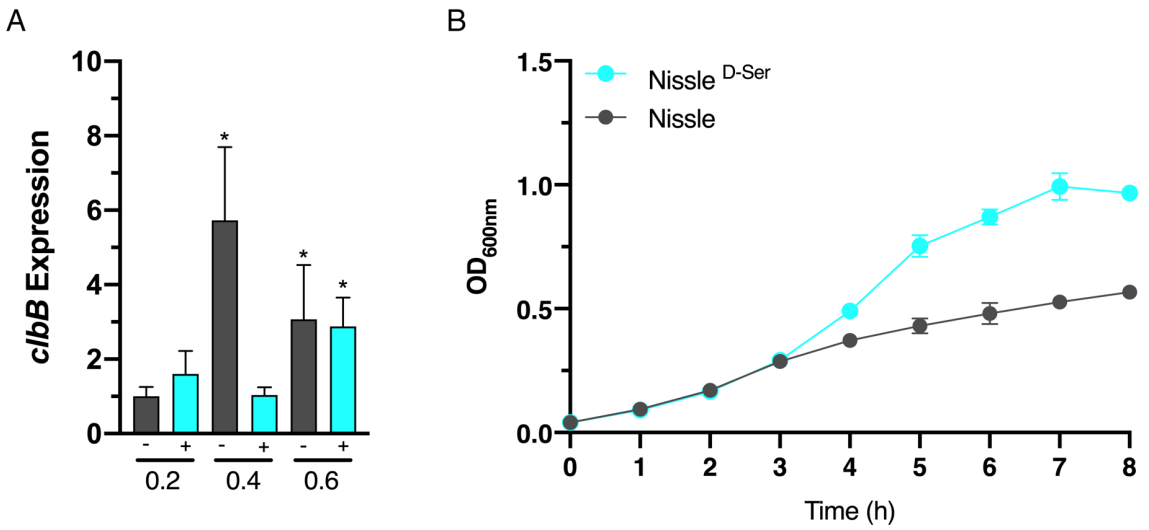


Fig 4-11 D-Serine modulates growth and affects gene expression during flask culture. Nissle was grown in MEM-HEPES in the presence (+) or absence (-) of D-Serine. RNA was extracted when cultures reached an $OD_{600\ nm}$ of 0.2, 0.4 and 0.6. (A) Relative expression of *clbB* was measured by RT-qPCR. Expression values were determined relative to $OD_{600\ nm}$ = 0.2 (without D-Serine) as the control. Statistical significance was determined from three biological replicates using a one-way ANOVA, * indicating significance, $P < 0.05$. (B) Growth curves were performed to compare the effects of D-Serine on growth.

Nissle was cultured in MEM-HEPES in the presence or absence of 1 mM D-Serine for 4.5 hours to allow cultures to reach mid-exponential phase before infecting HeLa cells. Transcript levels were measured throughout this period and expression of *clbB* was determined by RT-qPCR. Fig 4-11A shows that *clbB* expression significantly increased by 5.73-fold between untreated samples taken at an $OD_{600\ nm}$ = 0.2 and $OD_{600\ nm}$ = 0.4 (P = 0.042). This correlated with cells being in exponential growth phase as illustrated by the growth curve in Fig 4-11B. Interestingly, expression levels of *clbB* were relatively unchanged at an $OD_{600\ nm}$ = 0.2 between untreated and treated Nissle. However, when cells reached an $OD_{600\ nm}$ = 0.4 expression of *clbB* was 5.52-fold less between treated and untreated samples. Surprisingly, levels of *clbB* transcript decreased between an $OD_{600\ nm}$ = 0.4 and $OD_{600\ nm}$ = 0.6 in Nissle grown alone. However, growth media supplemented with D-Serine provided Nissle with an additional carbon source and resulted in enhanced growth, which became apparent at hour 4 (Fig 4-11B), perhaps explaining why *clbB* expression is similar between Nissle treated and untreated. Indeed, Nissle in the presence of D-Serine sampled at an $OD_{600\ nm}$ = 0.6 were in mid- exponential phase, whereas untreated Nissle sampled at the same OD were closer to late exponential phase. These data suggest expression of colibactin may be growth-phase dependent, and furthermore expression is highest when cells are in mid exponential growth.

4.2.6 Exposure to D-Serine reduces phosphorylation of histone H2AX

Transient infection of cultured epithelial cells with *E. coli* harbouring the *pks* island, can induce DNA double-strand breaks (DSBs) after a 4 h period (Nougayrede *et al.*, 2006). In response, the cell recruits the DNA damage response (DDR), and the ataxia telangiectasia mutated kinase (ATM) pathway is activated. The phosphorylation of histone H2AX (γ-

H2AX) is one of the first signals in response to DSBs, and can be detected 4 h after DNA damage has occurred (Stiff *et al.*, 2004). Thus, γ-H2AX is considered a marker for DNA DSBs in eukaryotic cells.

HeLa cells were infected with *pks*$^{+}$ *E. coli* for 4 h in the presence or absence of D-Serine. Next, cells were washed, and wells were replenished with media supplemented with gentamicin before incubating for a further 4 h. Cell lysates were extracted and proteins were separated by SDS PAGE. Detection of phosphorylated H2AX was determined by immunoblotting and normalised signal intensities from three experiments were compared to quantify levels of phosphorylation.

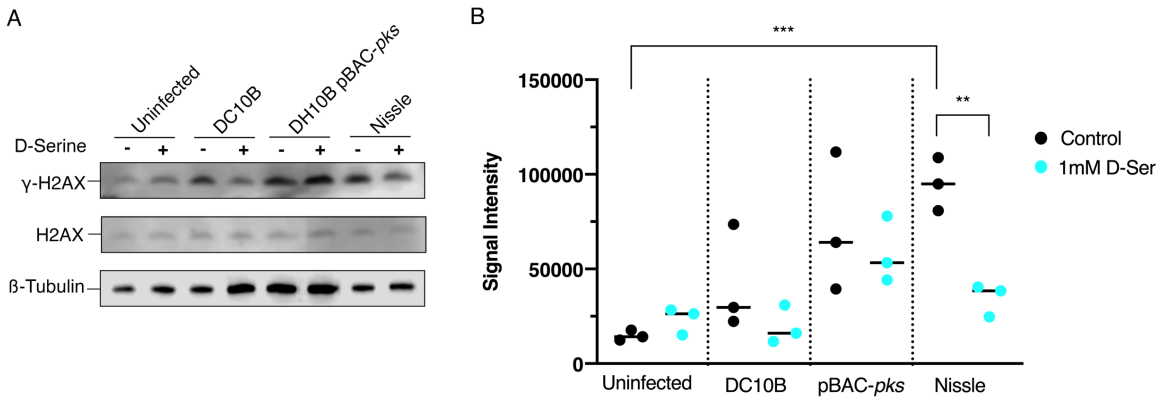


Fig 4-12 Repression of colibactin by D-Serine reduced DNA damage in HeLa cell infection. HeLa cells were infected for 4 h with *pks*$^{+}$ and *pks*$^{-}$ *E. coli* with an MOI = 400 bacteria per cell or left uninfected. Infections were performed in wells containing MEM-HEPES alone (-) or with media supplemented with 1mM D-Serine (+). (A) Immunoblot analysis of cell lysates extracted 4 h post infection. Phosphorylated histone (γ-H2AX) was employed as an indicator of double stranded DNA breaks and total histone (H2AX) was used as an internal control. β -Tubulin was used as a loading control. DH10B pBAC-*pks* and DC10B were used as *pks*$^{+}$ and *pks*$^{-}$ controls respectively. (B) Signal intensities of bands were measured using LI-COR Image Studio software. γ-H2AX signals were

corrected to account for any variation in loading using β-Tubulin signal intensity. Experimental signal was normalised so that the mean signal intensity of the eight samples was equivalent for each experiment. The experiment was carried out in triplicate. Columns represent mean +/- SEM with individual experimental observations indicated by data points. Statistical significance was assessed by unpaired Student's *t*-test with, ** and *** indicating $P < 0.01$ and 0.001, respectively.

Detection of γ-H2AX by immunoblotting revealed more phosphorylated proteins were present in cell lysates extracted from wells infected with untreated *pks*$^{+}$ *E. coli* strains, compared with the uninfected control. Fig 4-12A+B confirmed that levels of phosphorylated proteins remained relatively similar in cells infected with DH10B pBAC-*pks* in both the presence and absence of D-Serine, whereas the addition of D-Serine to the growth media during infection with Nissle resulted in a 4.42-fold reduction in phosphorylation compared with cells infected with untreated Nissle ($P = 0.0032$). These data suggest that the response to D-Serine may involve regulatory elements that are specific to natural colibactin-producing strains which extends beyond the circuitry of the pBAC-*pks* and could explain why the genotoxic activity was perturbed in treated Nissle but not in treated DH10B pBAC-*pks*. Signal intensities remained like the uninfected control in cells infected with the *pks*$^{-}$ strain DC10B (Fig 4-12A+B) and the addition of D-Serine did not significantly change levels of phosphorylation, therefore, confirming that D-Serine acted specifically on colibactin.

4.2.7 Nuclear foci are markedly reduced in HeLa cells infected with Nissle in the presence of D-Serine

The occurrence of a single DSB can lead to several hundred to thousand γ-H2AX proteins binding to the affected region of DNA (Rogakou *et al.*, 1999). The accumulation of these proteins appears as nuclear foci and can be visualised using a confocal microscope. HeLa cells were infected as described above and then stained with anti-γ H2AX antibody 8 h post-infection before visualising the cell nucleus.

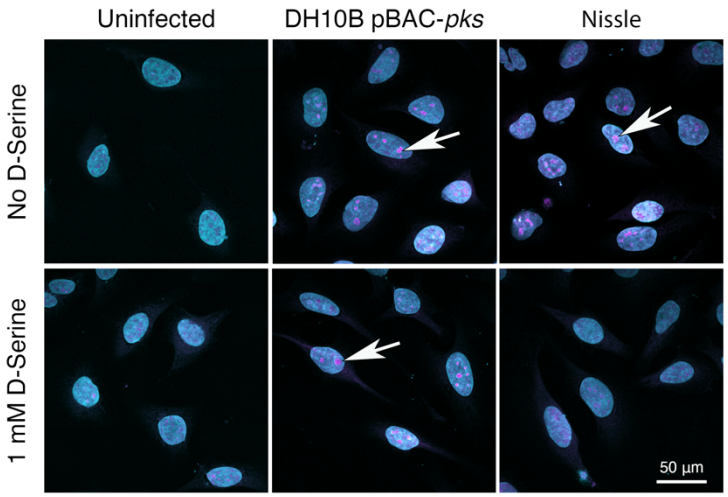


Fig 4-13 D-Serine reduced nuclear foci observed in HeLa cells. HeLa cells were infected for 4 h with *E. coli* Nissle or DH10B hosting BAC-*pks* (MOI=400). Infections were performed with and without the addition of 1 mM D-Serine to the growth media. At 8 h post infection, cells were washed, fixed and stained with anti-γ H2AX antibody. Cells were examined by confocal microscopy for DNA in cyan and phosphorylated histone H2AX protein in magenta. Images of uninfected, and *pks*$^{+}$ infected cells are shown, scale bar = 50 µm. White arrows highlight individual foci in the nucleus.

Distinct nuclear foci were detected in cells infected with strains harbouring the *pks* island. The white arrows in Fig 4-13 indicate individual foci and illustrate the assembly of γ-H2AX proteins at sites of DNA damage. Inclusion of D-Serine saw a marked reduction of foci in cells infected with treated Nissle and suggested that less DNA damage had occurred in these cells during cocultivation. Nuclei resembled uninfected cells and displayed little to no foci, whereas the number of foci remained unchanged in cells infected with DH10B pBAC-*pks* in both the presence and absence of D-Serine. Thus, exposure to D-Serine did not prevent the occurrence of DNA damage in cells infected by DH10B pBAC-*pks,* strengthening the hypothesis that the effects of D-Serine require factors that are specific to natural producing strains of colibactin.

Flow cytometry was utilised to measure the heterogeneity amongst cells exposed to Nissle treated or untreated with D-Serine. Stained cells were sorted using the BD FACS Aria and levels of γ-H2AX were detected.

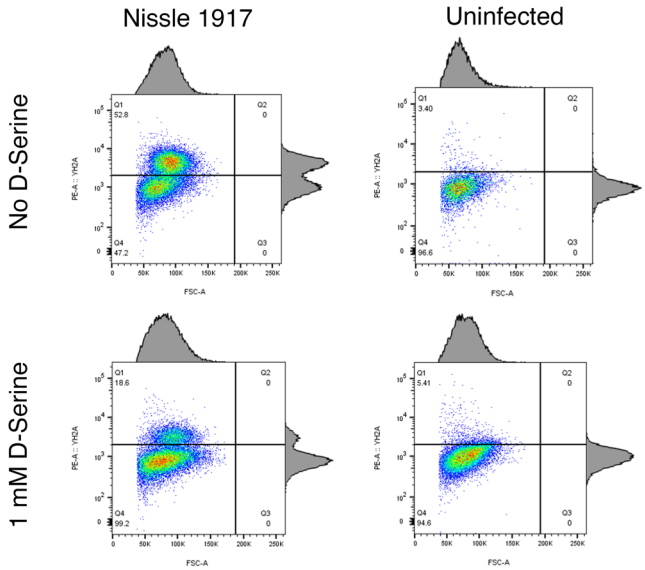


Fig 4-14 Treatment with D-Serine reduced phosphorylation of histone H2AX in HeLa cells. At 8 h post infection, cells were trypsinised and stained with anti-γ H2AX antibody

and intracellular levels of phosphorylated histone H2AX were measured by flow cytometry. Dot plots indicated the percentage of cells fluorescing in γ-H2AX channel, 100K events were analysed for each sample.

Single cell analysis was performed using FloJo Flow Cytometry software and fluorescence emission is presented in Fig 4-14. Detection of γ-H2AX revealed a 34.2% reduction in fluorescence emission between cells infected with Nissle exposed to D-Serine, compared with cells infected with Nissle alone. However, levels of γ-H2AX fluorescence did not fully return to that of the uninfected cells upon treatment with D-Serine. Taken together, these results indicate that exposure to D-Serine limits cytotoxicity in natively expressing *pks* strains, however, production of colibactin is not completely inhibited.

4.3 Discussion

4.3.1 L- and D- amino acids modulate expression of colibactin in *pks*$^{+}$ *E. coli* strains

Over recent years, there has been a growing appreciation for how the diet, and in particular, the role of amino acids, has on influencing bacterial gene expression. L-amino acids are indispensable because they are required for the biological and physiological functions of a myriad of proteins and enzymes in mammals. However, over the last few decades, D-amino acids have been increasingly recognised as natural biomolecules which play interesting and specific roles in both mammals and bacteria. D-amino acids can be found naturally in foods such as fruits and vegetables (Brückner and Westhauser, 2003). Interestingly, higher concentrations are more commonly associated with fermented and processed food; often owing to the activity of microorganisms during the fermentation

process (Marcone *et al.*, 2020). Free D-amino acids have also been found to exist in mammalian tissues. Indeed, D-Aspartic acid, D-Alanine and D-Serine, have been detected in the brain and it has been reported that they contribute to brain functionality (Hashimoto *et al.*, 1993; Panatier *et al.*, 2006; Kim *et al.*, 2010). Furthermore, due to its role in activating neurotransmitter receptors, D-Serine has been considered as a potential therapy for multiple neurological conditions, including post-traumatic stress disorder (Heresco-Levy *et al.*, 2009) and schizophrenia (MacKay *et al.*, 2019).

Amino acids are involved in important biological processes in mammals, however, there is a growing appreciation for the important regulatory role of certain amino acids in bacteria. Investigations into *E. coli* biofilm formation revealed the spatiotemporal regulation of L-Alanine metabolism is essential for cell viability and growth of colonies (Díaz-Pascual *et al.*, 2021) , whereas conversely, catalysis of the amino acid L-Tryptophan is implicated with the inhibition of biofilm formation (Shimazaki *et al.*, 2012). In the mammalian host, sensing these metabolites can serve as stimuli to trigger the expression of essential virulence genes. In response to the abundance of L-Arginine available in the gut, pathogenic EHEC strains signal the upregulation of LEE-encoded genes to facilitate site-specific colonisation of the host (Menezes-Garcia *et al.*, 2020). However, in contrast, exposure to D-Serine has been implicated in the downregulation of virulence genes in this *E. coli* pathotype, with our previous work demonstrating that D-Serine represses the type three secretion system (T3SS) (Connolly *et al.*, 2015). While our work has shown that D-Serine is present in trace concentrations in the gut (Connolly *et al.*, 2015), approximately 1000-fold lower than the concentration reported for the bladder (Anfora *et al.*, 2007), the production of D-Serine by members of the gut microbiome has been reported (Matsumoto *et al.*, 2018). As a result, the *E. coli* strains residing in the gut may encounter localised micro-niches rich in this metabolite, raising the possibility that it functions as a niche-specific regulator of diverse virulence genes in pathogenic *E. coli* (Connolly *et al.*, 2015).

This chapter discusses the comprehensive analysis of L- and D-amino acids and the regulatory effects exerted on *pks*-encoded *clbB*. The observations made in this chapter reveal that both L- and D-amino acids exert a modulatory effect on *clbB*, however, the regulatory effect exerted by D-amino acids was superior in *pks*$^{+}$ *E. coli* strains. D-Serine was identified as the most consistent repressor of *clbB* with expression downregulated 3.12- and 3.81-fold in CFT073 and Nissle respectively, suggesting there may be a conserved mechanism for *pks* regulation upon exposure to D-Serine. Interestingly, no modulatory effects were observed for L-Serine but repression of *clbB* was observed upon exposure to other L-amino acids. L-Isoleucine, L-Leucine and L-Valine are branched chain amino acids which share a common biosynthesis pathway and were each found to downregulate expression of *clbB* by 8.84, 3.61 and 2.61 respectively in CFT073. Serine is a precursor for the biosynthesis of Isoleucine, Leucine and Valine, suggesting that there may be multiple metabolic pathways involved in the response to amino acids. Moreover, the inhibitory effect was only mirrored in the D-enantiomer of Isoleucine where expression of *clbB* was downregulated 2.29-fold, but no changes were observed for D-Leucine and D-Valine in CFT073, supporting the growing evidence that L- and D-enantiomers of certain amino acids can possess distinct biological functions. Exposure to D-Isoleucine in Nissle saw expression of *clbB* upregulated by 2.49-fold, suggesting that D-Isoleucine could also have a strain specific regulatory response. Thus, these experiments have highlighted the modulatory effects exerted by amino acids and have highlighted their importance in regulating virulence in colibactin producing *E. coli.*

4.3.2 Exposure to D-Serine protects eukaryotic cells from colibactin-associated genotoxicity

In eukaryotic cells, colibactin-associated genotoxic activity can cause DNA damage and induce cell cycle arrest, triggering cellular senescence in affected cells. Therefore, to investigate the inhibitory effects of D-Serine further, colibactin-associated senescence was assessed using a cell culture infection model with *pks*$^{+}$ *E. coli.* In the absence of D-Serine, cells infected with Nissle presented with the characteristic cytopathic effect in which cellular and nuclear enlargement was evident (Fig 4-8). However, upon exposure to D-Serine, cells infected with treated Nissle displayed cell morphology that was markedly like that of the uninfected control (Fig 4-8). Detection of γ-H2AX by single cell analysis also revealed a 26% reduction in fluorescence emission between cells infected with D-Serine treated Nissle compared with cells infected with Nissle alone (Fig 4-14), although exposure to D-Serine did not fully return levels of phosphorylation to that of the uninfected control, the inhibiting effect exerted by D-Serine was apparent. Strikingly, this is the first study to report that exposure to D-Serine inhibits colibactin-associated genotoxicity.

There have been several other reports describing agents that can influence transcription and production of colibactin in *pks*$^{+}$ *E. coli*. Sadecki *et al*., 2021, identified that the chemical polymyxin B, a last resort antibiotic, increased both transcription of *pks* encoded genes and production of colibactin in the *E. coli* strains NC101 and Nissle. Furthermore, this study revealed that *E. coli* which displayed an increased tolerance to polymyxin, exhibited greater cytotoxicity in intestinal epithelial cells (Sadecki *et al.*, 2021), raising the concern that increased exposure to antibiotics could lead to enhanced virulence in these pro-carcinogenic *E. coli* strains. This is a particular worry for patients with chronic illnesses, who are commonly prescribed antibiotics such as polymyxin B, including individuals with

Cystic fibrosis who are already at an increased risk of developing CRC (Yamada *et al.*, 2018). The polyamine, spermidine, has also been linked to increased colibactin production. A recent study revealed that mutant strains devoid of the spermidine synthase SpeE, were unable to produce pre-colibactin and reported that genotoxic activity was impaired in Δ*speE* mutants (Chagneau *et al.*, 2019). Spermidine, in addition to being produced by intestinal bacteria such as *E. coli* (Tabor and Tabor, 1985), is also produced by eukaryotic cells (Johnson *et al.*, 2015) and can be acquired through the diet (Atiya Ali *et al.*, 2011). Therefore, the gut represents a rich source of spermidine that can be utilised by resident colibactin-producing bacteria, leading to increased genotoxicity in these strains. Indeed, spermidine has been linked to the development of cancer and has been found at high concentrations in CRC tissues (Gerner, Bruckheimer and Cohen, 2018). However, it is not yet known if spermidine supplementation would alter the host microbiota or enhance tumour progression, but with spermidine being identified as a colibactin inducer, it will be important for future studies to address this. Taken together, these data highlight that the diet and the misuse of antibiotics may play a significant role in controlling the production of colibactin and may create a pro-tumorigenic environment. Thus, there is an urgent need to identify novel agents that can control expression of *clb* genes and dampen the production of colibactin in *pks* harbouring strains.

4.3.3 D-Serine is a novel therapeutic with prophylactic potential against colibactin producing *E. coli*

Considering the association between colibactin activity and the development of CRC (Pleguezuelos-Manzano *et al.*, 2020), current research has focused on identifying agents capable of repressing the colibactin biosynthesis pathway. Indeed, studies have identified both chemical and non-chemical agents capable of targeting colibactin production.

Cougnoux *et al*., 2016, described the identification of two boron-based compounds which bound to the active site of the serine peptidase ClbP and suppressed the genotoxic activity of colibactin *in vitro* and in mice (Cougnoux *et al.*, 2016). The heightened focus surrounding colibactin research has helped elucidate the sequence of events that are involved in the biosynthesis of this elusive metabolite. Recently, Jobin and colleagues characterised and described the functional role of the *pks* encoded gene product ClbM. They revealed that ClbM, was a MATE transporter with a conserved transport system that was key for the translocation of pre-colibactin from the cytoplasm into the periplasmic space. Furthermore, structural characterisation of the protein revealed a pre-colibactin specific binding domain that could be an attractive target for small molecule inhibitors to prevent the synthesis of colibactin (Mousa *et al.*, 2016).

However, the increasing prevalence of drug resistance has led to the exploration of new and natural antimicrobial agents. The antibacterial properties of cinnamon and its essential oil (cinnamaldehyde) have been well studied (Singh *et al.*, 2007), with recent research revealing that treatment with these compounds induced downregulation of the *pks*-encoded *clbB* gene *in E. coli* strains isolated from patients with CRC (Kosari *et al.*, 2020). In addition, the biological activity of tannin, extracted from medicinal plants, was also shown to repress transcription of colibactin and prevent the associated genotoxic activity of colibactin producing *E. coli* (Kaewkod *et al.*, 2021). Contrastingly, a study by Santos and colleagues described increased colibactin production in *pks*-harbouring *E. coli* which was associated with exposure to inulin and galacto oligosaccharides, two commonly found oligosaccharides present in prebiotics (Oliero *et al.*, 2021). In view of these findings, it remains clear that colibactin production is strongly influenced by a range of compounds, however, it remains evident that the diet can play a key role in modulating colibactin expression *pks*$^{+}$ *E. coli* strains and further studies will be required to understand the regulatory mechanism(s) in action.

This chapter has highlighted the important role of naturally available amino acids, by identifying D-Serine, as a potent repressor of colibactin. The results presented have shown that exposure to D-Serine prevents the colibactin-associated cytopathic effects in eukaryotic cells, and that treatment with 1mM D-Serine is sufficient to induce prolonged protection. This suggests that D-Serine could have prophylactic potential, providing the host with long-lived protection against colibactin production by commensal *E. coli* residing in the gastrointestinal tract. Furthermore, D-Serine is a component of our diet and therefore presents an opportunity for dietary intervention in the treatment of gastrointestinal disease. Such treatments would particularly benefit high-risk patients, including individuals with inflammatory bowel disease (IBD), where the prevalence of *E. coli* belonging to the B2 phylogroup is high and the incidence of developing CRC is significantly greater (Bernstein *et al.*, 2001; Kotlowski *et al.*, 2007). Therefore, understanding the physiological implications *in vivo* will be key in further exploring the prophylactic potential of D-Serine.

5 Exploring the regulatory role of D-Serine on the *pks* island and beyond

5.1 Introduction

Bacteria encounter a myriad of chemical signals, and the ability to sense and respond to these cues is crucial for bacteria to survive and colonise distinct niches in the host. Indeed, many species harbour unique systems that can identify biomolecules and translate these signals into the appropriate transcriptional response. The D-Serine metabolism locus, *dsdCXA*, enables strains harbouring this pathway to catabolise the host metabolite D-Serine when cells are exposed to inhibitory levels, primarily in the bladder where D-Serine is reported to be at concentrations as high as 1 mM compared to typical concentrations of 1 μM in the gut (Norregaard-Madsen, McFall and Valentin-Hansen, 1995; Anfora *et al.*, 2007). The tolerance locus encodes a D-Serine deaminase (DsdA), a D-Serine inner membrane transporter (DsdX) and an essential transcriptional regulator (DsdC) that regulates the system (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). Studies have demonstrated that the *dsdCXA* locus is extensively carried by uropathogenic *E. coli* (UPEC) isolates (Anfora *et al.*, 2007) and other members of the *E. coli* B2 phylogroup (Connolly *et al.*, 2015). Furthermore, these strains can exploit D-Serine as a sole carbon source, potentially facilitating colonisation at nutrient deficient extraintestinal sites (Anfora and Welch, 2006).

Conversely, D-Serine has been shown to have toxic effects in strains lacking the complete metabolism locus, such as *E. coli* O157:H7, where exposure to D-Serine causes activation of the SOS response and downregulation of the Type 3 Secretin System (T3SS) (Connolly *et al.*, 2015). However, the effects of D-Serine on gene expression have only been studied in pathotypes lacking *dsdCXA* until recently, when exposure to D-Serine was found to incite distinct transcriptional responses among different pathotypes that possessed the tolerance locus (Connolly *et al.*, 2021). Interestingly, the transcriptional response in UPEC

highlighted differential expression in a cluster of genes involved in a non-ribosomal peptide synthetase pathway, encoding genes required for the biosynthesis of colibactin.

In the previous chapter, the D-Serine associated repression of *clb* genes and the subsequent inhibition of colibactin production was characterised *in vitro*. Exposure to D-Serine induced greater than a two-fold downregulation in *clb* genes and limited the colibactin associated cytopathic effects in HeLa cells. However, the regulatory pathway responsible for influencing the transcriptional changes in *pks*$^{+}$ *E. coli* was not alluded to. DsdC has been characterised as a D-Serine responsive LysR-type transcriptional regulator (LTTR) that is required for catabolism via activation of *dsdXA* (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). LTTR are a family of global transcriptional regulators that possess a conserved structure with an N-terminal DNA-binding helix-turn-helix motif, and extensive studies have shown that they can act either as activators or repressors of single or operonic genes (AK and P, 2006; I *et al.*, 2008; Maddocks and Oyston, 2008). Furthermore, LTTRs possess a C-terminal co-inducer-binding domain. Co-inducers have been recognised as important for the function of LTTRs, and often they involve a feedback loop in which the LTTR is activated or repressed in response to a product of a given metabolic/ synthesis pathway (Maddocks and Oyston, 2008). Therefore, it was hypothesised that the transcriptional changes in response to D-Serine were mediated by the LTTR, DsdC. Isogenic mutants were generated in two *pks*$^{+}$ strains, CFT073 and Nissle, and the effects on colibactin production in response to D-Serine treatment was assessed.

5.2 Results

Over recent years, it has emerged that amino acids play an important role in bacterial gene regulation. In Chapter 4, a range of L and D- amino acids were identified as modulators of the recently discovered genotoxin, colibactin. D-Serine was identified as one of the most consistent and potent repressors of *clbB* in two *pks*$^{+}$ *E. coli* strains: CFT073 and Nissle. Furthermore, treatment with D-Serine dampened the colibactin associated cytotoxic activity of Nissle during infection of host cells and conferred long-lasting protection in mammalian cells. However, the regulatory mechanism involved in D-Serine associated inhibition of colibactin, remains unknown.

5.2.1 The D-Serine metabolism locus is not essential for D-Serine associated repression of colibactin in *pks*$^{+}$ *E. coli*

Certain strains of *E. coli* possess the unique ability to metabolise D-Serine by encoding a specialised catabolic pathway, the *dsdCXA* locus (Metzler and Snell, 1952; Norregaard-Madsen, McFall and Valentin-Hansen, 1995). DsdC has been well studied as a D-Serine responsive transcriptional regulator that is required for catabolism through activation of *dsdXA* (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). Therefore, it was hypothesised that the repression of the *pks* island by D-Serine could be mediated by DsdC. To investigate the hypothesis, *dsdC* was deleted in Nissle and the effect on colibactin production was assessed.

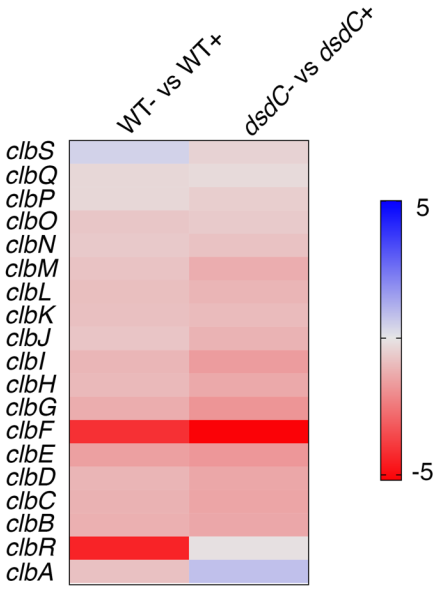


Fig 5-1 D-Serine induces down regulation of *clb* genes in WT and mutant Δ*dsdC* CFT073. Heat map, showing the EdgeR calculated $\log_2$ relative fold changes for each gene in the *pks* island with corresponding colour key adjacent. False discovery rate-corrected *P* values can be found in Appendix 2.

First, the RNA-Seq data from a previous study was re-analysed to compare the response to D-Serine in CFT073 WT and Δ*dsdC* CFT073 (Connolly *et al.*, 2021). The analysis of the RNA-Seq data revealed that exposure to D-Serine triggered similar fold changes in both WT and Δ*dsdC* CFT073 strains. The results are summarised in the heat map presented in Fig 5-1, which shows downregulation of colibactin synthesis genes upon addition of 1mM D-Serine in both the wild type and *dsdC* deletion mutant. To assess whether the effect of deleting DsdC was the same in Nissle, expression of *clbB* was measured by RT-qPCR for Nissle WT and Nissle Δ*dsdC* in the presence and absence of D-Serine.

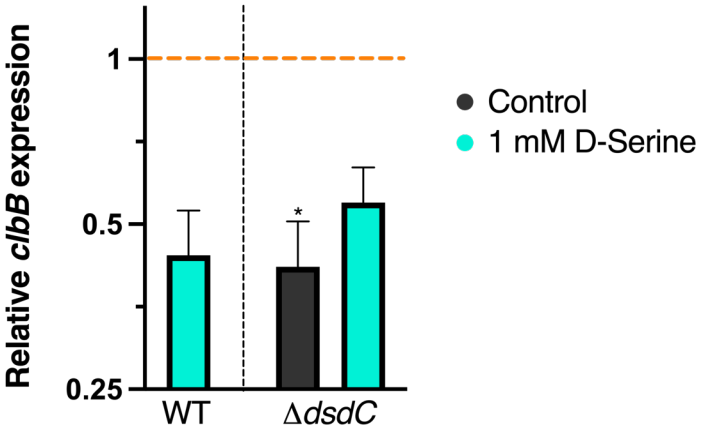


Fig 5-2 Exposure to D-Serine downregulates *clbB* in Nissle Δ*dsdC*. Relative *clbB* expression was measured by RT-qPCR. Nissle WT and Nissle Δ*dsdC* were grown in MEM-HEPES media supplemented with 1 mM D-Serine for 5 h. The orange dashed line indicates baseline expression for Nissle WT without D-Serine, with columns below this line representing downregulation. Statistical significance was determined from three biological replicates using a one-way ANOVA, * indicating significance $P < 0.05$.

The expression values presented in Fig 5-2 indicate that *clbB* was significantly downregulated by 2.38-fold in Nissle Δ*dsdC* ($P = 0.018$), compared to the WT control. Notably, the addition of D-Serine increased expression of *clbB* in Nissle Δ*dsdC* compared to the untreated mutant, however, *clbB* remained repressed with expression decreased 1.8-fold compared to Nissle WT. These results confirmed the observations made by RNA-Seq and indicated that DsdC does not mediate *clbB* transcription in response to D-Serine, however, DsdC may still contribute to regulation of the *pks* island.

Next, to assess whether the deletion of *dsdC* affected the genotoxic activity, HeLa cells were infected with Nissle Δ*dsdC*. Briefly, cells were infected for 4 h in the presence and absence of D-Serine, then wells were washed and replenished with media supplemented with gentamicin before incubating for a further 4 h. Cell lysates were extracted and

proteins were separated by SDS PAGE. To assess the level of DNA damage, detection of phosphorylated H2AX was determined by immunoblotting, and normalised signal intestines from three experiments were compared to quantify levels of phosphorylation.

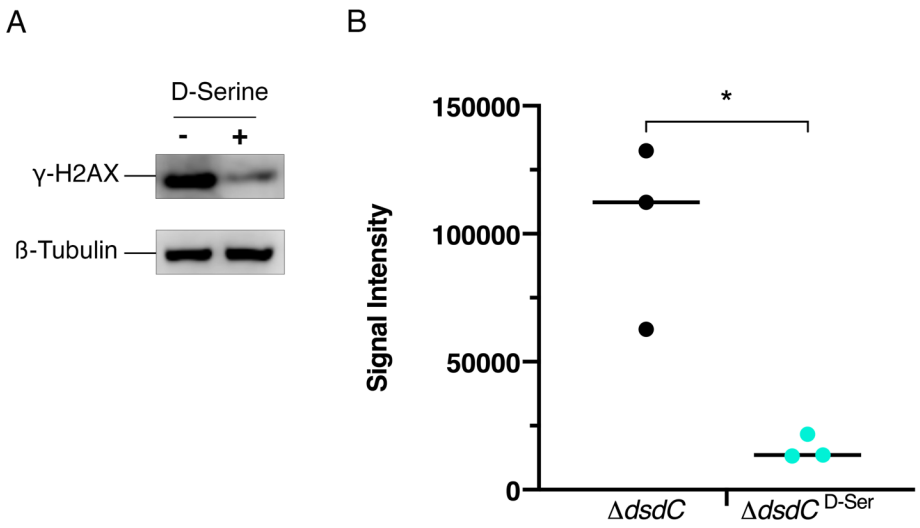


Fig 5-3 Genotoxic activity is reduced in HeLa cells infected with Nissle Δ*dsdC* treated with D-Serine. HeLa cells were infected for 4 h with live Nissle Δ*dsdC* with an MOI = 400 bacteria per cell or left uninfected. Infections were performed in wells containing MEM-HEPES alone (-) or with media supplemented with 1mM D-Serine (+). (A) Immunoblot analysis of cell lysates extracted 4 h post infection. Phosphorylated histone (γ-H2AX) was employed as an indicator of double stranded DNA breaks and β-Tubulin was used as a loading control. (B) Signal intensities of bands were measured using LI-COR Image Studio software. γ-H2AX signals were corrected to account for any variation in loading using β-Tubulin signal intensity. Experimental signal was normalised, so that the mean signal intensity of the two samples was equivalent for each experiment. The experiment was carried out in triplicate. Columns represent mean +/- SEM with individual experimental observations indicated by data points. Statistical significance was assessed by unpaired Student's *t*-test with * indicating $P < 0.05$.

Detection of γ-H2AX by immunoblotting revealed that more phosphorylated proteins were present in cell lysates extracted from wells infected with Nissle Δ*dsdC* alone compared with treated Nissle Δ*dsdC* (Fig 5-3A). Indeed, the signal intensities presented in Fig 5-3B illustrate that levels of phosphorylation were significantly reduced by 6.34-fold in cells infected with Nissle Δ*dsdC* exposed to D-Serine (P = 0.035). This indicated that deletion of DsdC did not interfere with D-Serine associated inhibition of colibactin induced genotoxicity.

Following the examination of genotoxic activity, the effect of D-Serine on colibactin-associated cellular senescence was assessed in HeLa cells infected with Nissle Δ*dsdC*. The experiment was performed as described in Chapter 4. Briefly, cells were infected for 4 h, then wells were washed and replenished with media supplemented with gentamicin. After a further 4 h incubation, wells were washed and media was replenished before cells were incubated for 72 h to allow for the development of the large cell phenotype.

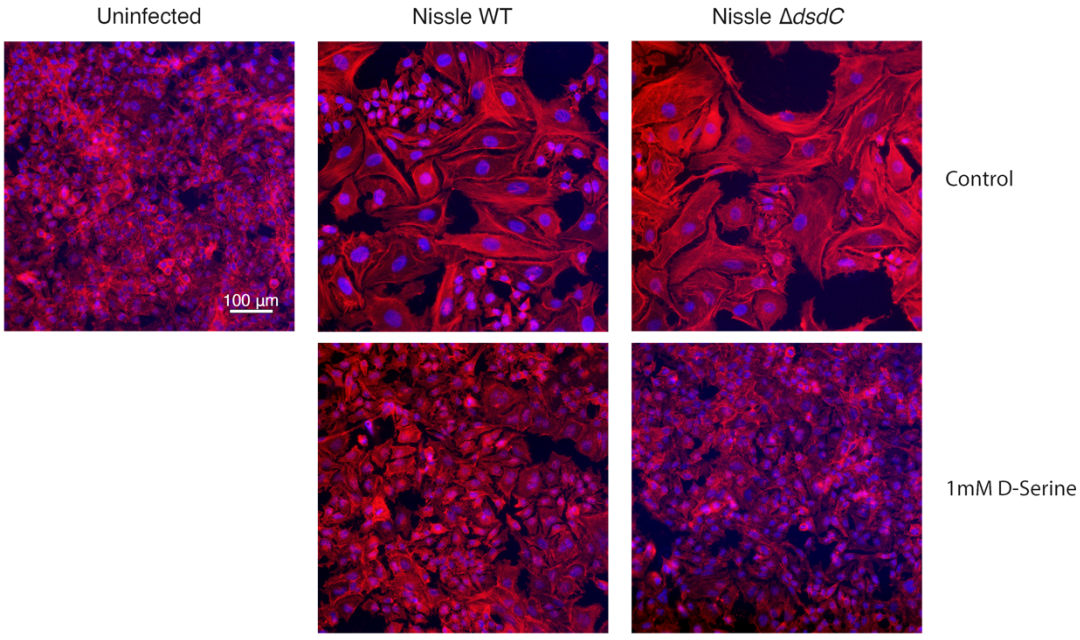


Fig 5-4 Treatment with D-Serine reduces the cytopathic effect in Nissle Δ*dsdC*. HeLa cells were infected for 4 h with Nissle or Nissle Δ*dsdC* (MOI = 400). Infections were

performed with and without the addition of 1 mM D-Serine to the growth media. At 8 h after infection, cells were washed and incubated for 72 h to allow for the megacell phenotype to develop. HeLa cell morphology was observed by wide field fluorescence. Actin cytoskeleton was stained with Phalloidin in red and DNA was counterstained with DAPI in blue at 72 h post infection.

Megalocytosis is a well characterised phenomenon that has been associated with colibactin and is observed when eukaryotic cells are infected with live *pks*$^{+}$ *E. coli* (Nougayrede *et al.*, 2006). Observations made at 72 h post-infection revealed that HeLa cells infected with either Nissle WT or Nissle Δ*dsdC* displayed the characteristic colibactin-associated megacell phenotype and enlargement of the cell body and nucleus was apparent. In accordance with the findings described in Fig 5-3, treatment with 1 mM D-Serine evoked the same response in Δ*dsdC* and Nissle WT. Fig 5-4 illustrates that cells infected with both mutant or WT Nissle displayed distinctively reduced cellular senescence, and cell morphology was evidently like that of the uninfected control. These findings suggest that exposure to D-Serine decreases the genotoxic activity of Nissle Δ*dsdC* to a similar extent as has previously been described in Nissle WT (see Chapter 4). Intriguingly, deletion of the D-Serine responsive regulator, DsdC, does not perturb the effects of D-Serine in Nissle. Therefore, these findings allude to an alternative mechanism involved in the D-Serine-associated downregulation of colibactin synthesis.

5.2.2 DsdC binds to ClbR, the key activator of colibactin biosynthesises

Since its discovery in 2006, much of the previous research on colibactin has focused on solving the molecular structure and understanding the mode of action of active colibactin.

However, more recently ClbR, which has been described as a LuxR-like protein with a helix-turn-helix binding motif (Homburg *et al.*, 2007), was characterised as the transcriptional activator of colibactin biosynthesis (Wallenstein *et al.*, 2020). Furthermore, Dobrindt and colleagues demonstrated that cultivation methods influenced *clbR* expression levels. Increased transcript levels were observed upon bacterial cultivation in M9 minimal media relative to rich media, suggesting that nutrient availability contributes to colibactin production. Incidentally, additional research conducted by the Roe lab has been investigating the regulatory role of DsdC, also a transcriptional activator. A chromatin-immunoprecipitation coupled with next generation sequencing (ChIP-Seq) experiment, was designed and performed by Dr Natasha Turner, in which CFT073 was cultured in M9 minimal media in the presence or absence of D-Serine. Strikingly, analysis of the binding sites revealed peaks for genes in the colibactin biosynthesis gene cluster.

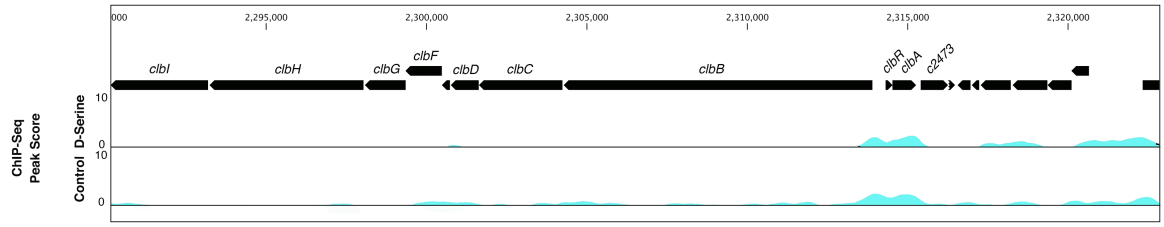


Fig 5-5 DsdC binds directly to *clbR* and *clbA* in the presence and absence of D-Serine. ChIP-Seq peaks from CFT073 grown in M9 media alone (labelled control) or supplemented with 1 mM D-Serine. Binding peaks, in cyan, were called using CLC genomics software and the corresponding genes are annotated above the track maps.

The track maps in Fig 5-5 illustrate a region of the genome encoding some of the colibactin biosynthesis genes (*clbI-clbA*). Strikingly, two binding peaks are visible for the regions encoding the genes *clbR* and *clbA* and potentially indicates direct DNA binding of DsdC. Therefore, this data suggests that DsdC is not solely restricted to activation of the *dsdXCA* locus, instead DsdC could potentially bind and regulate genes throughout the genome.

The ChIP-Seq experiment highlighted that DsdC could have a significant role in regulating colibactin associated genes through direct binding of *clbR* and *clbA*. Interestingly, these two genes are orientated in the opposite direction to the other *pks* encoded genes and encode gene products necessary for the regulation and activation of colibactin production (Homburg *et al.*, 2007). Therefore, to investigate the role of DsdC on gene expression, transcript levels of *clbR* and *clbA* were measured by RT-qPCR in Nissle WT and Nissle Δ*dsdC*.

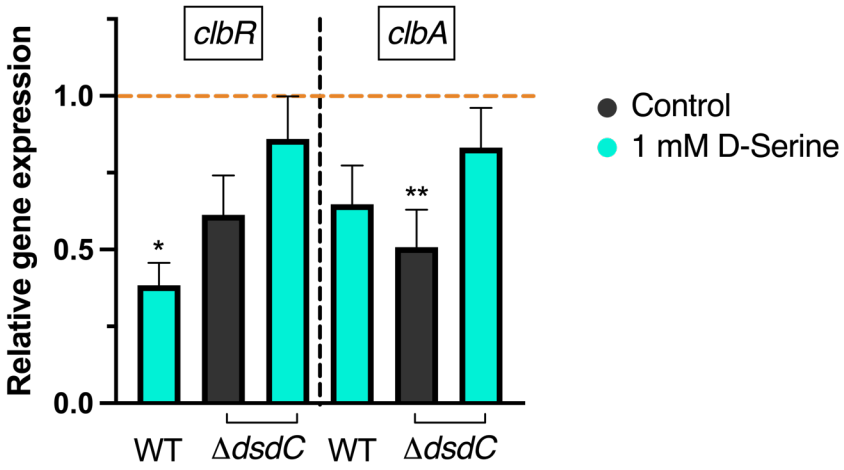


Fig 5-6 DsdC regulates expression of *clbA* and *clbR* in response to D-Serine. Relative *clbR* and *clbA* expression was measured by RT-qPCR. Nissle WT and Nissle Δ*dsdC* were grown in MEM-HEPES media supplemented with 1 mM D-Serine for 5 h. The orange dashed line indicates baseline expression for Nissle WT without D-Serine, with columns below this line representing downregulation. Statistical significance was determined from three biological replicates using a one-way ANOVA, * and ** indicating significance, $P <$ 0.05 and 0.01, respectively.

The expression values in Fig 5-6 indicate that transcripts of *clbR* and *clbA* were downregulated in the Δ*dsdC* background by 1.63-fold and 1.97-fold, respectively. Strikingly, upon exposure to D-Serine, the expression of *clbR* was increased from 0.61 to

0.86 and *clbA* expression was also increased from 0.51 to 0.83, thus suggesting that DsdC downregulates *clbR* and *clbA* in the presence of D-Serine.

The transcriptional activator, ClbR, directs the efficient regulation of the *pks* island and is essential in the production of active colibactin (Wallenstein *et al.*, 2020). Therefore, as it was indicated that DsdC could bind to this gene (Fig 5-5), it was hypothesised that DsdC could affect production of colibactin through regulation of ClbR. A fluorescence reporter assay was performed to measure the promoter activity of *pks* encoded *clbB* in 10 WT and 10 Δ*dsdC* Nissle single colonies. Overnight cultures were prepared for each single colony and then inoculated 1:100 into a 96-well plate containing MEM-HEPES alone or media supplemented with 1 mM D-Serine. The plate was incubated in a humidity chamber, shaking for 6 h, with fluorescence intensity was measured every hour.

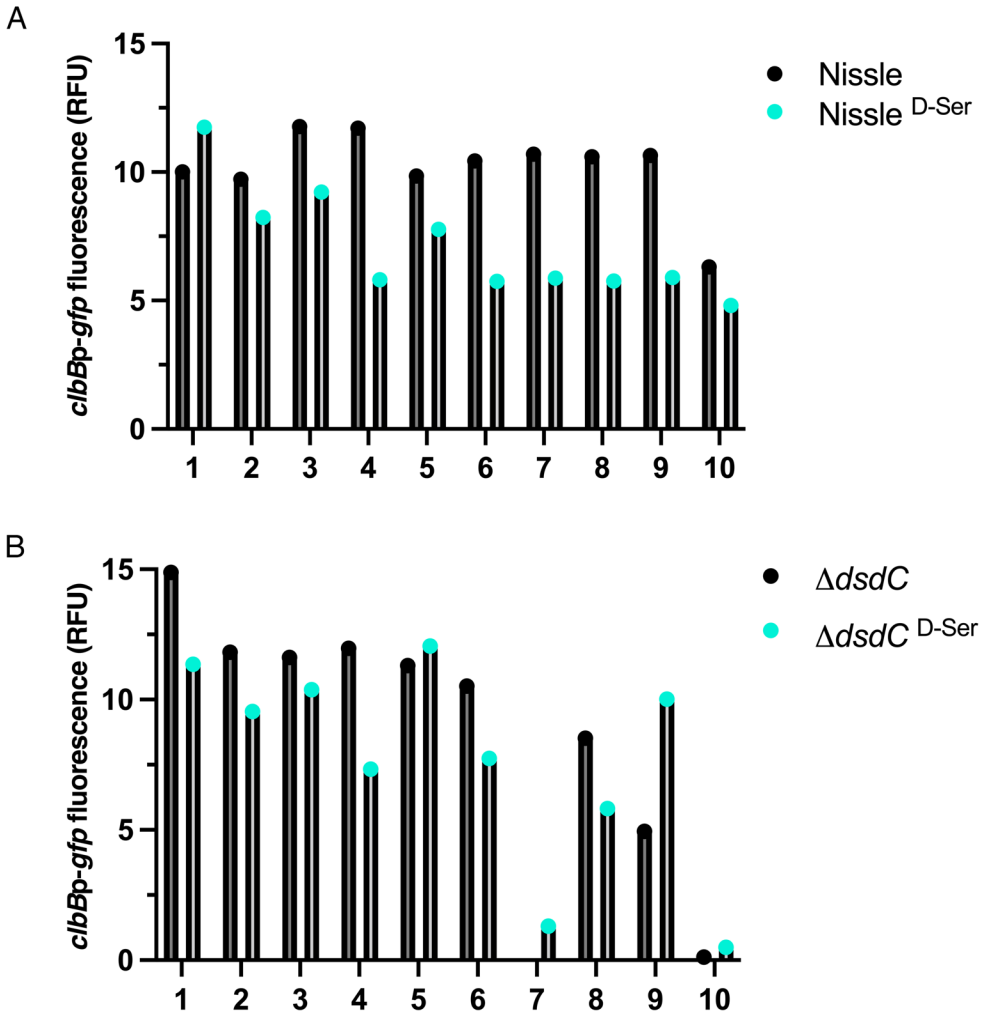


Fig 5-7 Expression of *clbB* in mutant and wild type isolates in response to D-Serine. The *clbBp*:*gfp* reporter activity of 10 WT and 10 Δ*dsdC* Nissle isolates was measured in MEM-HEPES in the presence and absence of 1mM D-serine. Fluorescence was measured at hourly intervals over a 6 h growth period. The data presented shows the relative fluorescence units (RFU) recorded at 5 h post-inoculation, and *clbB* expression was measured as GFP/$OD_{600\ nm}$ with negative values excluded from the data set.

Interestingly, expression of *clbB* was relatively consistent among the WT isolates and a range of 5.47 was determined from the data presented in Fig 5-7A. Conversely, *clbB* expression was more varied in the mutant isolates with the range calculated as 15.30 for the RFUs presented in Fig 5-7B. Similarly, expression of *clbB* in response to D-Serine was more consistent in WT isolates, with the range calculated as 6.93. In comparison, the

response to D-Serine among mutant isolates was variable and the range was higher at 10.86. Together these data demonstrate that expression of *clbB* is more variable in the Δ*dsdC* Nissle background and suggests that DsdC might have some regulatory function, potentially through regulation of ClbR.

The ChIP-Seq peak in Fig 5-5 revealed that DsdC bound to an intergenic region between *clbR* and *clbB.* Interestingly, this location encodes a region with variable numbers of tandem repeats (VNTR) and consists of an 8 base-pair nucleotide sequence, 5'-ACAGATAC-3' (Wallenstein *et al.*, 2020). The number of repeats can vary, although differences to the number of octanucleotide repeats has only been described between different strains (Putze *et al.*, 2009). Furthermore, it remains unclear whether the number of tandem repeats can affect colibactin expression. Therefore, to investigate if the variability in colibactin expression observed in the Δ*dsdC* Nissle mutants resulted from differences in the VNTR region, this section of genome was amplified from the same 10 WT and 10 Δ*dsdC* Nissle single colony isolates as described in Fig 5-7, and the products were sequenced.

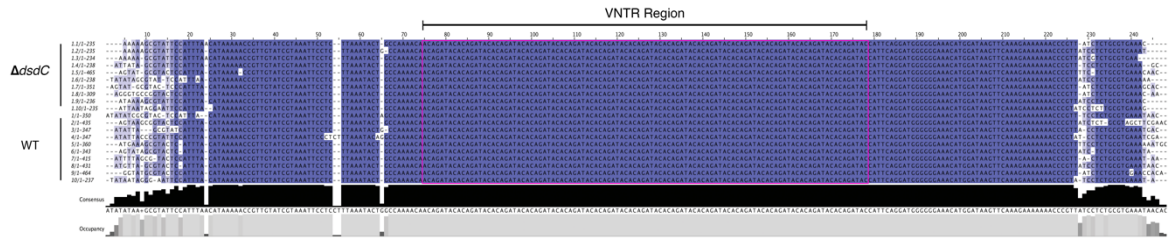


Fig 5-8 VNTR region in Δ*dsdC* and wild type Nissle. Sequences from 10 WT and 10 Δ*dsdC* isolates were aligned using the Clustal tool available in Jalview. Percentage identity is represented by colour intensity, with the most intense purple equal to 100%. The VNTR region is highlighted with a pink box.

The sequence alignments illustrated in Fig 5-8 revealed that there was 100% sequence identity in the VNTR region of 10 WT and 10 Δ*dsdC* single colony isolates. Indeed, it is apparent that this region is highly conserved. The total number of octanucleotide repeats was determined as 13 and this did not vary between isolate, confirming the observations made by *Putze et al,* 2009. Furthermore, deletion of DsdC did not have any deleterious effects on the VNTR region.

5.2.3 The role of D-Serine beyond the *pks* island

D-Serine has recently been implicated with inducing genome wide transcriptional changes in *E. coli* pathotypes, including UPEC strain CFT073 that also carries the *pks* cluster (Connolly *et al.*, 2021). Earlier in this chapter, the inhibitory effect of D-Serine on the expression of colibactin biosynthesis genes encoded on the *pks* island, was described (Fig 5-1). However, further analysis of the RNA-Seq data revealed several other genes of interest that were also differentially expressed in the presence of D-Serine, some of these genes are annotated in Fig 5-9A.

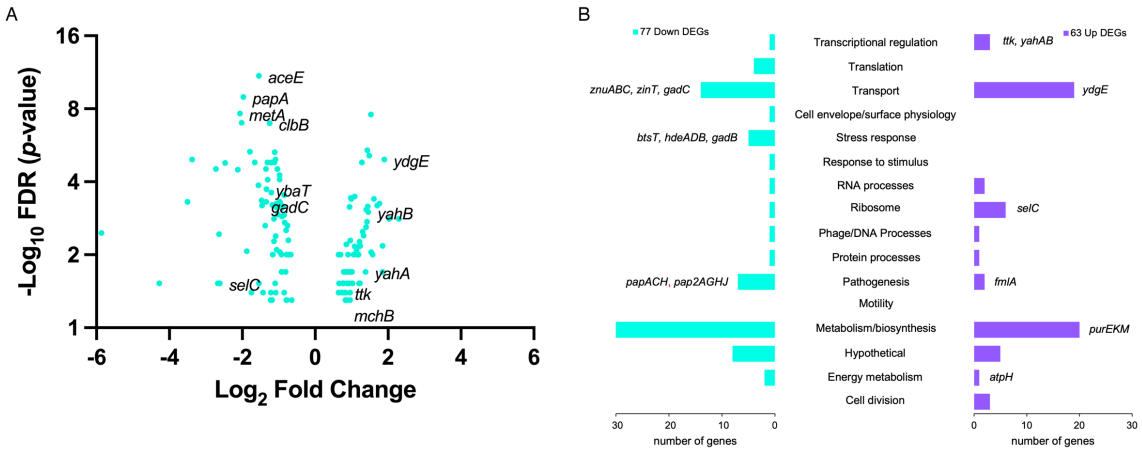


Fig 5-9 Transcriptional response to D-Serine in CFT073. (A) Volcano plot depicting differential gene expression in CFT073. (B) Functional category analysis of DEGs.

Samples used for RNA-Seq analysis were cultured for 5 h in M9 minimal media alone or media that was supplemented with 1 mM D-Serine 3 h post-inoculation.

Most of the differentially expressed genes (DEGs) in the data set belonged to the transport and metabolism/biosynthesis functional groups (Fig 5-10B). Notably, the upregulation of *mch* genes, which are involved in microcin biosynthesis, was of note. Interestingly, the interplay between the production of microcins and colibactin was recently investigated in Nissle, and it was shown that the probiotic properties of this strain could not be disassociated from its genotoxic activity (Olier *et al.*, 2012). Therefore, as D-Serine had exhibited a modulatory effect on microcin associated genes in CFT073, it was hypothesised that D-Serine could also affect the expression of microcin genes in Nissle. Thus, the expression of genes encoding the precursors to microcin M and H47, *mcmA* and *mchB*, were measured by RT-qPCR in response to D-Serine.

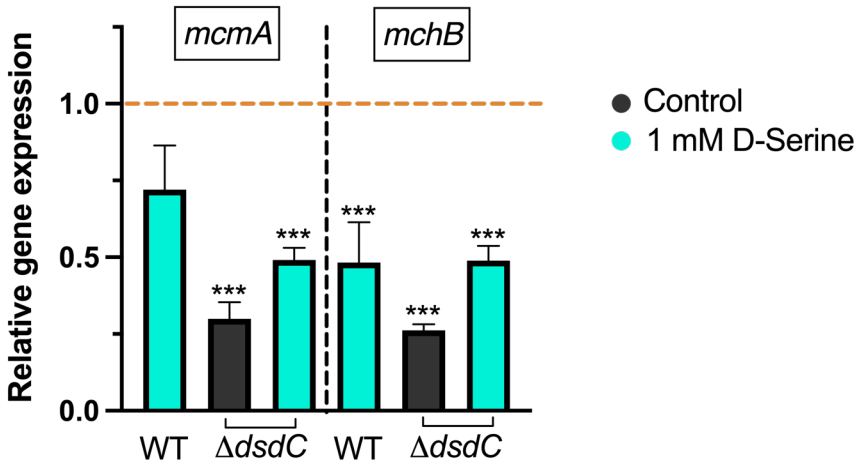


Fig 5-10 Expression of microcin genes is inhibited in the presence of D-Serine. Relative *mcmA* and *mchB* expression was measured by RT-qPCR. Nissle WT and Nissle Δ*dsdC* were grown in MEM-HEPES media supplemented with 1 mM D-Serine for 5 h. The orange dashed line indicates baseline expression for Nissle WT without D-Serine, with columns below this line representing downregulation. Statistical significance was

determined from three biological replicates using a one-way ANOVA, *** indicating significance, $P < 0.001$.

The expression values in Fig 5-10 revealed that D-Serine had the opposite effect on microcin genes in Nissle. Upon exposure to D-Serine, *mcmA* and *mchB* were repressed by 1.39-fold and 2.0-fold, respectively, in Nissle WT. The expression of microcin genes was also tested in the Δ*dsdC* background and strikingly the expression of both *mcmA* and *mchB* was significantly downregulated compared to the WT control ($P < 0.001$). Furthermore, in response to the addition of D-Serine, the expression values of *mcmA* and *mchB* increased. These data further elucidate the regulatory role of DsdC and demonstrate a potential novel function in the regulation of microcin biosynthesis.

Microcin M and microcin H47 have both been linked to the antibacterial activity of Nissle (Sassone-Corsi *et al.*, 2016). Therefore, as D-Serine had been shown to modulate the expression of microcin biosynthesis genes (Fig 5-10), it was hypothesised that D-Serine could also perturb the antibacterial activity associated with Nissle. A co-culture assay was performed in M9 minimal media with Nissle and the adherent and invasive *E. coli* strain LF82. The effect of D-Serine on the killing activity of Nissle, was assessed in both the WT and Δ*dsdC* backgrounds.

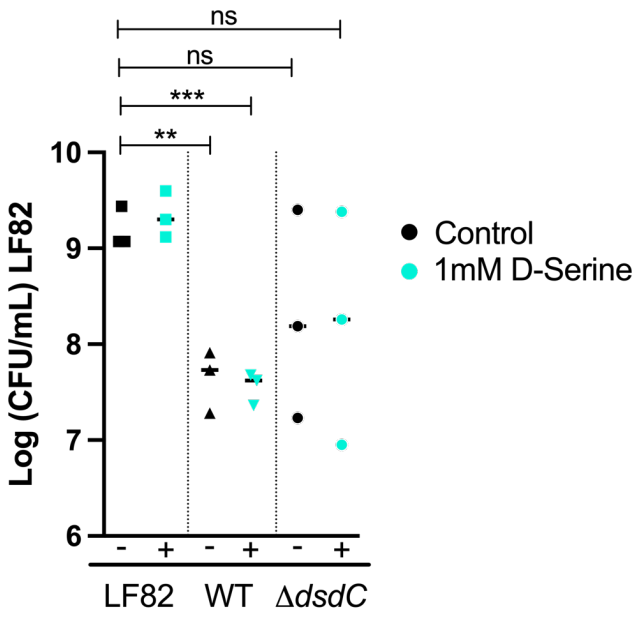


Fig 5-11 Exposure to D-Serine does not affect killing activity of Nissle. Log colony forming units (CFU) of LF82 following 24 h co-culture in M9 minimal media with Nissle WT and mutant Nissle Δ*dsdC*. LF82 was also cultured alone as control. Cocultures were performed in M9 media alone (control) or supplemented with 1 mM D-Serine. The medians and individual results from three independent experiments are shown. Statistical significance was determined by unpaired Student's *t*-test, **, *** indicating significance, $P < 0.01$ and 0.001, respectively.

Strikingly, despite D-Serine inducing downregulation of the individual microcin biosynthesis genes (Fig 5-10), killing activity of Nissle WT was not affected by exposure to D-Serine during co-culture with LF82. Indeed, CFUs for LF82 were significantly reduced 32.92-fold when co-cultured with Nissle and similarly, a significant reduction of 44.99-fold was observed in Nissle treated with D-Serine ($P = 0.0023$ and 0.005, respectively). Surprisingly, the killing activity observed in the Δ*dsdC* mutant was variable. However, when comparing the CFUs from untreated and treated Nissle Δ*dsdC* co-cultures, it appeared that the mutant was also unresponsive to D-Serine. These findings highlight that D-Serine does not perturb the killing activity of Nissle against LF82. Additionally, this

data suggests that DsdC could potentially contribute to the regulation of microcin biosynthesis in Nissle.

The binding profile of DsdC was investigated by analysing peaks in the region of the CFT073 genome, where microcin associated genes were encoded.

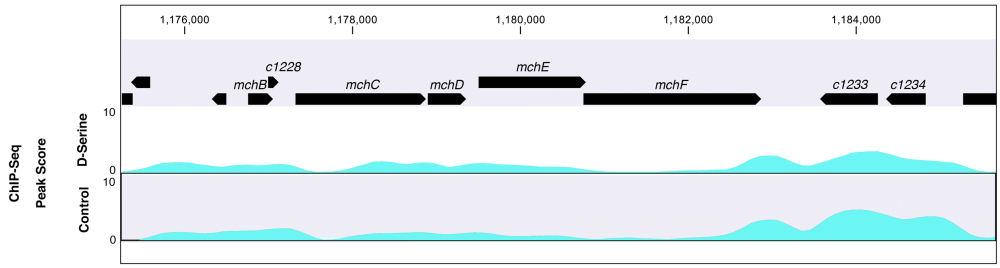


Fig 5-12 DsdC binds directly to microcin biosynthesis genes in the presence and absence of D-Serine. ChIP-Seq peaks from CFT073 grown in M9 media alone (labelled control) or supplemented with 1 mM D-Serine. Binding peaks, in cyan, were called using CLC software and the corresponding genes are annotated above the track maps.

Analysis of the ChIP-Seq peaks revealed a canonical binding motif typical of LTTRs downstream of the *mchF* gene (Fig 5-12). Although not annotated on the CFT073 genome used to map the ChIP reads, the region shares 100% similarity with the Nissle genome where the gene *mcmA* is encoded directly downstream of *mchF*. Therefore, it can be speculated that this peak corresponds to the microcin gene *mcmA,* and that binding of DsdC could be indicative of a potential regulatory role of DsdC in the regulation of microcin M biosynthesis.

5.3 Discussion

Chapter 4 revealed the unique ability of D-Serine to inhibit the expression of colibactin biosynthesis genes and dampen the genotoxic activity of the colibactin producing strain Nissle; however, the regulatory mechanism was not clear. Exposure to D-Serine has been demonstrated to induce the expression of the LTTR DsdC, the main function of which has been described to be the activation of *dsdXA* transcription (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). This cluster facilitates uptake and catabolism of D-Serine, respectively. Therefore, as DsdC is the primary transcriptional regulator induced in response to D-Serine, this chapter investigated whether the repression of *clb* genes was mediated by DsdC. Experiments were performed with Nissle Δ*dsdC* to investigate the expression of *pks* encoded genes in response to D-Serine. Further, the colibactin-associated cytopathic effects were assessed in eukaryotic cells infected with the Δ*dsdC* in the presence and absence of D-Serine. The results from these investigations were surprising and revealed a more complex regulatory pathway that extended beyond the *pks* island.

5.3.1 DsdC mediated utilisation of D-Serine is not essential for inhibition of colibactin

Experiments performed with Nissle Δ*dsdC* indicated that repression of colibactin by D-Serine was not directly facilitated by DsdC as observations revealed similar responses to D-Serine in both WT and Nissle Δ*dsdC* in terms of the reduction in cytopathic responses (Fig 5-3 & 5-4). Interestingly, analysis of RNA-Seq data comparing the effects of D-Serine on UPEC WT and UPEC Δ*dsdC* transcriptomes revealed the same observation, in that *clb* genes were downregulated in both WT and mutant conditions (Fig 5-1). The response to

D-Serine in both UPEC and Nissle Δ*dsdC* mutants was unexpected, however, it did highlight the possibility of a shared mechanism of action between these two *pks* encoding *E. coli* strains.

One of the roles of the D-Serine metabolism locus is to prevent *dsdCXA*$^{+}$ *E. coli* strains from succumbing to inhibitory concentrations of D-Serine in the urinary tract, by converting the substrate to ammonia and pyruvate (Cosloy and McFall, 1973; Norregaard-Madsen, McFall and Valentin-Hansen, 1995; Anfora *et al.*, 2007). Indeed, it is plausible that the D-Serine associated inhibitory effects on colibactin could be attributed to one of the compounds produced during catabolism. However, DsdC and D-Serine have been shown to be essential for the expression of the D-Serine deaminase, *dsdA* (Norregaard-Madsen, McFall and Valentin-Hansen, 1995) and thus, without degradation via DsdA, ammonia and pyruvate are not produced in the Δ*dsdC* background. Therefore, it is more likely that the inhibitory effects arise from exposure to D-Serine. Furthermore, D-Serine could enter the cell via an alternative transport system, as in addition to transport mediated by DsdX, D-Serine can enter the cell via CycA (Cosloy, 1973; Anfora and Welch, 2006). Thereby providing a potential explanation for why exposure to D-Serine repressed expression of the *clb* genes and reduced genotoxic activity in both Nissle Δ*dsdC* and in the WT backgrounds.

Analysis of CFT073 transcripts revealed several other regulators that were significantly differentially expressed in response to D-Serine (Fig 5-9). Three regulators of particular interest were *ttk* (aka *SlmA)* a nucleoid occlusion protein with a TetR transcriptional regulator domain, *yahB* a LTTR regulator and *yahA* also a DNA binding transcriptional regulator. Strikingly, all three of these genes were upregulated in the presence of D-Serine in both the CFT073 WT and Δ*dsdC* data sets. Indeed, it is likely that these proteins responded to D-Serine independently of activation by DsdC and therefore could potentially

exhibit unique regulatory functions. To investigate if these proteins exerted a regulatory role during D-Serine associated inhibition of colibactin, deletion mutants were attempted in CFT073, unfortunately no successful mutants were selected during this project. However, investigating the implications on *clb* expression by deleting these regulators could further illuminate the regulatory networks involved in D-Serine dependent repression of colibactin.

Recently, Wallenstein *et al.*, 2020, demonstrated that the *pks* encoded *clbR* gene is the transcriptional activator of the colibactin biosynthesis genes. ClbR has been described as containing a LuxR-type DNA-binding helix-turn-helix domain, which is typically found in response regulators of the LuxR family and thereby supports the view that it is involved in the regulation of gene expression (Homburg *et al.*, 2007; Wallenstein *et al.*, 2020). Intriguingly, unlike other members of the LuxR family, ClbR does not possess an N-terminal receiver (REC) domain and therefore it has been suggested that ClbR is not activated through ligand binding, instead activation requires an alternative molecular mechanism that has yet to be described (Wallenstein *et al.*, 2020). Interestingly, Dobrindt and colleagues also found that expression of *clbR* was influenced by iron availability with transcription being strongly repressed in iron abundant conditions compared with iron limited conditions (Wallenstein *et al.*, 2020). Therefore, it is possible that expression of *clbR* could be mediated by other nutrients available in the growth media including amino acids such as D-Serine.

5.3.2 Regulatory role of DsdC in colibactin biosynthesis and beyond

Understanding the regulation of colibactin has become an area of immense interest, particularly as the production of this secondary metabolite has been linked with the development of CRC (Raisch *et al.*, 2014). Therefore initially, Δ*dsdC* deletion mutants were generated to investigate whether D-Serine associated repression of colibactin was mediated by the transcriptional regulator DsdC. Assessment of the response to D-Serine in both CFT073 and Nissle revealed that downregulation of *clb* genes in the presence of D-Serine was independent of DsdC. However, analysis from a previous experiment in which the DNA binding pattern of DsdC in CFT073 was assessed, revealed two binding peaks at the region encoding colibactin genes *clbR* and *clbA*. Furthermore, DsdC binding peaks were present in both the untreated and treated conditions, suggesting that DsdC binds to the genome and is likely activated by its co-inducer, D-Serine, to become involved in transcription. However, direct physical interaction of DsdC with the promoter region of *clbR* remains to be determined.

Thus far, the transcriptional activator ClbR, is the first and only regulator to be associated with the regulation of colibactin production (Wallenstein *et al.*, 2020). However, the results discussed in this chapter suggest a potential role for DsdC that could be implicated in the regulation of ClbR. D-Serine associated colibactin repression was observed in both Nissle lacking DsdC and the WT, therefore it was apparent that DsdC did not mediate the response to D-Serine. Alternatively, it may be possible that DsdC is not involved in the expression of colibactin but may be involved in the expression of other colibactin pathway-dependent compounds. Indeed, research has found that the *pks* island contributes to the production of multiple secondary metabolites, which, in addition to colibactin, includes the

analgesic lipopeptide C12-Asn-GABA that can act on sensory neurones to influence pain in the host (Vizcaino *et al.*, 2014; Auvray *et al.*, 2021). Intriguingly, LysR-type regulators have been implicated with the regulation of biosynthetic gene clusters in other Gram-negative bacteria (Thapa and Grove, 2019). ScmR was identified as a master regulator in *Burkholderia thailandensis* E264 and displayed the ability to interact with pathway-specific transcriptional regulators to control and supress secondary metabolite production (Mao *et al.*, 2017). The global LTTR, MvfR, was also demonstrated to regulate expression of secondary metabolism by Rahme and colleagues, who identified MvfR as a putative activator of pyocyanin biosynthesis in *Pseudomonas aeruginosa* (Cao *et al.*, 2001). However, efforts to understand the physical protein interactions of DsdC with *clbR* are required, and electrophoretic mobility shift assays (EMSA) should be considered in future experiments to validate the ChIP binding.

In addition to *clbR*, ChIP-Seq analysis revealed that the protein DsdC bound to a region of the UPEC genome encoding microcin genes. BLAST queries identified that Nissle shared 100% similarity in this region, suggesting that DsdC had the potential to bind to the same genes in Nissle. Gene expression was determined by RT-qPCR in Nissle WT and Nissle Δ*dsdC* and revealed that DsdC exerted a regulatory role in the transcription of *mchB* and *mcmA*, with genes becoming suppressed in the mutant background. Intriguingly, microcin genes are believed to contribute to the probiotic activity of Nissle (Sassone-Corsi *et al.*, 2016), therefore it was interesting that DsdC appeared to exert a regulatory function on both colibactin and microcin associated biosynthesis genes. Recently, Massip *et al*., 2019, successfully decoupled the antimicrobial activity from the genotoxic activity of Nissle. The study revealed a surprise observation and demonstrated that the *pks* island was connected to the production of microcin H47 and microcin M, the precursors of which are encoded by *mchB* and *mcmA*, respectively (Massip *et al.*, 2019). Interestingly, deletion of *dsdC* resulted in Nissle demonstrating variable levels of killing activity against LF82 (Fig 5-11),

which could suggest that DsdC might play a role in the regulation of microcin production. However, D-Serine did not appear to affect the killing activity of Nissle WT, despite the earlier observations that expression of *mchB* and *mcmA* were downregulated in the presence of D-Serine (Fig 5-10). To explore these observations further, deletion mutants for *mchB* and *mcmA* were attempted in Nissle, with the prospect of identifying which microcin was responsible for LF82 killing and whether one microcin was more sensitive to D-Serine than the other. However, no successful mutants were selected, and this remains to be continued in future work.

Together, these findings suggest a potential role for the LTTR protein DsdC in controlling the synthesis of secondary metabolites in Nissle, but the role of D-Serine remains ambiguous. One function of D-Serine could be to act as a co-effector required for activation of DsdC during the biosynthesis of colibactin and microcins. However, this does not explain the D-Serine associated repression of colibactin in the Δ*dsdC* background and could suggest that the response is influenced by another regulatory pathway. Future work, to investigate alternative D-Serine responsive transcriptional regulators, will build on these initial findings and improve the current understanding.

6 Final discussion

Extensive research suggests that the intestinal microflora plays a key role in the pathophysiology of various human diseases. In particular, compositional changes to the microbiota have been linked to the development and progression of certain cancers. Cancer poses a significant burden on human health and life expectancy. Indeed, statistics from 2020 estimated that there was approximately 19.3 million cancer cases and 10 million cancer-related mortalities recorded worldwide (Sung *et al.*, 2021). Notably, colorectal cancer (CRC) was found to be the third most common form of malignancy and was associated with the second highest mortality rate compared with other cancers (Sung *et al.*, 2021). Many risk factors have been linked to the development of cancer, including inherited genetics and various lifestyle factors (Turnbaugh *et al.*, 2007). However, there is a growing number of infection-related cancer cases, with almost 20% of common malignancies attributed to infection with a pathogenic agent (Pagano *et al.*, 2004; Doocey *et al.*, 2022). Indeed, several oncogenic microorganisms such as *Helicobacter pylori*, *Fusobacterium nucleatum* and certain strains of *Escherichia coli* are currently under investigation because of their association with the development of CRC (Doocey *et al.*, 2022).

Independent studies have shown that tumours and mucosa of CRC patients are abnormally colonised by adherent *E. coli* strains belonging to the B2 phylogroup, compared to healthy controls (Swidsinski *et al.*, 1998; Raisch *et al.*, 2014). Concerningly, characterisation of these *E. coli* has revealed that most strains harbour the polyketide synthase (*pks*) locus responsible for the biosynthesis of the potent genotoxin colibactin (Buc *et al.*, 2013; Raisch *et al.*, 2014). In mammalian cells, colibactin alkylates host DNA causing DNA double-strand breaks and interstrand cross-links, which lead to cell-cycle arrest and cellular senescence (Bossuet-Greif *et al.*, 2018). Further, a recent breakthrough study identified a colibactin-specific mutational signature in CRC-derived tissues that directly implicates colibactin producing *E. coli* in the development of CRC (Dziubańska-Kusibab et al., 2020;

Pleguezuelos-Manzano et al., 2020). Thus, colibactin has been described as a *bona fide* virulence factor and *E. coli* producing this genotoxin represent an immediate threat to public health.

Recently, the effect of the human diet, and in particular, the role of amino acids, on bacterial gene expression has become an active topic of research. Therefore, as many colibactin producing *E. coli* reside in the gastrointestinal tract where they will encounter amino acids, it was considered that dietary amino acids could modulate the expression of colibactin. A comprehensive panel of L- and D-amino acids was screened in two *E. coli* strains harbouring the *pks* island, CFT073 and Nissle, and the expression level of *clbB* was determined by RT-qPCR. Intriguingly, both L- and D-amino acids were found to exhibit an inhibitory effect on *clbB* expression. However, notably, D-Serine was identified as the most consistent repressor with expression downregulated in CFT073 (3.12-fold) and Nissle (3.81-fold). Moreover, the similar fold changes observed in both strains suggested that there could be a conserved mechanism for *pks* regulation in response to D-Serine. In addition to encoding the *pks* island, both CFT073 and Nissle possess the *dsdCXA* locus, an enzymatic pathway that facilities the catabolism of D-Serine (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). Intriguingly, DsdC, the known regulator of this pathway, becomes activated in response to D-Serine. Therefore, it was hypothesised that DsdC could play a role in governing the regulatory pathway responsible for influencing the transcriptional changes observed in CFT073 and Nissle in response to D-Serine. However, exposure to D-Serine did not alter the inhibitory effects on colibactin production in an isogenic Δ*dsdC* mutant. Indeed, upon exposure to D-Serine, expression levels of *clbB* were downregulated and the genotoxic activity of colibactin was perturbed in HeLa cells infected with Nissle Δ*dsdC*; corresponding with the observations made in the WT Nissle in response to D-Serine. Although these findings were surprising, the results have revealed

that a more complex regulatory network may be at work and thus requires further investigating.

For over a century, Nissle has been used as the active ingredient of the probiotic pharmaceutical, Mutaflor®. Following its initial isolation from a healthy solider in 1917, Nissle has been considered as a commensal strain of *E. coli* and has been intensely studied for its unique ability to resolve gastrointestinal disorders. The probiotic activity exerted by Nissle is believed to be linked to the production of two antimicrobial peptides, microcin H47 (MccH47) and microcin M (MccM) (Sassone-Corsi *et al.*, 2016). However, it was recently demonstrated that the protective antimicrobial properties of Nissle cannot be dissociated from the genotoxic activity (Olier *et al.*, 2012), raising important questions concerning the safety of using this strain as an intestinal therapeutic. Indeed, the interplay between the genotoxic and antagonistic effects of Nissle demonstrates that there is a thin line between commensalism and pathogenicity in certain strains of bacteria, and further highlights the complex regulatory networks governing these activities. Intriguingly, this thesis revealed that exposure to D-Serine significantly downregulated expression of the *pks* island and inhibited colibactin-associated genotoxicity, whereas exposure to D-Serine did not alter the antimicrobial activity of Nissle. Thus, these data suggest that modulation of colibactin gene expression in response to D-Serine may involve a separate regulatory mechanism that does not affect the protective activities of Nissle. Therefore, dietary supplementation with D-Serine could present an opportunity to perturb the genotoxic activity associated with Nissle, whilst retaining the probiotic activity. The prophylactic potential of D-Serine has already been considered as a therapy for multiple neurological conditions, including post-traumatic stress disorder (Heresco-Levy *et al.*, 2009) and schizophrenia (MacKay *et al.*, 2019), thus demonstrating that D-Serine can be delivered to patients in safe and effective quantities. The recent correlation between *pks* harbouring *E. coli* and CRC has highlighted the need for the development of colibactin targeting

therapeutics. Therefore, deciphering the precise molecular mechanisms implicated in the colibactin-specific response to D-Serine will enhance our current understanding and aid in exploring its therapeutic potential to prevent colibactin-associated disease.

7 References

Abba, K., Sinfield, R., Hart, C.A and Garner, P. (2009). Pathogens associated with persistent diarrhoea in children in low and middle income countries: systematic review. *BMC Infectious Diseases*, 9, p 88. doi:10.1186/1471-2334-9-88.

Abe, A., de Grado, M., Pfuetzner, R.A., Sanchez-Sanmartin, C., Devinney, R, *et al.* (1999). Enteropathogenic *Escherichia coli* translocated intimin receptor, Tir, requires a specific chaperone for stable secretion. *Molecular Microbiology*, 33(6), pp. 1162–1175. doi:10.1046/J.1365-2958.1999.01558.x

Anfora, A.T., Haugen, B.J., Roesch, P., Redford, P and Welch, R.A. (2007). Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by *Escherichia coli* CFT073. *Infection and Immunity*, 75(11), pp. 5298–5304. doi:10.1128/IAI.00652-07.

Anfora, A.T and Welch, R.A. (2006). DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. *Journal of Bacteriology*, 188(18), pp. 6622–8. doi:10.1128/JB.00634-06.

Arthur, J.C., Perez-Chanona, E., Muhlbauer, M., Tomkovich, S., Uronis, J.M., *et al*. (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*, 338(6103), pp. 120-3. doi:10.1126/science.1224820.

Atiya Ali, M., Poortvliet, E., Stromberg, R., Yngve. (2011). Polyamines in foods: development of a food database. *Food & Nutrition Research*, 55. doi:10.3402/fnr.v55I0.5572.

Auvray, F., Perrat, A., Arimizu, Y., Chagneau, C.V., Bossuet-Greif, N., *et al*. (2021). Insights into the acquisition of the *pks* island and production of colibactin in the *Escherichia coli* population. *Microbial Geonomics*, 7(5), p. 000579. doi:10.1099/MGEN.0.000579.

Badea, L. Doughty, S., Nicholls, L., Sloan, J., Robins-Browne, R.M *et al*. (2003) Contribution of Efa1/LifA to the adherence of enteropathogenic *Escherichia coli* to epithelial cells. *Microbial Pathogenesis*, 34(5), pp. 205–215. doi:10.1016/S0882-4010(03)00026-3.

Baldelli, V., Scaldaferri, F., Putignani, L and Del Chierico, F. (2021). The role of *Enterobacteriaceae* in gut microbiota dysbiosis in Inflammatory Bowel Diseases. *Microorganisms*, 9(4), p. 697. doi:10.3390/microorganisms9040697.

Baldini, M.M., Kaper, J.P., Levine, M.M, Candy D.C and Moon, H.W. (1983). Plasmid-mediated adhesion in enteropathogenic *Escherichia coli*. *Journal of Paediatric Gastroenterology and Nutrition*, 2(3), pp. 534–538. doi:10.1097/00005176-198302030-00023

Baldwin, T.J., Ward, W., Aitkin, A., Knutton, S and Williams, P.H. (1991). Elevation of intracellular free calcium levels in HEp-2 cells infected with enteropathogenic *Escherichia coli*. *Infection and Immunity*, 59(5), pp. 1599–1604. doi:10.1128/IAI.59.5.1599-1604.1991.

Barletta, F., Ochoa, T.J., Mercado, E., Ruiz, J., Ecker, L., *et al*. (2011). Quantitative real-time polymerase chain reaction for enteropathogenic *Escherichia coli*: a tool for investigation of asymptomatic versus symptomatic infections. *Clinical Infectious Diseases,* 53(12), p. 1223-9. doi:10.1093/cid/cir730.

Berger, C.N., Sodha, S.V., Shaw, R.K., Griffin, P.M., Pink, D., *et al*. (2010), Fresh fruit and vegetables as vehicles for the transmission of human pathogens. *Environmental Microbiology*, 12(9), pp. 2385–2397. doi:10.1111/j.1462-2920.2010.02297.x.

Bernstein, C.N., Blanchard, J.F., Kliewer, E and Wajda, A. (2001). Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer*, 91(4), pp. 854-62. doi:10.1002/1097-0142(20010215)91:4<854::aid-cncr1073>3.0.co;2-z.

Bian, X., Fu, J., Plaza, A., Herrmann, J., Pistorius, D., *et al*. (2013). *In vivo* evidence for a prodrug activation mechanism during colibactin maturation. *Chembiochem*, 14(10), pp. 1194–1197. doi:10.1002/cbic.201300208.

Bielaszewska, M., Mellmann, A., Zhang, W., Kock, R., Fruth, A., *et al*. (2011). Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *The Lancet. Infectious Diseases*, 11(9), pp. 671–676. doi:10.1016/S1473-3099(11)70165-7.

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., *et al*. (1994) Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infection and Immunity*, 62(2), pp. 606–614. doi:10.1128/iai.62.2.606-614.1994.

Bolukaoto, J.Y., Singh, A., Alfinete, N and Barnard, T.G. (2021). Occurrence of hybrid diarrhoeagenic *Escherichia coli* associated with multidrug resistance in environmental water, Johannesburg, South Africa. *Microorganisms,* 9(10), p. 2163. doi:10.3390/MICROORGANISMS9102163.

Bonocora, R.P., Smith, C., Lapierre, P and Wade, J.T. (2015). Genome-scale mapping of *Escherichia coli* σ54 reveals widespread, conserved intragenic binding. *PLOS Genetics*, 11(10), p. e1005552. doi:10.1371/journal.pgen.1005552.

Bossuet-Greif, N., Vignard, J., Taieb, F., Mirey, G., Dubois, D., *et al*. (2018). The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *mBio*, 9(2), p.e02393-17. doi:10.1128/mBio.02393-17.

Bray, J. (1945). Isolation of antigenically homogeneous strains of *Bact. coli neapolitanum* from summer diarrhœa of infants. *The Journal of Pathology and Bacteriology*, 57(2), pp. 239–247. doi:10.1002/path.1700570210.

Brotherton, C.A., Wilson, M., Byrd, G and Balskus, E.P. (2015). Isolation of a metabolite from the *pks* island provides insights into colibactin biosynthesis and activity. *Organic. Letters*, 17(6), pp. 1545–1548. doi:10.1021/acs.orglett.5b00432.

Brotherton, C.A and Balskus, E.P. (2013). A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *Journal of the American Chemical Society,* 135(9), pp. 3359-62. doi:10.1021/ja312154m.

Brückner, H and Westhauser, T. (2003). Chromatographic determination of L- and D-amino acids in plants. *Amino acids*, 24(1–2), pp. 43–55. doi:10.1007/s00726-002-0322-8.

Buc, E., Dubois, D., Sauvanet, P., Raisch, J., Delmas, J., *et al.* (2013). High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS ONE*, 8(2), p. e56964. doi:10.1371/journal.pone.0056964.

Cao, H., Krishnan, G., Goumnerov, B., Tsongalis, J., Tompkins, R., *et al.* (2001). A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. *Proceedings of the National Academy of Sciences*, 98(25), pp. 14613–14618. doi:10.1073/pnas.251465298.

Cava, F. *et al.* (2011). Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cellular and Molecular Life Sciences*, pp. 817–31. doi:10.1007/s00018-010-0571-8.

Cepeda-Molero, M., Berger, C.N., Walsham, A.D.S., Ellis, S.J., Wemyss-Holden, S., *et al.* (2017). Attaching and effacing (A/E) lesion formation by enteropathogenic *E. coli* on human intestinal mucosa is dependent on non-LEE effectors. *PLoS Pathogens*, 13(10). P. e1006706. doi:10.1371/journal.ppat.1006706.

Chagneau, C.V., Garcie, C., Bossuet-Greif, N., Tronnet, S., Brachmann, A.O., *et al.* (2019) The polyamine spermidine modulates the production of the bacterial genotoxin colibactin," *mSphere*, 4(5), p. e00414-19. doi:10.1128/mSphere.00414-19.

Chagneau, C.V., Massip, C., Bossuet-Greif, N., Fremez, C., Motta, J.P., *et al.* (2021) Uropathogenic *E. coli* induces DNA damage in the bladder. *PLoS Pathogens*, 17(2), p. e1009310. doi:10.1371/journal.ppat.1009310.

Chang, J.W., Sato, Y., Ogawa, T., Arakawa, T., Fukai, S., *et al.* (2018). Crystal structure of the central and the C-terminal RNase domains of colicin D implicated its translocation pathway through inner membrane of target cell. *The Journal of Biochemistry*, 164(5), pp. 329–339. doi:10.1093/jb/mvy056.

Chen, H.D and Frankel, G. (2005). Enteropathogenic *Escherichia coli*: unravelling pathogenesis. *FEMS Microbiology Reviews*, 29(1), pp. 83–98. doi:10.1016/j.femsre.2004.07.002.

Chen, S.L., Wu, M., Henderson, J.P., Hooton, T.M., Hibbing, M.E., *et al.* (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). doi:10.1126/scitranslmed.3005497.

Choi, H.H., and Cho, Y.S. (2016). Fecal microbiota transplantation: current applications, effectiveness, and future perspectives. *Clinical Endoscopy*, 49(3), pp. 257–265. doi:10.5946/ce.2015.117.

Clermont, O., Christenson, J.K., Denamur, E and 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), pp. 58–65. doi:10.1111/1758-2229.12019.

Connolly, J.P.R., Goldstone, R.J., Burgess, K., Cogdell, R.J and Beatson, S.A. (2015). The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *ISME Journal*, 9(4), pp. 1039–1051. doi:10.1038/ismej.2014.242.

Connolly, J.P.R., Turner N.C.A., Hallam, J.C., Rimbi, P.T., Flett, T., *et al.* (2021) D-Serine induces distinct transcriptomes in diverse *Escherichia coli* pathotypes. *Microbiology (Reading)*, 167(10). doi:10.1099/MIC.0.001093.

Cosloy, S.D. (1973). D-serine transport system in *Escherichia coli* K-12. *Journal of Bacteriology*, 114(2), pp. 679–684. doi:10.1128/jb.114.2.679-684.1973.

Cosloy, S.D and McFall, E. (1973). Metabolism of D-Serine in *Escherichia coli* K-12: mechanism of growth inhibition. *Journal of Bacteriology*, 114(2), p. 685-94. Doi:10.1128/jb.114.2.685-694.1973

Cougnoux, A., Gibold, L., Robin, F., Bubois, D., Pradel, N., *et al.* (2012). Analysis of structure–function relationships in the colibactin-maturating enzyme ClbP. *Journal of Molecular Biology*, 424(3–4), pp. 203–214. doi:10.1016/j.jmb.2012.09.017.

Cougnoux, A., Delmas, J., Gibold, L., Fais, T., Romagnoli, C., *et al.* (2016). Small-molecule inhibitors prevent the genotoxic and protumoural effects induced by colibactin-producing bacteria. *Gut*, 65(2), pp. 278–285. doi:10.1136/gutjnl -2014-307241.

Cravioto, A., Gross, R.J., Scotland, S.M and Rowe, B. (1979). An adhesive factor found in strains of *Escherichia coli* belonging to the traditional infantile enteropathogenic serotypes. *Current Microbiology*, 3, pp. 95–99.

Croxen, M.A and Finlay, B.B. (2009). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology,* 8(1), pp. 26–38. doi:10.1038/nrmicro2265.

Cuevas-Ramos, G., Petit, C.R., Marcq, I., Boury, M., Oswald, E., *et al.* (2010). *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), pp. 11537–11542. doi:10.1073/pnas.1001261107.

Cutler, S.A., Lonergan, S.M., Cornick, N., Johnson, A.K and Stahl, C.H. (2007). Dietary inclusion of colicin E1 is effective in preventing postweaning diarrhea caused by F18-positive *Escherichia coli* in pigs. *Antimicrobial Agents and Chemotherapy*, 51(11), p. 3830-5. doi:10.1128/AAC.00360-07.

Datsenko, K.A and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12), pp. 6640–5. doi:10.1073/pnas.120163297.

Deng, K and Hansen, E.J. (2003). A CdtA-CdtC complex can block killing of HeLa cells by Haemophilus ducreyi cytolethal distending toxin. *Infection and Immunity*, 71(11), pp. 6633–6640. doi:10.1128/IAI.71.11.6633-6640.2003.

Díaz-Pascual, F., Lempp, M., Nosho, K., Jeckel, H., Jo, J.K *et al.* (2021). Spatial alanine metabolism determines local growth dynamics of *Escherichia coli* colonies. *Elife*, 10, p. e70794. doi:10.1101/2021.02.28.433255.

Donnenberg, M.S and Kaper, J.B. (1992). Enteropathogenic *Escherichia coli*. *Infection and Immunity*, 60(10), pp. 3953–3961.

Doocey, C.M., Finn, K., Murphy, C and Guinane, C.M. (2022). The impact of the human microbiome in tumorigenesis, cancer progression, and biotherapeutic development. *BMC Microbiology*, 22(1), p. 53. doi:10.1186/s12866-022-02465-6.

Duong, V.T., Tu, L.T.P., Tuyen, H.T., Nhi, L.T.Q., Campbell, J.I., *et al.* (2020). Novel multiplex real-time PCR assays reveal a high prevalence of diarrhoeagenic *Escherichia coli* pathotypes in healthy and diarrhoeal children in the south of Vietnam. *BMC Microbiology*, 20(1), p. 192. doi:10.1186/s12866-020-01878-5.

Dziubańska-Kusibab, P.J., Berger, H., Battistini, F., Bouwman, B.A.M., Iftekhar, A., *et al.* (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nature Medicine*, 26(7), pp.1063-69. doi:10.1038/s41591-020-0908-2.

Elliott, S.J., Wainright, L.A., McDaniel, T.K., Jarvis, K.G, Deng, L.C., *et al.* (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Molecular Microbiology*, 28(1), pp. 1–4. doi:10.1046/j.1365-2958.1998.00783.x.

Faïs, T., Delmas, J., Barnich, N., Bonnet, R and Dalmasso, G. (2018). Colibactin: more than a new bacterial toxin. *Toxins (Basel)*, 10(4), p. 151. doi:10.3390/toxins10040151.

Finlay, B.B., Rosenshine, I., Donnenberg, M.S and Kaper, J.B. (1992). Cytoskeletal composition of attaching and effacing lesions associated with enteropathogenic *Escherichia coli* adherence to HeLa cells. *Infection and Immunity*, 60(6), pp. 2541–3. doi:10.1128/iai.60.6.2541-2543.1992.

Fischbach, M.A and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews*, 106(8), pp. 3468–96. doi:10.1021/cr0503097.

Foubister, V., Rosenshine, I., Donnenberg, M.S and Finlay, B.B. (1994). The *eaeB* gene of enteropathogenic *Escherichia coli* is necessary for signal transduction in epithelial cells. *Infection and Immunity*, 62(7), pp. 3038–40. doi:10.1128/iai.62.7.3038-3040.1994.

Foubister, V., Rosenshine, I. and Finlay, B.B. (1994). A diarrheal pathogen, enteropathogenic *Escherichia coli* (EPEC), triggers a flux of inositol phosphates in infected epithelial cells. *Journal of Experimental Medicine*, 179(3), pp. 993–8. doi:10.1084/jem.179.3.993.

Francescangeli, F., de Angelis, M.L. and Zeuner, A. (2019). Dietary factors in the control of gut homeostasis, intestinal stem cells, and colorectal cancer. *Nutrients*, 11(12), p. 2936. doi:10.3390/nu11122936.

Frankel, G., Phillips, A.D., Rosenshine, I., Dougan, G., Kaper, J.B., *et al.* (1998) Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Molecular Microbiology*, 30(5), pp. 911–21. doi:10.1046/j.1365-2958.1998.01144.x.

Kauffmann, F (1947). The serology of the *coli* group. *The Journal of Immunology*, 57(1), pp. 71-100.

Garrett, W.S., Gordon, J.I. and Glimcher, L.H. (2010). Homeostasis and inflammation in the intestine. *Cell*, 140(6), p. 859-70. doi:10.1016/j.cell.2010.01.023.

Genchi, G. (2017). An overview on d-amino acids. (2017). *Amino Acids*, 49(9), pp. 1521–1533. doi:10.1007/S00726-017-2459-5.

Gerner, E.W., Bruckheimer, E. and Cohen, A. (2018). Cancer pharmacoprevention: targeting polyamine metabolism to manage risk factors for colon cancer. *The Journal of Biological Chemistry*, 293(48), pp. 18770–78. doi:10.1074/jbc.TM118.003343.

Gruenheid, S., DeVinney, R., Bladt, F., Goosney, D., Gelkop, S., *et al.* (2001). Enteropathogenic *E. coli Tir binds Nck* to initiate actin pedestal formation in host cells. *Nature Cell Biology*, 3(9), pp. 856-9. doi:10/1038/ncb0901-856.

Guerra, B., Junker, E., Schroeter, A., Helmuth, R., Guth, B.E.C., *et al.* (2006). Phenotypic and genotypic characterization of antimicrobial resistance in *Escherichia coli* O111 isolates. *Journal of Antimicrobial Chemotherapy*, 57(6), pp. 1210–4. doi:10.1093/jac/dkl127.

Guntaka, N.S., Healy, A.R., Crawford, J.M., Herzon, S.D., Bruner, S.D. (2017). Structure and functional analysis of ClbQ, an unusual intermediate-releasing thioesterase from the colibactin biosynthetic pathway. *ACS Chemical Biology*, 12(10), pp. 2598–2608. doi:10.1021/acschembio.7b00479.

Hancock, V., Dahl, M., and Klemm, P. (2010). Probiotic *Escherichia coli* strain Nissle 1917 outcompetes intestinal pathogens during biofilm formation. *Journal of Medical Microbiology*, 59(Pt 4), pp. 392–99. doi:10.1099/jmm.0.008672-0.

Hancock, V., Vejborg, R.M. and Klemm, P. (2010). Functional genomics of probiotic *Escherichia coli* Nissle 1917 and 83972, and UPEC strain CFT073: comparison of

transcriptomes, growth and biofilm formation. *Molecular Genetics and Genomics,* 284(6), pp. 437–54. doi:10.1007/s00438-010-0578-8.

Hao, R., Qiu, S., Wang, Y., Yang, G., Su, W., *et al.* (2012). Quinolone-resistant *Escherichia coli* O127a:K63 serotype with an extended-spectrum-beta-lactamase phenotype from a food poisoning outbreak in China. *Journal of Clinical Microbiology*, 50(7), pp. 2450–2451. doi:10.1128/JCM.00276-12.

Hashimoto, A., Nishikawa, T., Konno, R., Niwa, A., Yasumura, Y., *et al.* (1993). Free D-serine, D-aspartate and D-alanine in central nervous system and serum in mutant mice lacking D-amino acid oxidase. *Neuroscience Letters*, 152(1–2), pp. 33–60. doi:10.1016/0304-3940(93)90476-2.

Heresco-Levy, U., Vass, A., Bloch, B., Wolosker, H., Dumin, E., *et al.* (2009). Pilot controlled trial of D-serine for the treatment of post-traumatic stress disorder. *The International Journal of Neuropsychopharmacology*, 12(9), pp. 1275–82. doi:10.1017/S1461145709000339.

Hering, N.A., Richter, J.F., Fromm, A., Wieser, A., Hartmann, S., *et al.* (2014). TcpC protein from *E. coli* Nissle improves epithelial barrier function involving PKCζ and ERK1/2 signalling in HT-29/B6 cells. *Mucosal Immunology*, 7(2), pp. 369–378. doi:10.1038/mi.2013.55.

Hernández-Lucas, I., Gallego-Hernández, A.L., Encarnacion, S., Fernández-Mora, M., Martinez-Batallar, A.G., *et al.* (2008). The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica serovar Typhi. Journal of Bacteriology*, 190(5), pp. 1658–1670. doi:10.1128/JB.01649-07.

Heroven, A.K., and Dersch, P. (2006). RovM, a novel LysR-type regulator of the virulence activator gene *rovA*, controls cell invasion, virulence and motility of *Yersinia pseudotuberculosis*. *Molecular Microbiology*, 62(5), pp. 1469–1483. doi:10.1111/j.1365-2958.2006.05458.x.

Hodges, K., Alto, N.M., Ramaswamy, K., Dudeja, P.K. and Hecht, G. (2008). The enteropathogenic *Escherichia coli* effector protein EspF decreases sodium hydrogen exchanger 3 activity. *Cellular Microbiology*, 10(8), pp. 1735–1745. doi:10.1111/j.1462-5822.2008.01163.x.

Holmes, A., Muhlen, S., Roe, A.J. and Dean, P. (2010). The EspF effector, a bacterial pathogen's swiss army knife. *Infection and Immunity*, 78(11), pp. 4445–4453. doi:10.1128/IAI.00635-10

Homburg, S., Oswald, E., Hacker, J. and Dobrindt, U. (2007). Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli. FEMS Microbiology Letters*, 275(2), pp. 255-62. doi:10.1111/j.1574-6968.2007.00889.x.

Hu, J. and Torres, A.G. (2015). Enteropathogenic *Escherichia coli:* foe or innocent bystander? *Clinical Microbiology and Infection*, 21(8), pp. 279-34. doi:10.1016/j.cmi.2015.01.015

Hueck, C.J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews*, 62(2), p.p 379-433. doi:10.1128/MMBR.62.2.379-433.1998.

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., *et al.* (2009). Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *Journal of Bacteriology*, 191(1), pp. 347–354. doi:10.1128/JB.01238-08.

Ingle, D.J., Tauschek, M., Edwards, D.J., Hocking, D.M., Pickard, D.J., *et al*. (2016). Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nature Microbiology*, 1(2), pp. 1–9. doi:10.1038/nmicrobiol.2015.10.

Iwata, Y., Sakai, N., Yoneda, I., Senda, Y., Sakai-Takemori, Y., *et al.* (2021). D-Serine inhibits the attachment and biofilm formation of methicillin-resistant *Staphylococcus aureus*. *Biochemical and Biophysical Research Communications*, 537, pp. 50–56. doi:10.1016/j.bbrc.2020.12.078.

Jakes, K.S. and Finkelstein, A. (2010). The colicin Ia receptor, Cir, is also the translocator for colicin Ia. *Molecular Microbiology*, 75(3), pp. 567-78. doi:10.1111/j.1365-2958.2009.06966.x.

Jarvis, K.G., Giron, J.A., Jerse, A.E., McDaniel, T.K., Donnenberg, M.S., *et al.* (1995). Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. *Proceedings of the National Academy of Sciences (USA),* 92(17), pp. 7996–8000. doi:10.1073/pnas.92.17.7996.

Johnson, C.H., Dejea, C.M., Edler, D., Hoang, L.T., Santidrian, A.F., *et al.* (2015) Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metabolism*, 21(6), pp. 891–7. doi:10.1016/j.cmet.2015.04.011.

Kaewkod, T., Tobe, R., Tragoolpua, Y. and Mihara, H. (2021). Medicinal plant extracts protect epithelial cells from infection and DNA damage caused by colibactin-producing *Escherichia coli*, and inhibit the growth of bacteria. *Journal of Applied Microbiology*, 130(3), pp. 769–785. doi:10.1111/jam.14817.

Kalman, D., Weiner, O.D., Goosney, D.L., Sedat, J.W., Finlay, B.B., *et al.* (1999) Enteropathogenic *E. coli* acts through WASP and Arp2/3 complex to form actin pedestals. *Nature Cell Biology*, 1(6), pp. 389-91. doi.10.1038/14087.

Kaper, J.B., Nataro, J.P. and Mobley, H.L.T. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology,* 2(2), pp. 123–140. doi:10.1038/nrmicro818.

Kenny, B., Lai, L.C., Finlay, B.B. and Donnenberg, M.S. (1996). EspA, a protein secreted by enteropathogenic *Escherichia coli*, is required to induce signals in epithelial cells. *Molecular Microbiology*, 20(2), pp. 313–323. doi:10.1111/j.1365-2958.1996.TB02619.x.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., *et al.* (1997) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell*, 91(4), pp. 511–520. doi:10.1016/s0092-8674(00)80437-7.

Kenny, B. and Finlay, B.B. (1995). Protein secretion by enteropathogenic *Escherichia coli* is essential for transducing signals to epithelial cells. *Proceedings of the National Academy of Sciences (USA),* 92(17), pp. 7991-5. doi.10.1073/pnas.92.17.7991.

Kim, P.M., Duan, X., Huang, A.S., Liu, C.Y., Ming, G.L., *et al.* (2010). Aspartate racemase, generating neuronal D-aspartate, regulates adult neurogenesis. *Proceedings of the National Academy of Sciences (USA)*, 107(7), pp. 3175–3179. doi:10.1073/pnas.0914706107.

Knutton, S., Baldwin, T., Williams, P.H. and McNeish, A.S. (1989). Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infection and Immunity*, 57(4), pp. 1290-8. Doi.10.1128/iai.57.4.1290-1298.1989.

Konaté, A., Dembele, R., Guessennd, N.K., Kouadio, F.K., Kouadio, I.K *et al.* (2017). Epidemiology and antibiotic resistance phenotypes of diarrheagenic *Escherichia coli* responsible for infantile gastroenteritis in Ouagadougou, Burkina Faso. *European Journal of Microbiology & Immunology*, 7(3), pp. 168-175. doi:10.1556/1886.2017.00014.

Kosari, F., Taheri, M., Moradi, A., Alni, R.H. and Alikhani, M.Y. (2020). Evaluation of cinnamon extract effects on *clbB* gene expression and biofilm formation in *Escherichia coli* strains isolated from colon cancer patients. *BMC Cancer,* 20(1), p. 267. doi:10.1186/s12885-020-06736-1.

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., *et al.* (2012). Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Research*, 22(2), pp. 292-8. doi:10.1101/gr.126573.111.

Kotloff, K.L., Nataro, J.P., Losonsky, G.A., Wasserman, S.S., Hale, T.L., *et al*. (1995). A modified *Shigella* volunteer challenge model in which the inoculum is administered with

bicarbonate buffer: clinical experience and implications for *Shigella* infectivity. *Vaccine*, 13(16), pp. 1488–1494. doi:10.1016/0264-410x(95)00102-7.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., *et al.* (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet*, 382(9888), pp. 209-22. doi:10.1016/S0140-6736(13)60844-2.

Kotlowski, R., Bernstein, C.N., Sepehri, S. and Krause, D. (2007). High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut*, 56(5), pp. 669-75. doi:10.1136/gut.2006.099796.

Krishnan, S., Alden, N. and Lee, K. (2015). Pathways and functions of gut microbiota metabolism impacting host physiology. *Current Opinion in Biotechnology*, 36, pp. 137–145. doi:10.1016/j.copbio.2015.08.015.

Kruis, W., Fric, P., Pokrotnieks, J., Lukas, M, Fixa, B., *et al*. (2004). Maintaining remission of ulcerative colitis with the probiotic *Escherichia coli* Nissle 1917 is as effective as with standard mesalazine. *Gut*, 53(11), pp. 1617–1623. doi:10.1136/gut.2003.037747.

Lam, H., Oh, D-C., Cava, F., Takacs, C.N., Clardy, J., *et al.* (2009). D-amino acids govern stationary phase cell wall remodelling in bacteria. *Science*, 325(5947), pp. 1552–1555. doi:10.1126/science.1178123.

Lane, M.C. and Mobley, H.L.T. (2007). Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney International*, 72(1), pp. 19–25. doi:10.1038/sj.ki.5002230.

Langendorf, C., Le Hello, S., Moumouni, A., Gouali, M., Mamaty, A.A., *et al.* (2015). Enteric bacterial pathogens in children with diarrhea in Niger: diversity and antimicrobial resistance. *PLOS ONE*, 10(3), p. e0120275. doi:10.1371/journal.pone.0120275.

Law, C.J., Penfold, C.N., Walker, D.C., Moore, G.R., James, R *et al.* (2003). OmpF enhances the ability of BtuB to protect susceptible *Escherichia coli* cells from colicin E9 cytotoxicity. *FEBS Letters*, 545(2–3), pp. 127–132. doi:10.1016/s0014-5793(03)00511-8.

Levine, M.M., Bergquist, E.J., Nalin, D.R., Waterman, DH., Hornick, R.B., *et al.* (1978). *Escherichia coli* strains that cause diarrhoea but no not produce heat-labile or heat-stable enterotoxins and are non-invasive. *The Lancet*, 1(8074), pp. 1119–1122. doi:10.1016/s0140-6736(78)90299-4.

Levine, M.M. and Edelman, R. (1984). Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: epidemiology and pathogenesis. *Epidemiology Reviews*, 6, pp. 31-51. doi.10.1093/oxfordjournal.epirev.a036274.

Li, Z.R., Li, J., Gu, J.P., Lai, J.Y.H., Duggan, B.M., *et al.* (2016). Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nature Chemical Biology*, 12(10), pp. 773-5. doi:10.1038/nchembio.2157.

Logue, C.M., Doetkott, C., Mangiamele, P., Wannemuehler, Y.M., Johnson, T.J., *et al.* (2012). Genotypic and phenotypic traits that distinguish neonatal meningitis-associated

Escherichia coli from fecal *E. coli* isolates of healthy human hosts. *Applied and Environmental Microbiology*, 78(16), pp. 5824–5830. doi:10.1128/AEM.07869-11.

Lopez, L.R., Bleich, R.M. and Arthur, J.C. (2021). Microbiota effects on carcinogenesis: initiation, promotion, and progression. *Annual Review of Medicine.* 72, pp. 243-261. doi:10.1146/annurev-med-080719-091604.

MacKay, M.-A.B, Kravtsenyuk, M., Thomas, R., Mitchell, N.D., Dursun, S.M., *et al.* (2019). D-Serine: potential therapeutic agent and/or biomarker in schizophrenia and depression? *Frontiers in Psychiatry*, 10, p. 25. doi:10.3389/fpsyt.2019.00025.

Maddocks, O.D.K., Short, A.J., Donnenberg, M.S., Bader, S. and Harrison, D.J. (2009). Attaching and effacing *Escherichia coli* downregulate DNA mismatch repair protein *in vitro* and are associated with colorectal adenocarcinomas in humans. *PloS One*, 4(5), p. e5517. doi:10.1371/journal.pone.0005517.

Maddocks, S.E. and Oyston, P.C. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (Reading)*, 154(Pt 12), pp. 3609–3623. doi:10.1099/mic.0.2008/022772-0.

Mao, D., Bushin, L.B., Moon, K., Wu, Y and Seyedsayamdost, M.R. (2017). Discovery of *scmR* as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264. *Proceedings of the National Academy of Sciences*, 114(14), pp. E2920–E2928. doi:10.1073/pnas.1619529114.

Marchès, O., Ledger, T.N., Boury, M., Ohara, M.S., Tu, X., *et al* (2003). Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks

cell cycle G2/M transition. *Molecular Microbiology*, 50(5), pp. 1553–1567. doi:10.1046/j.1365-2958.2003.03821.x.

Marcone, G.L., Rosini, E., Crespi, E. and Pollegioni, L. (2020). D-amino acids in foods. *Applied Microbiology and Biotechnology*, 104(2), pp. 555–574. doi:10.1007/s00253-019-10264-9.

Martin, H.M., Campbell, B.J., Anthony Hart, C., Mpofu, C., Nayar, M., *et al.* (2004). Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology*, 127(1), pp. 80–93. doi:10.1053/j.gastro.2004.03.054.

Massip, C., Branchu, P., Bossuet-Greif, N., Chagneau, C.V., Gaillard, D., *et al.* (2019). Deciphering the interplay between the genotoxic and probiotic activities of *Escherichia coli* Nissle 1917. *PLoS Pathogens*, 15(9), p. e1008029. doi:10.1371/journal.ppat.1008029.

Matsumoto, M., Kunisawa, A., Hattori, T., Kawana, S., Kitada, Y., *et al.* (2018). Free D-amino acids produced by commensal bacteria in the colonic lumen. *Scientific Reports,* 8(1), pp. 1–7. doi:10.1038/s41598-018-36244-z.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S. and Kaper, J.B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proceedings of the National Academy of Sciences (USA),* 92(5), pp. 1664–1668. doi:10.1073/pnas.92.5.1664.

McDaniel, T.K. and Kaper, J.B. (1997). A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli*

K-12. *Molecular Microbiology*, 23(2), pp. 399–407. doi:10.1046/j.1365-2958.1997.2311591.x.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., *et al.* (2018). CellProfiler 3.0: next-generation image processing for biology. *PLoS Biology*, 16(7), p. e2005970. doi:10.1371/journal.pbio.2005970.

Mellies, J.L., Elliot, S.J., Sperandio, V, Donnenberg, M.S., and Kaper, J.B. (1999). The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Molecular Microbiology*, 33(2), pp. 296–306. doi:10.1046/j.1365-2958.1999.01473.x.

Menezes-Garcia, Z., Kumar, A., Zhu, W., Winter, S.E. and Sperandio, V. (2020). L-Arginine sensing regulates virulence gene expression and disease progression in enteric pathogens. *Proceedings of the National Academy of Sciences (USA)*, 117(22), pp. 12387–12393. doi:10.1073/pnas.1919683117.

Metzler, D.E. and Snell, E.E. (1952). Deamination of serine. II. D-Serine dehydrase, a vitamin B6 enzyme from *Escherichia coli*. *The Journal of Biological Chemistry*, 198(1), pp. 363–73. doi:10.1016/s0021-9258(18)55590-3.

Moon, H.W., Whip, S.C., Argenzio, R.A., Levine, M.M. and Giannella, R.A. (1983). Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. *Infection and Immunity*, 41(3), pp. 1340-51. doi:10.1128/iai.41.4.1340-1351.1983.

Mousa, J.J., Yang, Y., Tomkovich, S., Shima, A., Newsome, R.C., *et al.* (2016). MATE transport of the *E. coli*-derived genotoxin colibactin. *Nature Microbiology,* 1, p. 15009. doi:10.1038/nmicrobiol.2015.9.

Muniesa, M., Hammerl, J.A., Hertwig, S., Appel, B. and Brussow, H. (2012). Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Applied and Environmental Microbiology*, 78(12), pp. 4065–4073. doi:10.1128/AEM.00217-12.

Nataro, J.P., Maher, K.O., Macjier, P. and Kaper, J.B. (1987). Characterization of plasmids encoding the adherence factor of enteropathogenic *Escherichia coli*. *Infection and Immunity*, 55(10), p. 2370-7. doi:10.1128/iai.55.10.2370-2377.1987.

Nataro, J.P. and Kaper, J.B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, 11(1), pp. 142–201. doi:10.1128/CMR.11.1.142.

Norregaard-Madsen, M., McFall, E. and Valentin-Hansen, P. (1995). Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. *Journal of Bacteriology*, 177(22), pp. 6456–61. doi:10.1128/jb.177.22.6456-6461.1995.

Nougayrede, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz., *et al.* (2006). *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells, *Science,* 313(5788), pp. 848-51. doi:20.2226/science.1127059.

Nougayrede, J.-P., Changeau, C.V., Motta, J-P., Bossuet-Greif, N., Belloy, M., *et al.* (2021) A toxic friend: genotoxic and mutagenic activity of the probiotic strain *Escherichia coli* Nissle 1917. *mSphere*, 6(4), p. e0062421. doi:10.1128/mSphere.00624-21.

Ochoa, T.J., Barletta, F., Contreras, C. and Mercado, E. (2008). New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(9), pp. 852–856. doi:10.1016/j.trstmh.2008.03.017.

Ogunbanwo, S.T., Sanni, A.I. and Onilude, A.A. (2004). Influence of bacteriocin in the control of *Escherichia coli* infection of broiler chickens in Nigeria. *World Journal of Microbiology and Biotechnology*, 20, pp. 51–56. doi:10.1023/B:WIBI.0000013311.43842.74.

O'Hara, A.M. and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7), pp. 688-93. doi:10.1038/sj.embor.7400731.

Olier, M., Marcq, I., Salvador-Cartier, C., Secher, T., Dobrindt, U., *et al.* (2012). Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity. *Gut Microbes*, 3(6), pp. 501–9. doi:10.4161/gmic.21737.

Oliero, M., Calve, A., Fragoso, G., Cuisiniere, T., Hajjar, R., *et al.* (2021). Oligosaccharides increase the genotoxic effect of colibactin produced by *pks*+ *Escherichia coli* strains. *BMC Cancer,* 21(1), p. 172. doi:10.1186/s12885-021-07876-8.

Orskov, I., Oskov, F, Jann, B. and Jann, K. (1977). Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriological Reviews*, 41(3), pp. 667-710. doi:10.1128/br.41.3.667-710.1977.

Pagano, J.S., Blaser, M., Buendia, M-A., Damania, B., Khalili, K., *et al.* (2004). Infectious agents and cancer: criteria for a causal relation. *Seminars in Cancer Biology*, 14(6), pp. 453–471. doi:10.1016/j.semcancer.2004.06.009.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A and Brown, P.O. (2007). Development of the human infant intestinal microbiota. *PLOS Biology*, 5(7), p. e177. doi:10.1371/journal.pbio.0050177.

Panatier, A., Theodosis, D.T., Mothet, J-P., Touquet, B., Pollegioni, L., *et al.* (2006). Glia-derived D-serine controls NMDA receptor activity and synaptic memory. *Cell*, 125(4), pp. 775–784. doi:10.1016/j.cell.2006.02.051.

Pawłowska, B. and Sobieszczańska, B.M. (2017). Intestinal epithelial barrier: the target for pathogenic *Escherichia coli*. *Advances in Clinical and Experimental Medicine,* 26(9), pp. 1437–1445. doi:10.17219/acem/64883.

Platenkamp, A. and Mellies, J.L. (2018). Environment controls LEE regulation in enteropathogenic *Escherichia coli*. *Frontiers in Microbiology*, 9, p. 1694. doi:10.3389/fmicb.2018.01694.

Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H.M., *et al.* (2020). Mutational signature in colorectal cancer caused by genotoxic *pks* + *E. coli*. *Nature,* 580(7802), pp. 269-273. doi:10.1038/s41586-020-2080-8.

Putze, J., Hennequin, C., Nougayrede, J-P., Zhang, W., Homburg, S., *et al.* (2009). Genetic structure and distribution of the colibactin genomic island among members of the family *Enterobacteriaceae*. *Infection and Immunity*, 77(11), pp. 4696–4703. doi:10.1128/IAI.00522-09.

Raisch, J., Buc, E., Bonnet, M., Sauvanet, P., Vazeille, E., *et al.* (2014). Colon cancer-associated B2 *Escherichia coli* colonize gut mucosa and promote cell proliferation. *World Journal of Gastroenterology,* 20(21), pp. 6560-72. doi:10.3748/wjg.v20.i21.6560.

Reeves, P. (1972). *The Bacteriocins in Nature*. In: The Bacteriocins. Molecular Biology Biochemistry and Biophysics, 11, pp. 81-86. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-46289-4_7.

Robins-Browne, R.M., Tokhi, A.M., Adams, L.M., Bennett-Wood, V., Moisidis, A.V., *et al.* (1994). Adherence characteristics of attaching and effacing strains of *Escherichia coli* from rabbits. *Infection and Immunity*, 62(5), pp. 1584–1592. doi:10.1128/iai.62.5.1584-1592.1994.

Roesch, P.L., Redford, P., Batchelet, S., Moritz, R.L., Pellet, S., *et al.* (2003). Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract. *Molecular Microbiology*, 49(1), pp. 55–67. doi:10.1046/j.1365-2958.2003.03543.x.

Rogakou, E.P., Boon, C., Redon, C. and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *The Journal of Cell Biology*, 146(5), pp. 905–16. doi:10.1083/jcb.146.5.905.

Rooks, M.G. and Garrett, W.S. (2016). Gut microbiota, metabolites and host immunity. *Nature Reviews Immunology,* 16(6), pp. 341–352. doi:10.1038/nri.2016.42.

Rosenshine, I., Donnenberg, M.S., Kaper, J.B. and Finlay, B.B. (1992). Signal transduction between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells: EPEC induces tyrosine phosphorylation of host cell proteins to initiate cytoskeletal rearrangement and bacterial uptake. *The EMBO Journal*, 11(10), pp. 3551–3560. doi:10.1002/j.1460-2075.1992.tb05438.x.

Rothbaum, R., McAdams, A.J., Giannella, R. and Partin, J.C. (1982). A clinicopathologic study of enterocyte-adherent *Escherichia coli*: a cause of protracted diarrhea in infants. *Gastroenterology*, 83(2), pp. 441–454. PMCID:7044882.

Sadecki, P.W., Balboa, S.J., Lopez, L.R., Kedziora, K.M. and Arthur, J.C. (2021). Evolution of polymyxin resistance regulates colibactin production in *Escherichia coli*. *ACS Chemical Biology*, 16(7), pp. 1243–1254. doi:10.1021/acschembio.1C00322.

Saldaña, Z., Erdem, A.L., Schuller, S., Okeke, I.N., Lucas, M., *et al.* (2009). The *Escherichia coli* common pilus and the bundle-forming pilus act in concert during the formation of localized adherence by enteropathogenic *E. coli*. *Journal of Bacteriology*, 191(11), pp. 3451–3461. doi:10.1128/JB.01539-08.

Salomon, R.A. and Farias, R.N. (1992). Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli*. *Journal of Bacteriology*, 174(22), pp. 7428–7435. doi:10.1128/jb.174.22.7428-7435.1992.

Santos, A.C.M, Santos, F.F., Silva, R.M. and Gomes, T.A.T. (2020). Diversity of hybrid- and hetero-pathogenic *Escherichia coli* and their potential implication in more severe diseases. *Frontiers in Cellular and Infection Microbiology*, 10, p. 339. doi:10.3389/fcimb.2020.00339.

Santos, A.S. and Finlay, B.B. (2015). Bringing down the host: enteropathogenic and enterohaemorrhagic *Escherichia coli* effector-mediated subversion of host innate immune pathways. *Cellular Microbiology*, 17(3), pp. 318–332. doi:10.1111/cmi.12412.

Sassone-Corsi, M., Nuccio, S-P., Liu, H., Hernandez, D., Vu, C.T.*, et al.* (2016). Microcins mediate competition among *Enterobacteriaceae* in the inflamed gut. *Nature*, 540(7632), pp. 280–283. doi:10.1038/nature20557.

Scaletsky, I.C.A., Michalski, J., Torres, A.G., Dulguer, M.V. and Kaper, J.B. (2005) Identification and characterization of the locus for diffuse adherence, which encodes a novel afimbrial adhesin found in atypical enteropathogenic *Escherichia coli*. *Infection and Immunity*, 73(8), pp. 4753–4765. doi:10.1128/IAI.73.8.4753-4765.2005.

Scaletsky, I.C.A., Silva, M.L.M. and Trabulsi, L.R. (1984). Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. *Infection and Immunity*, 45(2), pp. 534-6. doi:10.1128/iai.45.2.534-536.1984

Secher, T., Brehin, C. and Oswald, E. (2016). Early settlers: which *E. coli* strains do you not want at birth? *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 311(1), pp. 123–129. doi:10.1152/ajpgi.00091.2016.

Shimazaki, J., Furukawa, S., Ogihara, H. and Morinaga, Y. (2012). L-Tryptophan prevents *Escherichia coli* biofilm formation and triggers biofilm degradation. *Biochemical and Biophysical Research Communications*, 419(4), pp. 715–8. doi:10.1016/j.bbrc.2012.02.085.

Shrivastava, A.K., Kumar, S., Mohakud, N.K., Suar, M. and Sahu, P.S. (2017). Multiple etiologies of infectious diarrhea and concurrent infections in a pediatric outpatient-based screening study in Odisha, India. *Gut pathogens*, 9(1), p. 16. doi:10.1186/s13099-017-0166-0.

Singh, G., Maurya, S., DeLampasona, M.P. and Catalan, C.A.N. (2007). A comparison of chemical, antioxidant and antimicrobial studies of cinnamon leaf and bark volatile oils, oleoresins and their constituents. *Food and Chemical Toxicology: an international journal published for the British Industrial Biological Research Association*, 45(9), pp. 1650–1661. doi:10.1016/j.fct.2007.02.031.

Sobhani, I., Tap, J., Roudot-Thoraval-F., Roperch, J.P., Letulle, S., et a*l.* (2011). Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS ONE*, 6(1), p. e16393. doi:10.1371/journal.pone.0016393.

Staley, T.E., Jones, E.W. and Corley, L.D. (1969). Attachment and penetration of *Escherichia coli* into intestinal epithelium of the ileum in newborn pigs. *The American Journal of Pathology*, 56(3), p. 371.

Stiff, T., O'Driscoll, M., Rief, N., Iwabuchi, K., Lobrich, M., *et al*. (2004). ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. *Cancer Research*, 64(7), pp. 2390–2600. doi:10.1158/0008-5472.can-03-3207.

Stone, K.D., Zhang, H.Z., Carlson, L.K. and Donnenberg, M.S. (1996). A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. *Molecular Microbiology*, 20(2), pp. 325–337. doi:10.1111/j.1365-2958.1996.tb02620.x.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., *et al*. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249. doi:10.3322/caac.21660.

Swidsinski, A., Khilkin, M., Kerjaschki, D., Schreiber, S., Ortner, M., *et al.* (1998). Association between intraepithelial *Escherichia coli* and colorectal cancer. *Gastroenterology*, 115(2), pp. 281–286. doi:10.1016/S0016-5085(98)70194-5.

Swimm, A., Bommarius, B., Li, Y., Cheng, D., Reeves, P., *et al.* (2004). Enteropathogenic *Escherichia coli* use redundant tyrosine kinases to form actin pedestals. *Molecular Biology of the Cell*, 15(8), pp. 3520–3529. doi:10.1091/mbc.e04-02-0093.

Tabor, C.W. and Tabor, H. (1985). Polyamines in microorganisms. *Microbiological Reviews*, 49(1), pp. 81–99. doi:10.1128/mr.49.1.81-99.1985.

Teixeira, N.B., Rojas. T.C.G., da Silveira, W.D., Matheus-Guimaraes, C., Silva, N.P., *et al.* (2015). Genetic analysis of enteropathogenic *Escherichia coli* (EPEC) adherence factor (EAF) plasmid reveals a new deletion within the EAF probe sequence among O119 typical EPEC strains. *BMC Microbiology*, 15(1), p. 200. doi:10.1186/s12866-015-0539-9.

Terlizzi, M.E., Gribaudo, G. and Maffei, M.E. (2017). Uropathogenic *Escherichia coli* (UPEC) infections: virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial strategies. *Frontiers in Microbiology*, 8, p. 1566. doi:10.3389/fmicb.2017.01566.

Thapa, S.S. and Grove, A. (2019). Do global regulators hold the key to production of bacterial secondary metabolites? *Antibiotics (Basel)*, 8(4), p. 160. doi:10.3390/antibiotics8040160.

Trabulsi, L.R., Keller, R. and Gomes, T.A.T. (2002). Typical and atypical enteropathogenic *Escherichia coli*. *Emerging Infectious Diseases*, 8(5), pp. 508–513. doi:10.3201/eid0805.010385.

Trautman, E.P., Healy, A.R., Shine, E.E., Herzon, S.B and Crawford, J.M. (2017). Domain-targeted metabolomics delineates the heterocycle assembly steps of colibactin biosynthesis. *Journal of the American Chemical Society*, 139(11), pp. 4195–4201. doi:10.1021/jacs.7b00659.

Tree, J.J., Roe, A.J., Flockhart, A., McAteer, S.P., Xu, X., *et al.* (2011). Transcriptional regulators of the GAD acid stress island are carried by effector protein-encoding prophages and indirectly control type III secretion in enterohemorrhagic *Escherichia coli* O157:H7. *Molecular Microbiology*, 80(5), p. 1349-1365. doi:10.1111/j.1365-2958.2011.07650.x.

Tripathi, P., Shine, E.E., Healy, A.R., Kim, C.S., Herzon, S.B., *et al.* (2017). ClbS is a cyclopropane hydrolase that confers colibactin resistance. *Journal of the American Chemical Society*, 139(49), pp. 17719-17722. doi:10.1021/jacs.7b09971.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., *et al.* (2007). The human microbiome project. *Nature,* 449(7164), pp. 804–810. doi:10.1038/nature06244.

Turner, N.C.A. (2021). The role of the transcriptional regulator, DsdC, in *Escherichia coli* pathogenicity. University of Glasgow, United Kingdom.

Ukena, S.N., Singh, A., Dringenberg U., Engelhardt, R., Seidler, U., *et al.* (2007). Probiotic *Escherichia coli* Nissle 1917 inhibits leaky gut by enhancing mucosal integrity. *PLoS ONE*, 2(12), p. e1308. doi:10.1371/journal.pone.0001308.

Ulshen, M.H. and Rollo, J.L. (1980). Pathogenesis of *Escherichia coli* gastroenteritis in man — another mechanism, *New England Journal of Medicine*, 302(2), pp. 99–101. doi:10.1056/NEJM198001103020207.

UNICEF. (2016) *Unicef:* One is too many: ending child deaths from pneumonia and diarrhoea, [Online] Available at: https://scholar.google.com/scholar_lookup?title=One+Is+too+Many:+Ending+Child+Deaths+from+Pneumonia+and+Diarrhoea&author=The+United+Nations+International+Children%E2%80%99s+Emergency+Fund&publication_year=2016 [Accessed: November 22, 2021].

UNICEF and WHO. (2019). Progress on household drinking water, sanitation and hygiene I 2000-2017. Special focus on inequalities. [Online] Available at: https://washdata.org [Accessed: November 23, 2021].

Ursell, L.K., Metcalf, J.L., Parfrey, L.W. and Knight, R. (2012). Defining the human microbiome. *Nutrition Reviews*, 70 Suppl1 1(Suppl 1), pp. S38-44. doi:10.1111/j.1753-4887.2012.00493.x.

Vejborg, R.M., Friis, C., Hancock, V., Schembri, M.A and Klemm, P. (2010). A virulent parent with probiotic progeny: comparative genomics of *Escherichia coli* strains CFT073, Nissle 1917 and ABU 83972. *Molecular Genetics and Genomics: MGG*, 283(5), pp. 469–484. doi:10.1007/S00438-010-0532-9.

Vizcaino, M.I., Engel, P., Trautman, E. and Crawford, J.M. (2014). Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *Journal of the American Chemical Society,* 136(26), pp. 9244-9270. doi:10.1021/ja503450q.

Vizcaino, M.I. and Crawford, J.M. (2015). The colibactin warhead crosslinks DNA. *Nature Chemistry*, 7(5), pp. 411-417. doi:10.1038/nchem.2221.

Wallenstein, A., Rehm, N., Brinkmann, M., Selle, M., Bossuet-Greif, N., *et al.* (2020). ClbR is the key transcriptional activator of colibactin gene expression in *Escherichia coli*. *mSphere,* 5(4), p. e00591-20. doi:10.1128/mSphere.00591-20

Wang, H., Zhong, Z., Luo, Y., Cox, E. and Devriendt, B. (2019). Heat-stable enterotoxins of enterotoxigenic *Escherichia coli* and their impact on host immunity. *Toxins (Basel)*, 11(1), p. 24. doi:10.3390/toxins11010024.

Wernke, K.M., Xue, M., Tirla, A., Kim, C.S., Crawford, J.M., *et al.* (2020). Structure and bioactivity of colibactin. *Bioorganic & Medicinal Chemistry Letters*, 30(15), p. 127280. doi:10.1016/j.bmcl.2020.127280.

Wijetunge, D.S.S., Gongati, S., DebRoy, C., Kim, K.S., Couraud, P.O., *et al.* (2015). Characterizing the pathotype of neonatal meningitis causing *Escherichia coli* (NMEC). *BMC Microbiology*, 15, p. 211. doi:10.1186/S12866-015-0547-9

Wiles, T.J. and Mulvey, M.A. (2013). The RTX pore-forming toxin α-hemolysin of uropathogenic *Escherichia coli*: progress and perspectives. *Future Microbiology*, 8(1), pp. 73–84. doi:10.2217/fmb.12.131.

Williams, J.G., Roberts, S.E., Ali, M.F., Cheung, W.Y., Cohen., D.R., *et al*. (2007) Gastroenterology services in the UK. The burden of disease, and the organisation and delivery of services for gastrointestinal and liver disorders: a review of the evidence. *Gut*, 56 Suppl1 (Suppl 1), pp. 1–113. doi:10.1136/gut.2006.117598.

Wilson, M.R., Jiang, Y., Villalta, P.W., Stornetta, A., Boudreau, P.D., *et al.* (2019). The human gut bacterial genotoxin colibactin alkylates DNA. *Science*, 363(6428), p. eaar7785. doi:10.1126/science.aar7785.

Wong, A.R.C., Pearson, J.S., Bright, M.D., Munera, D., Robinson, K.S., *et al.* (2011) Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Molecular Microbiology*, 80(6), pp. 1420–1438. doi:10.1111/j.1365-2958.2011.07661.x.

Xue, M., Kim, C.S., Healy, A.R., Wernke, K.M., Wang, Z., *et al.* (2019). Structure elucidation of colibactin and its DNA cross-links. *Science*, 365(6457). p.eaax2685. doi:10.1126/SCIENCE.AAX2685.

Yamada, A., Komaki, Y., Komaki, F., Micic, D., Zullow, S., *et al.* (2018). Risk of gastrointestinal cancers in patients with cystic fibrosis: a systematic review and meta-analysis. *The Lancet Oncology*, 19(6), pp. 758–767. doi:10.1016/S1470-2045(18)30188-8.

Yang, S.C., Lin, C.H., Sung, C and Fang, J.Y. (2014). Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in Microbiology*, 5, p. 241. doi:10.3389/FMICB.2014.00241

Yang, T., Richards, E.M., Pepine, C.J. and Raizada, M.K. (2018). The gut microbiota and the brain-gut-kidney axis in hypertension and chronic kidney disease. *Nature Reviews. Nephrology*, 14(7), pp. 442–456. doi:10.1038/s41581-018-0018-2.

Yang, X., Bai., X., Zhang, J., Sun, H., Fu, S., *et al.* (2020). *Escherichia coli* strains producing a novel Shiga toxin 2 subtype circulate in China. *International Journal of Medical Microbiology*, 310(1), p. 151377. doi:10.1016/j.ijmm.2019.151377.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., *et al.* (2012). Human gut microbiome viewed across age and geography. *Nature*, 486(7402), pp. 222-270. doi:10.1038/nature11053.

de Zamaroczy, M., Mora, L., M., Lecuyer, A., Geli, V. and Buckingham, R.H. (2001) Cleavage of colicin D is necessary for cell killing and requires the inner membrane peptidase LepB. *Molecular Cell*, 8(1), pp. 159–168. doi:10.1016/s1097-2765(01)00276-3.

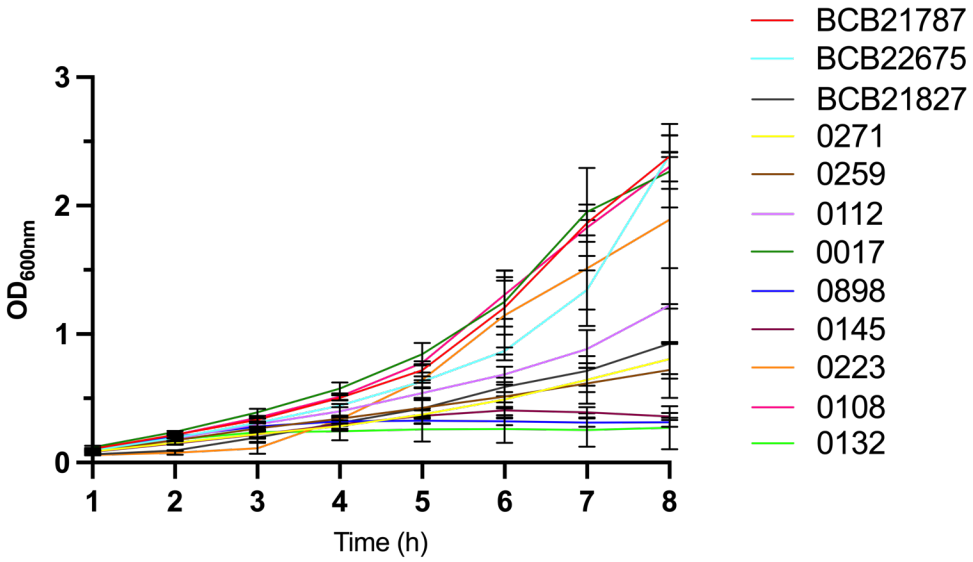
Zha, L., Wilson, M.R., Brotherton, C.A. and Balskus, E.P. (2016). Characterization of polyketide synthase machinery from the *pks* island facilitates isolation of a candidate precolibactin. *ACS Chemical Biology*, 11(5), pp. 1287–1295. doi:10.1021/acschembio.6B00014.

Zha, L., Jiang, Y., Henke, M.T., Wilson, M.R., Wang, J.X., *et al.* (2017). Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. *Nature Chemical Biology*, 13(10), pp. 1063-1065. doi:10.1038/nchembio.2448.

Zhang, S.X., Zhou, Y.M., Xu, W., Tian, L.G., Chen, J.X., *et al.* (2016). Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. *Infectious Diseases of Poverty*, 5(1), p. 64. doi:10.1186/s40249-016-0157-2.

Zhou, Y., Zhu, X., Hou, H., Lu, Y., Yu, J., *et al.* (2018). Characteristics of diarrheagenic *Escherichia coli* among children under 5 years of age with acute diarrhea: a hospital based study. *BMC Infectious Diseases*, 18(1), p. 63. doi:10.1186/s12879-017-2936-1.

8 Appendices



Appendix 1 Growth curve of clinical EPEC strains grown in DMEM media.

		WT vs WT +D-Serine		*ΔdsdC* vs Δ*dsdC* +D-Serine	
Feature ID	**Name**	**Fold Change**	**FDR Corrected p- value**	**Fold Change**	**FDR Corrected p- value**
c2450	clbS	1.34	0.53	-1.38	7.00E-02
c2451	clbQ	-1.28	0.54	-1.22	4.10E-01
c2452	clbP	-1.28	0.56	-1.48	1.00E-02
c2453	clbO	-1.67	5.10E-03	-1.57	1.02E-03
c2455	clbN	-1.59	0.01	-1.78	6.42E-06
c2456	clbM	-1.74	3.01E-03	-2.47	3.59E-11
c2457	clbL	-1.88	1.33E-03	-2.19	1.60E-08
c2458	clbK	-1.84	6.87E-04	-1.99	9.00E-08
c2459	clbJ	-1.69	2.31E-03	-2.26	1.62E-10
c2460	clbI	-2.16	5.06E-06	-3.22	0.00E+00
c2463	clbH	-2.05	3.08E-05	-2.66	2.91E-13
c2464	clbG	-2.48	1.55E-05	-3.65	4.75E-19
c2465	clbF	-17.84	0.4	-36.67	2.97E-03
c2466	clbE	-3.06	0.5	-3.5	1.00E-01
c2467	clbD	-2.19	5.31E-03	-2.74	1.06E-07
c2468	clbC	-2.29	1.55E-05	-2.82	1.96E-12
c2470	clbB	-2.38	1.08E-07	-2.72	0
c2471	clbR	-22.11	0.4	-1.1	1
c2472	clbA	-1.81	1	1.75	7.80E-01

Appendix 2 Differentially expressed genes comparing UPEC WT with UPEC WT + D-Serine and UPEC Δ*dsdC* with UPEC Δ*dsdC* + D-Serine by RNA-Seq.

Amino Acid	Relative *clbB* Expression to No Amino Acid Control in CFT073	Positive Error	Negative Error	*P*-Value	Relative Fold Change
D-Alanine	0.44	0.32	0.19	0.019	2.28
D-Arginine	0.84	0.04	0.04	0.153	1.19
D-Asparagine	0.64	0.22	0.16	0.053	1.55
D-Aspartic Acid	0.80	0.05	0.05	0.120	1.25
D-Cysteine	0.11	0.04	0.03	0.009	9.11
D-Glutamic Acid	0.26	0.15	0.09	0.007	3.84
D-Glutamine	0.47	0.17	0.12	0.011	2.11
D-Histidine	0.71	0.14	0.12	0.068	1.41
D-Isoleucine	0.44	0.02	0.02	0.023	2.29
D-Leucine	0.69	0.24	0.18	0.077	1.44
D-Lysine	0.91	0.18	0.15	0.298	1.10
D-Methionine	1.11	0.55	0.37	0.301	1.11
D-Phenylalanine	0.96	0.59	0.36	0.480	1.05
D-Proline	0.66	0.22	0.17	0.059	1.51
D-Serine	0.32	0.2	0.12	0.006	3.12
D-Threonine	0.62	0.16	0.13	0.038	1.62
D-Tryptophan	0.39	0.12	0.09	0.013	2.54
D-Tyrosine	0.43	0.02	0.02	0.022	2.34
D-Valine	1.53	1.01	0.61	0.158	1.53
Glycine	1.09	0.15	0.13	0.316	0.91
L-Alanine	0.94	0.48	0.32	0.462	1.07
L-Arginine	0.98	1.06	0.51	0.385	1.02
L-Asparagine	1.29	0.46	0.34	0.159	1.29
L-Aspartic Acid	0.13	0.14	0.07	0.003	7.52
L-Cysteine	0.71	0.66	0.34	0.310	1.40
L-Glutamic Acid	0.92	0.13	0.11	0.261	1.09
L-Glutamine	1.01	0.12	0.1	0.493	0.99
L-Histidine	0.65	0.13	0.11	0.045	1.55
L-Isoleucine	0.11	0.02	0.01	0.007	8.84
L-Leucine	0.28	0.09	0.07	0.005	3.61
L-Lysine	5.33	11.83	3.67	0.162	5.33
L-Methionine	1.45	0.86	0.54	0.170	1.45
L-Phenylalanine	0.25	0.08	0.06	0.007	3.94
L-Proline	0.76	0.32	0.22	0.161	1.32
L-Selenocysteine	0.16	0.05	0.04	0.010	6.17
L-Serine	1.04	0.08	0.08	0.448	0.97
L-Threonine	0.93	0.72	0.41	0.490	1.08
L-Tryptophan	0.84	0.37	0.26	0.271	1.19
L-Tyrosine	0.82	0.19	0.15	0.158	1.23
L-Valine	0.37	0.28	0.16	0.009	2.67

Appendix 3 Relative *clbB* expression in CFT073 treated with D- and L-amino acids by RT- qPCR.

Amino Acid	Relative *clbB* Expression to No Amino Acid Control in Nissle 1917	Positive Error	Negative Error	*P*-Value	Relative Fold Change
D-Alanine	0.42	0.23	0.15	0.050	2.39
D-Arginine	0.74	0.42	0.27	0.214	1.35
D-Aspartic Acid	0.56	0.54	0.28	0.150	1.77
D-Cysteine	0.48	0.39	0.21	0.076	2.07
D-Glutamic Acid	0.58	0.03	0.03	0.086	1.73
D-Glutamine	0.32	0.01	0.01	0.043	3.08
D-Isoleucine	2.49	0.39	0.34	0.005	2.49
D-Lysine	0.62	0.61	0.31	0.188	1.61
D-Methionine	0.30	0.37	0.17	0.037	3.35
D-Serine	0.26	0.08	0.06	0.036	3.81
D-Tryptophan	0.65	0.09	0.08	0.108	1.54
D-Tyrosine	0.32	0.02	0.02	0.043	3.08
L-Histidine	0.52	0.15	0.12	0.069	1.93
L-Leucine	0.32	0.02	0.02	0.042	3.12
L-Phenylalanine	1.01	0.56	0.36	0.479	1.01
L-Selenocysteine	0.18	0.08	0.05	0.029	5.64
L-Threonine	0.76	0.68	0.36	0.346	1.32

Appendix 4 Relative *clbB* expression in Nissle 1917 treated with D- and L-amino acids by RT- qPCR.