University of Glasgow

A molecular and biochemical analysis of a novel D-serine sensory system in *Escherichia coli* O157:H7

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

James PR Connolly BSc (Hons), MRes

Submitted February 2015

Institute of Infection, Immunity and Inflammation
College of Medical, Veterinary and Life Sciences
University of Glasgow

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 institution. 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.

James PR Connolly

February 2015

Acknowledgements

First and foremost, I would like to thank my supervisor Dr Andrew Roe for his constant support and enthusiasm throughout my PhD. I have learned more than I could have anticipated from him and appreciate his desire to help me develop into an independent scientist. I will fondly remember our conversations and the many great ideas that came to life over many great cups of coffee. Secondly, I wish to thank Professor Richard Cogdell, my secondary supervisor, for the help and advice he offered throughout my PhD.

My thanks go to the rest of the Roe group, past and present, for their help running the show. Special mention goes to Kate Beckham for putting up with me in the lab over the last three years. I also want to thank everyone in the GBRC who ever helped me out, in particular Rhys Grinter and the Walker group for being lifesavers with protein purification help and letting me steal reagents. I also have to thank Professor Jose Penades for his passionate and fruitful discussions regarding my work, they have been invaluable.

Furthermore, I would like to acknowledge my collaborators, both in Glasgow and further afield, for their enlightening discussions and contributions to the work presented in this thesis. In particular, thanks go to Professor David Smith, Dr Robert Goldstone, Dr Scott Beatson, Professor Waldemar Vollmer, Dr Gillian Douce and everyone at University of Glasgow Polyomics. I hope to collaborate again in the future. Also, a huge thanks to Professors David Gally, Rodney Welch, Brett Finlay, Steven Goodman and Kwang Sik Kim for graciously supplying bacterial strains and reagents.

Many thanks go to the Medical Research Council for funding my studentship and the University of Glasgow for hosting me in such an amazing environment.

It goes without saying that I owe a huge thanks to my family and friends, both in work and out, who have always supported me, in particular my parents Patrick and Mary, who always believed in me. Last, but by no means least, I need to thank Alys for her constant support over the last few years. You always keep me sane and make me laugh no matter how tough the day has been. Here's to the future…

Abstract

Escherichia coli is a diverse bacterial species found largely as a harmless commensal in the gastrointestinal tract of warm-blooded mammals. However, an array of highly adapted *E. coli* pathotypes have evolved over time capable of causing a variety of niche specific diseases, both intestinally and extraintestinally. This ability to cause disease results from the adaptation of the core genome to the host and the acquisition of horizontally acquired virulence factors. Furthermore, these foreign virulence factors are integrated into the regulatory network of the cell allowing niche specific competitive advantages for the emerged pathogen. Enterohaemorrhagic *E. coli* O157:H7 is a dangerous pathogen capable of causing haemorrhagic colitis and the potentially fatal haemolytic uremic syndrome. This pathogen is a harmless coloniser of ruminants whereas humans are extremely susceptible with transmission via the faecal-oral route, commonly associated with contaminated food products.

It is becoming apparent that the carriage of virulence factors is only one element to the adaptive nature of *E. coli* pathogens, with the specific response to niche specific signals governing when and where these virulence factors are utilised. O157:H7 utilise a type 3 secretion system (T3SS) as their major colonisation factor, which facilitates subversion of and subsequent intimate attachment to the human intestinal epithelium. This T3SS is encoded on a pathogenicity island known as the locus of enterocyte effacement (LEE) that encodes all the necessary components of a T3SS and a number of effector proteins. The LEE is tightly regulated in a temporal manner that is highly responsive to components of the intestinal physiology - namely nutrients, metabolites and hormone-like signals. Interestingly, despite the LEE encoded T3SS being capable of mediating attachment to a diverse variety of cell types *in vitro,* this colonisation factor is exclusively utilised by intestinal *E. coli* pathotypes but the mechanistic reasoning behind this is unknown.

In this work, the impact of the host metabolite D-serine on O157:H7 is explored in detail. D-serine is highly abundant in sites commonly colonised by extraintestinal *E. coli* pathotypes, such as the urinary tract and the brain. Furthermore, it acts as a positive fitness trait and regulator of virulence factors within these niches. Conversely, intestinal strains of *E. coli* are largely unable to metabolise D-serine but its effects on their gene

expression has not previously been investigated. Here, it is demonstrated that D-serine selectively affects gene expression in O157:H7 by repressing the LEE encoded T3SS and activating the SOS stress response. The toxicity of D-serine was entirely dependent on intracellular accumulation of this metabolite however the ability of D-serine to repress the LEE was found to be independent of its ability to be metabolised. Comparative genomic analysis revealed that carriage of both the LEE and the D-serine tolerance locus (*dsdCXA*) is an extremely rare event attributed to the apparent incompatibility between the two loci. It is proposed that the negative effects of D-serine on the LEE limit pathotypes such as O157:H7 to the gastrointestinal tract by forcing the evolutionary loss of *dsdCXA*, demonstrating the importance of co-operation between horizontally acquired and core pathogenic elements in defining niche specificity. A novel D-serine sensing system was also identified and characterised in O157:H7. This system includes a D-serine transporter, YhaO, and a LysR-type transcriptional regulator, YhaJ, which are absolutely required for expression of the LEE in O157:H7. This system is highly conserved in all *E. coli* and further demonstrates the adaptive ability of the core genome to perceive and respond to important environmental signals that define specific niches.

Collectively, this thesis describes the mechanistic basis of D-serine sensing in O157:H7 and the physiological relevance of D-serine sensing for diverse *E. coli* pathotypes. This work provides a strong framework for further research both by revealing novel insights into bacterial evolution and also creating potential targets for anti-bacterial therapeutics.

Table of contents

List of figures

List of tables

Abbreviations

Δ	Deletion
AA	Amino acid
A/E	Attaching and effacing
Amp	Ampicillin
AIEC	Adherent and invasive *E. coli*
ATP	Adenosine triphosphate
BBB	Blood brain barrier
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complimentary DNA
CFU	Colony forming units
Chl	Chloramphenicol
DAEC	Diffusely adherent *E. coli*
DAP	Diaminopimelate
D-AA	D-amino acid
D-ala	D-alanine
DEG	Differentially expressed gene
D-glu	D-glutamate
DIG	Digoxigenin
D-ser	D-serine
DNA	Deoxyribonucleic acid
Eae	Intimin
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
Epi	Epinepherine
Ery	Erythromycin
ETEC	Enterotoxigenic *E. coli*
ETT2	*E. coli* type 3 secretion system 2
EMSA	Electrophoretic mobility shift assay
ExPEC	Extraintestinal pathogenic *E. coli*
FDR	False discovery rate
Fis	Factor for inversion stimulation
g	Centrifugal force/Grams
GAD	Glutamate decarboxylase
Gb3	Globatrioaosylceramide 3
gDNA	Genomic DNA
GlcNAc	*N*-acetyl glucosamine
GrlA	Global regulator of *ler* activation
GrlR	Global regulator of *ler* repression
GSNO	Nitroso-glutathione
H-antigen	Flagellar antigen
H-NS	Histone like nucleoid structuring protein
HC	Haemorrhagic colitis
HGT	Horizontal gene transfer

His	Histidine
HPLC	High pressure liquid chromatography
HRP	Horseradish-peroxidase
HUS	Haemolytic uremic syndrome
IHF	Integration host factor
IM	Inner membrane
IMAC	Immobilised metal affinity ion chromatography
InPEC	Intestinal pathogenic *E. coli*
IPTG	β-D-1-thiogalactopyranoside
IRTKS	Insulin receptor tyrosine kinase substrate
Kan	Kanamycin
K-antigen	Capsular antigen
kb	Kilo base pairs
kDa	Kilo Dalton
l/L	Litre
L-AAs	L-amino acids
LEE	Locus of enterocyte effacement
Ler	LEE encoded regulator
LT	Lytic transglycosylase
LTTR	LysR type transcriptional regulator
lux	Luciferase
M	Molar
mM	Millimolar
MBP	Maltose binding protein
mg	Milligram
MGE	Mobile genetic element
ml	Milliliter
MMC	Mitomycin C
MNEC	Meningitis associated *E. coli*
mRNA	Messenger RNA
MS	Mass spectrometry
MurNAc	*N*-acetyl muramic acid
NAP	Nucleoid associated protein
ng	Nanogram
NCDAA	Non canonical D-amino acid
NEpi	Norepinephrine
NLE	Non LEE encoded effector
nM	Nanomolar
NO	Nitric oxide
NRIR	NLE regulatory inverted repeat
N-WASP	Neuronal Wiskott-Aldrich syndrome protein
O-antigen	Lipopolysaccharide antigen
OD^{600}	Optical density at 600 nm
OM	Outer membrane
ORF	Open reading frame
PAI	Pathogenicity Island
PBP	Penicillin binding protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde

PG	Peptidoglycan
qRT-PCR	Quantitative real time PCR
^{R}	Resistance
RFU	Relative fluorescence units
RLU	Relative luminescence units
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT-PCR	Reverse transcriptase PCR
SCFA	Short chain fatty acid
SEC	Size exclusion chromatography
sRNA	Small RNA
STEC	Shiga toxigenic *E. coli*
Stx	Shiga toxin
ssDNA	Single stranded DNA
T3S	Type 3 secretion
T3SS	Type 3 secretion system
TCA	Trichloroacetic acid
Tir	Translocated intimin receptor
tRNA	Transfer RNA
µg	Microgram
µl	Microliter
µM	Micromolar
UTI	Urinary tract infection
UPEC	Urinary pathogenic *E. coli*

1 Introduction

1.1 Escherichia coli

Escherichia coli was first discovered in 1885 by the German-Austrian scientist and paediatrician Theodor Escherich, under the name *Bacterium coli commune*, and it was not until 1954 that its now commonly used name was recognised (Cowan 1954, Escherich 1885). *E. coli* is widely regarded as a harmless commensal found in the gastrointestinal tract of humans and warm-blooded animals forming part of the natural intestinal flora soon after birth (Kaper *et al.,* 2004). A member of the *Enterobacteriaceae, E. coli* is a rod-shaped, Gram-negative bacterium. Key traits of this facultative anaerobe are that it grows ideally at 37°C, has the ability to ferment sugars and lactose, the latter in certain strains, and can be motile or non-motile depending on the presence of peritrichous flagella.

E. coli are typically typed using serological methods based on their lipopolysaccharide O-antigen, flagellar H-antigen, capsular K-antigen and the presence of surface fimbriae however not all strains can be classified in this way so alternative techniques, such as multilocus sequence typing, have been developed and used to classify strains based on genetic information (Croxen *et al.,* 2013). The dawn of next-generation sequencing has led to a wealth of information on the *E. coli* genome and thus, its evolution. No longer are researchers limited to phenotypic characterisation of strains, analysis based on a small number of housekeeping genes and reference only to a small number of sequenced genomes. Now, old and new isolates are routinely sequenced leading to an extremely detailed comparative view of the vast array of *E. coli* strains. At the time of this work more than 1500 sequenced *E. coli* genomes were publically available (Goldstone *et al.,* 2014).

1.2 Evolution of *E. coli* pathotypes

In terms of population genetics, *E. coli* can be divided into a number of phylogenetically distinct groups, or phylogroups for short (Figure 1-1). Recent studies have suggested that as many as seven phylogroups of *E. coli* (A, B1, B2, C, D, E and F) can be defined in the evolutionary timeline with the closely related *Shigella* species forming an eighth

accessory group (S) (Clermont *et al.,* 2013). The fine details of the *E. coli* phylogeny varies slightly from study to study depending largely on the method of analyses applied however this overall structure remains accepted as a framework for phylogenetic studies on *E. coli* (Clermont *et al.,* 2013, Ogura *et al.,* 2009, Sims & Kim 2011, Touchon *et al.,* 2009).

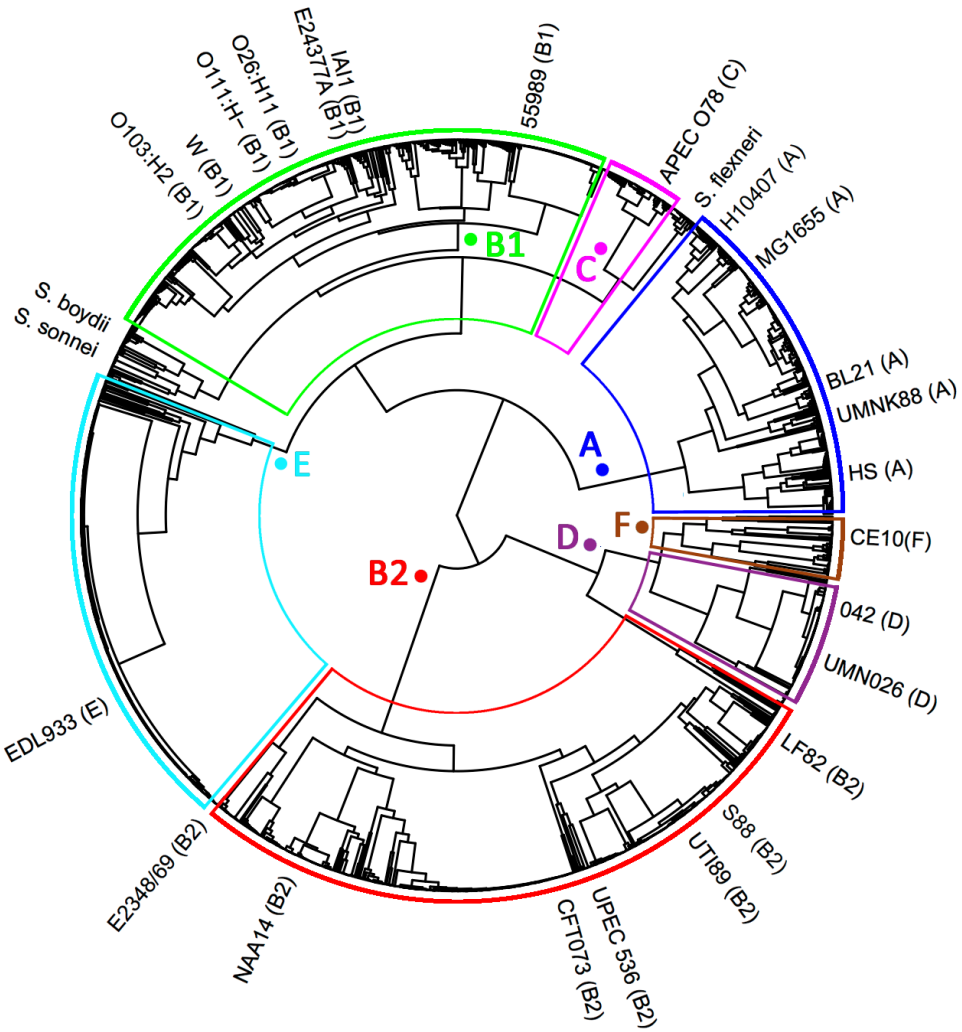

Figure 1-1 Phylogenomic tree of the *E. coli* lineage. Circularised maximum likelihood dendrogram built from a comparative analysis of the nucleotide sequences of 159 *E. coli* 'core genes' (Goldstone *et al.,* 2014). The sequences of 1591 *E. coli* and *Shigella sp.* isolates were obtained from NCBI as of 3/6/2014 and used in the analysis. Prototypical strains with their corresponding phylogroup are labelled according to their position on the tree. Phylogroups are colour coded: A = Blue; B1 = Green; B2 = Red; C = Violet; D = Purple; E = Cyan; F = Brown. The point at which each phylogroup diverged is indicated by a dot corresponding to the relevant colour group. Figure courtesy of Dr Robert Goldstone.

Such genetic diversity occurring in the *E. coli* species can be largely attributed to the plasticity of the *E. coli* genome (Touchon *et al.,* 2009). It is well known that bacteria are subject to widespread horizontal gene transfer (HGT) through the ability of being able to move DNA between bacterial hosts and that this plays a key role in the emergence of pathogenic strains (Croxen *et al.,* 2013). Individual and clusters of virulence genes (pathogenicity islands or PAI) are often encoded on mobile genetic elements (MGEs) such as plasmids, transposons, insertion sequences and bacteriophage. MGEs can either self replicate within the host or integrate into the host genome, often near tRNA genes, allowing expression of virulence specific loci within an otherwise commensal strain (Croxen & Finlay 2010). Non-virulence associated genes can also be carried on MGEs allowing bacteria to occupy new niches in a non-virulent manner as well. Genome minimalism can contribute to pathogen evolution, in the same respect as acquisition of MGEs, by optimizing the emerged pathogen towards a specific niche (Moran 2002). A recent study by Touchon *et al.* illustrated the impact HGT has on the plasticity of *E. coli* by highlighting that the average size of the *E. coli* genome is somewhere in the region of ~4700 genes, with the core genome being only ~2000 genes and the pan genome extending to ~18000 genes based on a number of diverse strains selected (Figure 1-2) (Touchon *et al.,* 2009).

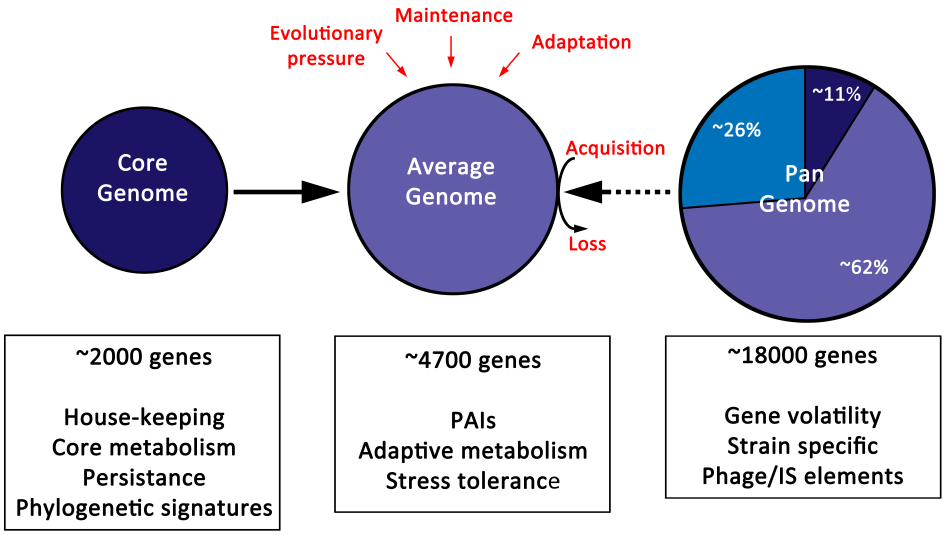

Figure 1-2 Genome plasticity of *E. coli*. Schematic representation illustrating an approximate size comparison size of the *E. coli* 'core', 'average' and 'pan' genomes. The core genome is comprised of so-called 'persistent' genes largely involved in housekeeping and core-metabolism. Acquisition or loss of genetic elements to facilitate evolutionary pressure, niche adaptation or maintenance results in an inflation of the genome. The 'pan' genome is comprised of 11% core elements, 62% persistent genes and a final 26% referred to as 'volatile' elements. This figure was adapted from Elsas *et al.,* 2010.

Over the course of history, numerous pathogenic *E. coli* have evolved via the mechanisms described above and group largely into a number of defined 'pathotypes' (Kaper *et al.,* 2004). *E. coli* pathotypes commonly lead to varying degrees of enteric pathogenesis. Included in this group of intestinal pathogenic *E. coli* (InPEC) are seven major pathotypes: shiga-toxigenic *E. coli* (STEC), enteropathogenic *E. coli* (EPEC), enteroaggregatative *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and adherent-invasive *E. coli* (AIEC). However, certain pathotypes of *E. coli* that have evolved the ability to passage through the intestinal tract and disseminate to extraintestinal niches (extraintestinal *E. coli* or ExPEC) are also prevalent causes of disease (Croxen & Finlay 2010). Urinary tract infections (UTI) are one of the most common infections in humans worldwide with significant rates of morbidity and mortality (Totsika *et al.,* 2012). Uropathogenic *E. coli* (UPEC) is the most common representative of the ExPEC and is the primary cause of UTI being responsible for the majority of all cases (Totsika *et al.,* 2012). An emerging ExPEC is meningitis associated *E. coli* (MNEC) that is capable of causing sepsis and meningitis, primarily in newborns, with high fatality rates (Croxen & Finlay 2010). These ExPEC have evolved the ability to cross the blood-brain barrier gaining access to the central nervous system of the host. In this work three pathotypes will be discussed. Primarily, enterohaemorrhagic *E. coli* (EHEC), a subset of the STEC group, will be introduced in section 1.2. Additionally, two members of the ExPEC, UPEC and MNEC, will also be introduced more briefly in section 1.3.

1.3 Enterohaemorrhagic *E. coli* – a subset of the STEC

EHEC is a subset of the STEC pathotype. STEC are broadly defined as *E. coli* pathotypes that harbour the Shiga toxin (Stx) 1 or 2 genes (stx_1 or stx_2) (Schüller 2011). STEC strains are diverse causing disease that ranges from mild diarrhoea to haemorrhagic colitis (HC) and haemolytic uremic syndrome (HUS) in extreme cases. Subsets of STEC vary in their serotype and virulence factor reservoir. Typically EHEC are a LEE-positive attaching and effacing (A/E) pathogen. The LEE (locus of enterocyte effacement) is a large PAI that encodes a type 3 secretion system (T3SS) and is responsible for the A/E lesion phenotype (McDaniel *et al.,* 1995). However, numerous serotypes are capable of causing HC and HUS

in a LEE-independent manner (Croxen *et al.,* 2013). For instance, the 2011 German outbreak of *E. coli* was caused by the O104:H4 strain, a largely EAEC-like pathogen that acquired the Shiga toxin, resulting in the emergence of a new hybrid STEC that caused over 900 cases of HUS and 50 deaths (Grad *et al.,* 2012). STEC strains appear to have evolved on multiple occasions and fall into four clonal lineages: EHEC 1, EHEC 2, STEC 1 and STEC 2 (Table 1-1).

Table 1-1 Clonal lineages of the STEC pathotype and associated serotypes.

STEC Lineage	Associated characteristics	Common serotypes
EHEC 1	LEE/O157-positive	O157:H7
EHEC 2	O157-negative; LEE-positive	O111:H8, O26:H11
STEC 1	LEE/O157-negative	O113:H21, O91:H21
STEC 2	LEE-positive; O157-negative	O45:H2, O103:H2/H6

*For details see Croxen *et al.,* 2013.

The most common EHEC serogroup associated with regular outbreaks is *E. coli* O157:H7, a LEE-positive, Shiga toxin producing STEC that falls into the EHEC 1 lineage. EHEC O157:H7 has been studied extensively as the prototypical EHEC and a vast array of literature describing its prevalence and molecular mechanisms of disease exists, which will be discussed in the following sections.

1.3.1 *Origin, occurrence and prevalence of E. coli O157:H7*

Riley *et al.* described the first recognised outbreaks of O157:H7 in the United States in 1983. They reported an unusual gastrointestinal illness defined by severe abdominal pain and initial watery diarrhoea culminating in extreme bloody diarrhoea. The outbreak affected 47 people in the Oregon and Michigan areas and was attributed to a rare isolate of *E. coli* that was found to have contaminated ground beef distributed to the same fast food chain in both locations (Riley *et al.,* 1983). An investigation by the Centre for Disease Control and Prevention (CDC) subsequently revealed that the O157:H7 serotype was indeed an emerging pathogen and its isolation was extremely rare prior to the 1970's.

Since this initial outbreak, O157:H7 has become known as a major cause of foodborne illness worldwide, particularly in the United States where in 2011 there were 463 cases reported and a further 521 cases associated with non-O157 STEC infections (FoodNet

2011 surveillance report; CDC). Although the incidence of O157:H7 cases has dropped in recent years, particularly in comparison to non-O157 STEC, it is still regarded as a major pathogen (Croxen *et al.,* 2013). Hospitalization rates and case-fatality rates are almost twice as high for O157:H7 in comparison to non-O157 STEC. Reports from the CDC have also highlighted the association of O157:H7 with fresh produce and not just contaminated meat products. In 2013, an outbreak associated with ready to eat salads infected 33 people in four US states (Arizona, California, Texas and Washington) resulting in seven hospitalizations and two cases of HUS. Cases associated with contaminated meat products also continue to be reported annually, most recently in 2014 when a multistate outbreak was associated with contaminated ground beef (www.cdc.gov). The European CDC 2013 epidemiological report highlighted the rise in O104 associated STEC cases since the 2011 outbreak, however O157:H7 still remains the major cause of STEC HUS across all age groups.

The major reservoir for *E. coli* O157:H7 is ruminants, primarily cattle, which the organism colonizes asymptomatically at the recto-anal junction (Naylor *et al.,* 2003). Ruminants are therefore regarded as the major shedding source for O157:H7 resulting in contamination of crop and water sources in the nearby farm area or meat during slaughter by direct contact with contaminated faeces. Cattle faeces have been reported to contain up to 10^6 colony-forming units (CFU) of O157:H7 per gram (Chase-Topping *et al.,* 2007). Illness can also be caused by direct contact with the infected animal or its faeces (Hale *et al.,* 2012).

1.3.2 *Pathogenesis of E. coli O157:H7*

E. coli O157:H7 is the most extensively studied EHEC, and STEC for that matter, but the carriage of similar virulence factors between members of other lineages and even *E.* coli pathotypes suggest shared mechanisms. The repertoire of virulence factors in O157:H7 includes a T3SS and its associated effector proteins, the Shiga toxin, the pO157 plasmid and the flagella (Nataro & Kaper 1998). Evidence for horizontal acquisition of virulence factors was highlighted in a study characterizing the complete genome sequence of *E. coli* O157:H7 (Perna *et al.,* 2001). Analysis of the genome sequence showed that it contains hundreds of introgressed segments or 'O-islands', many of which contain known virulence factors such as the LEE PAI encoding a T3SS. They also described genomic regions related to known bacteriophage, including BP-933W which encodes *stx*$_2$, and a selection of

cryptic prophage that encode various virulence related proteins but lack all necessary phage-like genes (Perna *et al.,* 2001).

1.3.3 *The LEE encoded type 3 secretion system*

E. coli O157:H7 forms A/E lesions on host epithelial cells characterised by destruction of brush border microvilli, intimate attachment, reorganisation of the host actin cytoskeleton and bacterial 'pedestal' formation (McDaniel *et al.,* 1995). This phenotype is not restricted to O157:H7 however. Other A/E pathogens include EPEC and the mouse pathogen *Citrobacter rodentium* (Schauer & Falkow 1993). The A/E phenotype has been attributed to the carriage of the large LEE PAI (Figure 1-3). This ~35 kb PAI contains all the necessary genes for the formation of a functional T3SS, as well seven secreted effector proteins, chaperones and the master regulators of the system (Wong *et al.,* 2011). Indeed, a study by McDaniel and Kaper described how they cloned the entire LEE from the EPEC strain E2348/69 on a plasmid and successfully used this to confer the A/E phenotype to the non-pathogenic K-12 strain (McDaniel & Kaper 1997). Genetically distinct T3SSs are widely used by a range of Gram-negative animal and plant pathogens (over 25 species including species of *Salmonella, Shigella, Yersinia, Burkholderia, Chlamydia* and *Pseudomonas*) to facilitate attachment to the host (Cornelis 2006). The LEE has a GC content of 38.3%, significantly lower than the 50.8% that makes up the *E. coli* chromosome, suggesting that the LEE was acquired by HGT (Frankel *et al.,* 1998, McDaniel & Kaper 1997). In O157:H7 the LEE is located at the *selC* tRNA locus and similarly in EPEC. This site is also the insertion point for a large PAI in UPEC suggesting an active hotspot for acquisition of virulence genes (Blum *et al.,* 1994). However, alternative insertion sites for the LEE have been identified suggesting acquisition of the LEE at multiple stages of the EHEC and EPEC evolutionary timeline (Frankel *et al.,* 1998).

The LEE contains 41 open reading frames (ORFs), which can be organized into five operons, namely LEE1 through LEE5 (Elliott *et al.,* 1998, Wong *et al.,* 2011). LEE1 encodes the master regulator Ler (LEE encoded regulator) through which activation of LEE2 to LEE5 is mediated (Haack *et al.,* 2003, Mellies *et al.,* 1999, Sánchez-SanMartín *et al.,* 2001). The second master regulator, GrlRA, is encoded between the LEE1 and LEE2 operons, forming a regulatory feedback loop on *ler* expression (Deng *et al.,* 2004). The *esc* genes encoding the basal T3SS machinery are located on LEE1/2/3/4, the Sep secretion

machinery are located on LEE2/3/4, the T3SS needle filament gene (*espA*) and the translocon pore genes (*espB/D*) are encoded on LEE4, the major adhesin intimin (*eae*) and its cognate 'translocated intimin receptor' (*tir*) are located on LEE5, the lytic transglycosylase (*etgA*) is located between LEE1 and LEE2, and numerous secreted effectors and chaperones are located throughout LEE1 through LEE5 (Elliott *et al.,* 1998, Wong *et al.,* 2011).

The structure of the T3SS is complex (Figure 1-4). The basal apparatus of the T3SS spans the inner membrane (IM), periplasm and outer membrane (OM) of the bacterial cell (Büttner 2012, Cornelis 2006). EscC, a secretin, forms an oligomeric pore structure in the OM (Burghout *et al.,* 2004). This is fused to EscJ, which forms the periplasmic portion of the basal apparatus (Crepin *et al.,* 2005). EscU, EscT, EscS and EscR form the multi-protein complex that spans the IM with EscJ composing the inner ring. EscV forms a cytoplasmic docking domain for substrates and SepQ forms the C-ring that surrounds the docking domain (Büttner 2012, Cornelis 2006). EscI is thought to form the inner rod of the basal apparatus (Sal-Man *et al.,* 2012). SepL and SepD are thought to act as gatekeepers for the T3SS prior to full assembly of the needle complex (Deng *et al.,* 2005). EscN is an ATPase that powers the T3SS machinery and interacts with SepQ and EscI (Zarivach *et al.,* 2007). The T3SS needle protein, EscF, forms a helical polymerised base extending from the OM outwards (Wilson *et al.,* 2001). EspA subunits then form the needle filament that extends to reach the host cell surface making contact via the T3SS translocon proteins, EspB and EspD. EspB and EspD form an active pore in the host cell membrane allowing translocation of effectors through the T3SS and into the host (Ide *et al.,* 2001).

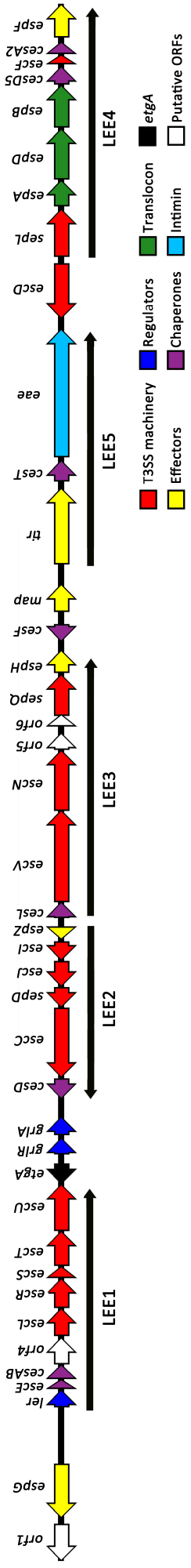

Figure 1-3 Genetic organisation of the LEE island from EHEC EDL933. Schematic representation of the entire locus of enterocyte effacement (LEE) from EHEC. Each of the 41 ORFs are labelled above the corresponding gene. Each gene is colour coded as indicated in the legend: Red = T3SS machinery *esc* and *sep* genes; Blue = Core regulators of the LEE; Green = the *esp* translocon genes; Black = *etgA*; Yellow = Secreted effectors; Purple = Chaperones; Cyan = Intimin; White = uncharacterised ORFs. The LEE is further divided into 5 operonic subsections, namely LEE1 through LEE5, and these are indicated by black arrows below the LEE island. This figure was adapted from Wong et al., 2011

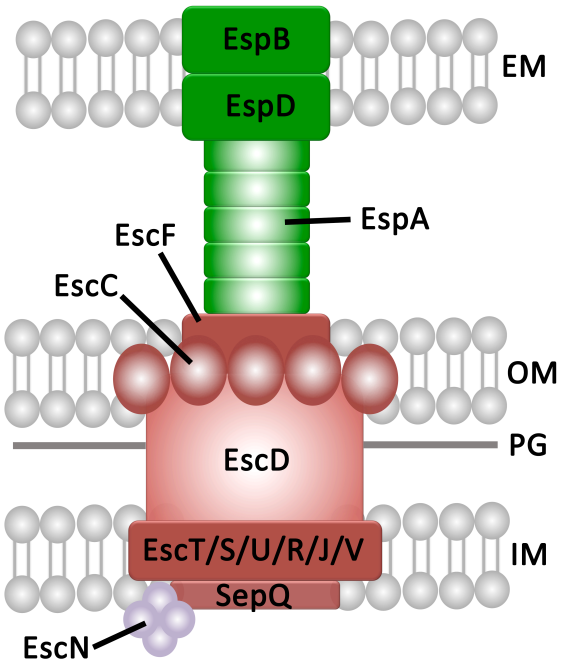

Figure 1-4 The T3SS injectisome. Schematic representation of the EHEC T3SS apparatus. The basal body that spans the bacterial inner membrane (IM), periplasmic peptidoglycan layer (PG) and outer membrane (OM) is illustrated in red. The EscN ATPase at the cytoplasmic side of the basal body is illustrated in purple whereas the needle filament, composed of EspA subunits, and the translocon pore, which contacts the eukaryotic host membrane (EM) are illustrated in green. This figure was adapted from Büttner 2012.

1.3.4 *Intimate attachment and translocation of effector proteins*

Initial attachment of O157:H7 to the intestinal epithelium is not fully understood. There are currently a number of fimbrial-like factors that potentially play a role in attachment but a definitive mechanism is not known (Croxen & Finlay 2010, Low *et al.,* 2006). Others include the *E. coli* common pilus, the more recently described haemorrhagic coli pilus and long polar fimbriae (Croxen *et al.,* 2013). The EHEC H7 flagella have also been implicated to play a role in initial attachment. In this study, the authors elegantly describe the adhesive properties of the H7 flagella and time-dependent expression of flagella during adherence to bovine epithelial cells, whereby at earlier time points attached bacteria express flagella but do not form A/E lesions and in the latter stages of infection expression of flagella is only seen on individually adhered bugs that have not formed A/E lesions. This study supports an efficient model of EHEC colonisation demonstrating a

controlled expression of virulence factors during infection (Mahajan *et al.,* 2009). The H7 flagella is also known to be highly immunogenic stimulating a strong toll-like receptor 5 response from the immune system (Hayashi *et al.,* 2001). Temporal expression of flagella may also be a mechanism of modulating the immune system during intimate attachment.

As described in section 1.2.3, the LEE encodes the major adherence factor Intimin and its receptor Tir (Kenny *et al.,* 1997). Once the EspB/D translocon has formed, Tir is translocated to the host cell cytoplasm via the T3SS where it relocates to the epithelial cell membrane, exposing itself to the cell surface. Here, it can interact with Intimin and ultimately lead to tight attachment (Figure 1-5). Tir has also been shown to play a role in A/E pedestal formation by recruiting cytoskeletal proteins to the site of attachment. Briefly, Tir becomes linked to insulin receptor tyrosine kinase substrate (IRTKS), which binds another translocated effector protein $EspF_U$ (also known as Tir cytoskeleton-coupling protein or TccP) (Campellone *et al.,* 2004, Weiss *et al.,* 2009). $EspF_U$ then interacts with Neuronal Wiskott-Aldrich syndrome protein (N-WASP) activating localized actin assembly resulting in pedestal formation (Sallee *et al.,* 2008).

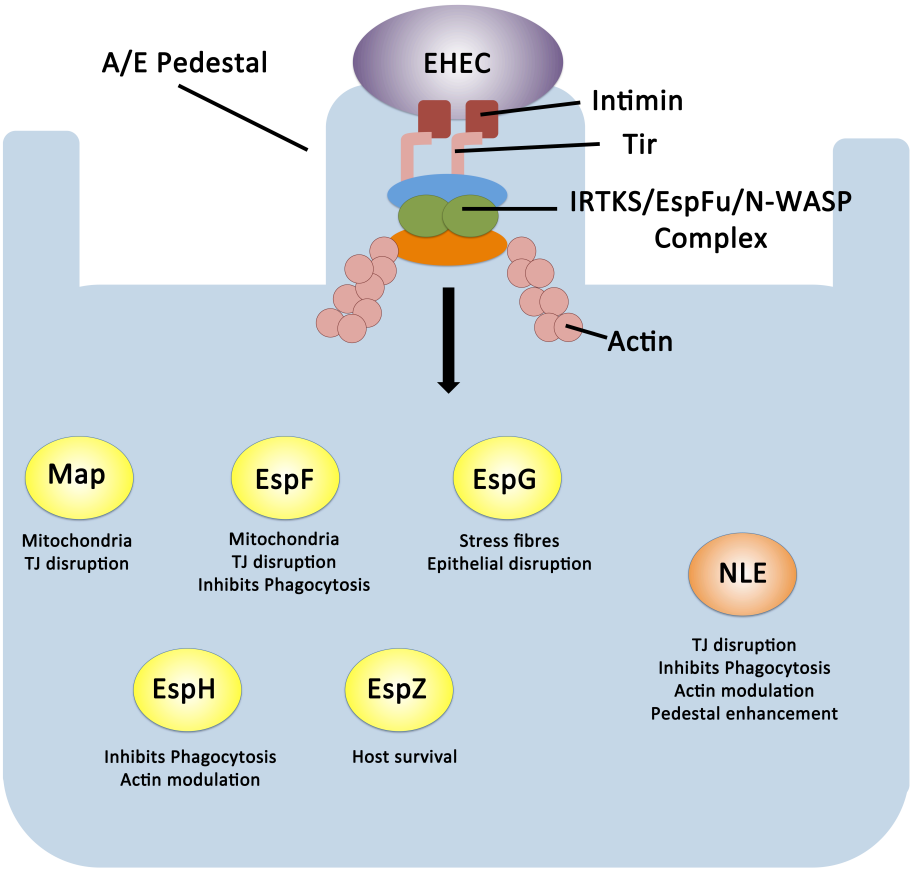

Figure 1-5 Pathogenesis of EHEC via A/E lesion formation. Schematic illustration of EHEC forming an attaching and effacing (A/E) lesion on the host epithelium (blue). EHEC translocates Tir into the host cell, which acts as the receptor for the surface adhesion Intimin. Tir forms a complex with the host cell factors IRTKS and N-WASP as well as another translocated effector protein named $EspF_U$, subverting the normal function of host cell actin and reorganising the cytoskeleton to form a pronounced pedestal upon which the bacterial cell resides. Numerous effector proteins are subsequently translocated via the T3SS and carry out multifactorial subversion of the host cell. LEE encoded effectors and their functions are highlighted in yellow whereas the non-LEE encoded effector (NLE) "suite" functions are highlighted by the orange NLE illustration. This figure was adapted from Croxen and Finlay 2010.

The LEE encodes 6 other effector proteins (Map, EspF, EspG, EspZ, EspH and EspB) that play a part in the formation of A/E lesions. These effector proteins have been the subject of many studies in EPEC and EHEC and often play multifactorial roles in host cell modulation. Map localizes to the mitochondria and mediates disruption of intestinal tight junctions (Dean & Kenny 2004, Kenny & Jepson 2000). EspF targets the mitochondria also but has other roles in disruption of tight junctions (Marchès *et al.,* 2006). EspG induces the formation of actin stress fibres and disrupts the epithelium, whereas EspZ promotes

survival of the host cell (Dean *et al.,* 2010, Shames *et al.,* 2010). EspF, EspB and EspH are all involved in inhibition of phagocytosis as well as having other important roles in modulation of the actin cytoskeleton (Dong *et al.,* 2010, Marchès *et al.,* 2008). LEE encoded effectors are listed in Table 1-2.

As discussed above, $EspF_U$ is translocated into the host cell via the T3SS but this protein is not encoded on the LEE, instead its gene is found on the cryptic prophage CP-933U and is known as a non-LEE encoded effector protein (NLE) (Campellone *et al.,* 2004). The repertoire of NLEs in *E. coli* and *C. rodentium* is extensive with a recent report summarising that EPEC and *C. rodentium* encoded 19 and 28 NLEs respectively, whereas EHEC encoded a staggering 55 NLEs (Table 1-2) (Tobe *et al.,* 2006). Not all of these are fully characterised however with a number of pseudogenes being present in each pathotype (6, 12 and 6 respectively for EPEC, EHEC and *C. rodentium*). The genomic context of NLEs is largely associated with lambdoid prophage, the majority of which are found on horizontally acquired genetic islands scattered throughout the genome. Indeed, this comprehensive study by Tobe *et al.* concluded that the likely major function of lambdoid prophage in EHEC was to carry T3S effectors (Tobe *et al.,* 2006). Functionally, NLEs are diverse with confirmed roles including inhibition of phagocytosis, disruption of tight junctions, actin reorganisation and host immune modulation, which have been summarised elsewhere in detail (Wong *et al.,* 2011).

Table 1-2. List of all known secreted effector proteins in EHEC.

Genomic Context	Name
LEE-encoded	EspB, Tir, Map, EspF, EspG, EspH, EspZ
Non-LEE encoded	EspI/NleA, EspJ, EspK, EspL*, EspM*, EspN, EspO*, EspR, EspS, EspT, EspV, EspW, EspX*, EspY*, NleB*, NleC, NleD, NleE, NleF, NleG/NleI*, NleH*, NleK, Cif, Ibe, $EspF_U$*

*Genome contains multiple homologs. For details see Wong *et al.,* 2011.

1.3.5 *The Shiga toxin*

As mentioned in section 1.2, carriage of the *stx*$_1$ or *stx*$_2$ genes encoding the potent Shiga-like cytotoxin is the main defining feature of the STEC pathotype (Croxen *et al.,* 2013). Stx1 and Stx2 encoding genes are located on lambdoid bacteriophage (CP-933V and BP-933W respectively) and Stx expression is highly dependent on inducing conditions, such as low iron for Stx1 and the lytic phase of the phage life cycle for the more potent Stx2 (Neely & Friedman 1998, Schüller 2011). Classically the lytic phase is induced by DNA damage and subsequent activation of the bacterial SOS response but can also be influenced by factors produced by the human host cell (Wagner *et al.,* 2001). The SOS response is a bacterial stress response network that involves a large regulon of genes involved in DNA repair. These genes are under the control of the repressor protein LexA. Upon DNA damage, an antirepressor of LexA, known as RecA, is transcribed and relieves LexA repression of the SOS regulon while also initiating DNA repair (Figure 1-6A) (Michel 2005). This is performed by RecA forming filament structures with single stranded DNA (ssDNA) that mediate auto cleavage of LexA and also DNA damage repair. The SOS response is also known to be induced by exposure to many antibiotics, thus eliminating the possibility of treating STEC infections in this way as Stx expression is likely to be induced through RecA mediated antirepression of the phage cI repressor protein (Croxen *et al.,* 2013, McGannon *et al.,* 2010, Miller *et al.,* 2004, Schüller 2011).

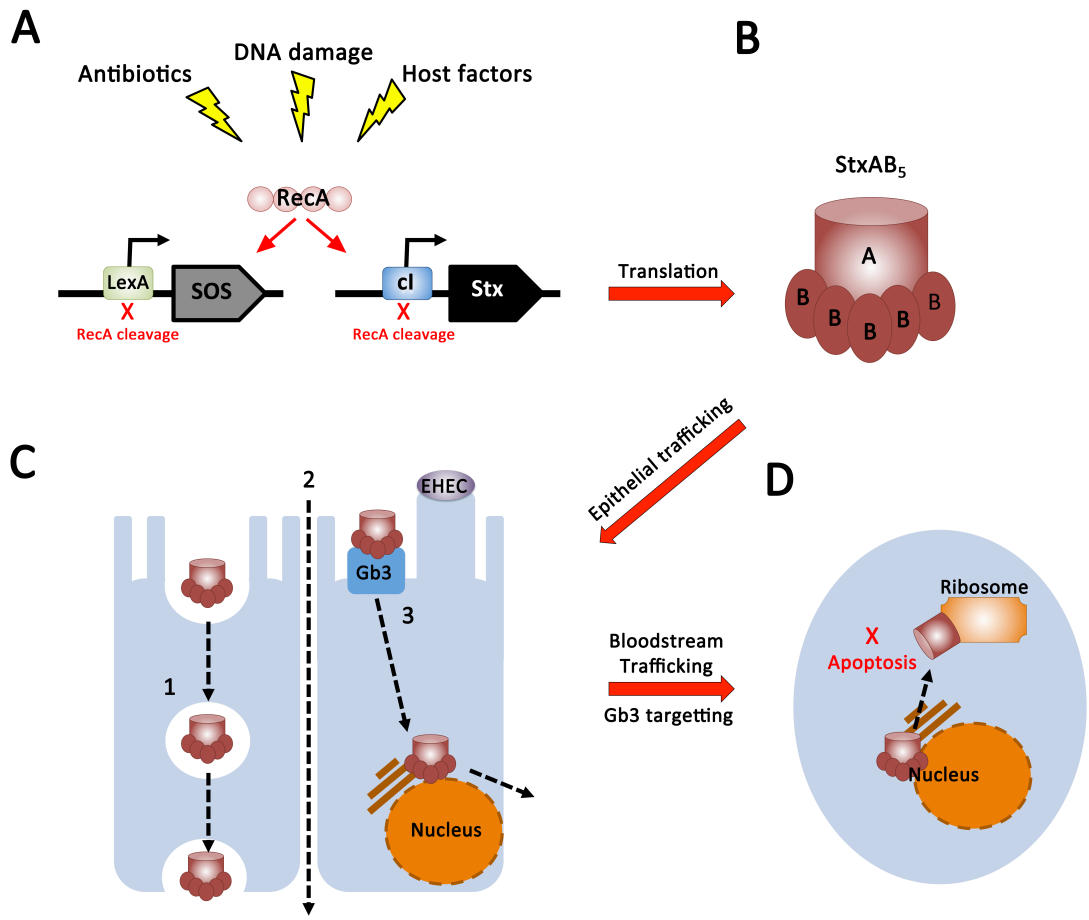

Figure 1-6 SOS activation and pathogenesis of Stx. The sequence of events the lead to Stx pathogenesis in the host. (A) The SOS response triggers RecA mediated auto-catalytic cleavage of LexA and prophage cI controlled genes. (B) Stx is an AB_5 type cytotoxin comprised of an active A subunit non-covalently linked to the B subunit pentamer. (C) Stx is trafficked across the intestinal epithelium by a number of proposed mechanisms. Receptor independent transctyosis, transport through tight junctions and retrograde transport via EHEC induced Gb3 expression on the intestinal epithelium (labelled 1 – 3 respectively). (D) Ribosomal inhibition and subsequent apoptosis in Gb3 positive kidney cells. This figure was adapted from Schüller 2011.

Stx is the virulence factor responsible for HUS in extreme STEC infections. The toxin is an AB_5 type cytotoxin that is comprised of the enzymatically active A subunit which is non-covalently linked to the B subunit pentamer (Figure 1-6B) (Fraser *et al.,* 2004). The B subunit recognises the globatrioaosylceramide 3 (Gb3) receptor on epithelial cell surfaces and internalizes the toxin via this recognition. The active A subunit is then transported by the retrograde pathway to the cytoplasm where it ultimately binds the ribosome and cleaves an essential adenine residue (A-4324) from the 28S rRNA, leading to inhibition of protein synthesis, the ribotoxic stress response and apoptosis (Endo *et al.,* 1988, Sandvig

et al., 1992, Smith *et al.,* 2003). The specificity of Stx to Gb3 receptors may explain why cattle are colonized asymptomatically by *E. coli* O157:H7, as they lack these receptors (Pruimboom-Brees *et al.,* 2000).

Stx is transported across the intestinal epithelium after bacterial release although the mechanism of this transport is unclear (Figure 1-6C). It is thought that disruption of the epithelium by bacterial adhesion and subsequent inflammation may facilitate this process. This theory is bolstered by the fact that patients with bloody diarrhoea are more likely to develop HUS than those without (Nataro & Kaper 1998). More specific mechanisms of Stx transport have been reviewed elsewhere (Schüller 2011). In brief, these mechanism include transport of Stx across the intestinal epithelium by transcytosis or through epithelial tight junctions that have been disrupted by the A/E infection process and transmigration of neutrophils (Hurley *et al.,* 2001, Philpott *et al.,* 1997), interaction with infection-induced Gb3 on the intestinal epithelial surface (Schüller *et al.,* 2007) or transcytosis via intestinal microfold-cells (Schüller 2011). Once Stx has breached the epithelial barrier and accesses the bloodstream it then reaches the surface of the kidney epithelium to which it binds via Gb3 (Figure 1-6D) (Schüller 2011).

1.3.6 *The bacterial flagella*

The flagella are a bundle of multiple tail-like appendages known individually as a flagellum that are assembled on the surface of the bacterial cell in many arrangements, peritrichously for *E. coli*. Many bacteria employ flagella as a means of motility for navigating and scavenging environments (Chevance & Hughes 2008). The bacterial flagellum shares structural similarities with the LEE-encoded T3SS and these systems are often referred to as flagellar or non-flagellar T3SSs. Various studies have argued over which system evolved first. It was once believed that both systems evolved from a common ancestor system however it is now more generally accepted that the flagellum structure was the ancestral basis for the non-flagellar T3SS, which evolved as a means of protein delivery to the host or extracellular environment (Abby & Rocha 2012, Gophna *et al.,* 2003).

The structure of the flagellum is complex and will only be briefly discussed but has been reviewed in detail elsewhere (Büttner 2012, Chevance & Hughes 2008, Cornelis 2006,

Erhardt *et al.,* 2010). It can be divided into three main components – the basal body or engine that shares many similarities to the T3SS basal apparatus, the propeller filament and the hook structure that joins them together (Figure 1-7). Like a non-flagellar T3SS the basal body spans the bacterial IM and OM. An MS-ring composed of 26 FliF units sits in the IM, fuses to FliE/FlhH and the periplasmic spanning rod-structure composed of FlgB/C/F/G/H/I and FliL. The integral-membrane motor proteins MotA/B surround the MS-ring in the IM whereas FliG/M/N make up the C-ring rotor on the cytoplasmic side. FliH/I/J form an ATPase complex that powers the system in a proton motive force dependent manner. FlhA/B and Flk are involved in substrate control and specificity of the system. ~120 FlgE subunits make up the hook structure and the flagellar propeller, a 20 nm long filament composed of ~20000 FliC subunits, extends from this (Büttner 2012, Chevance & Hughes 2008, Cornelis 2006, Erhardt *et al.,* 2010).

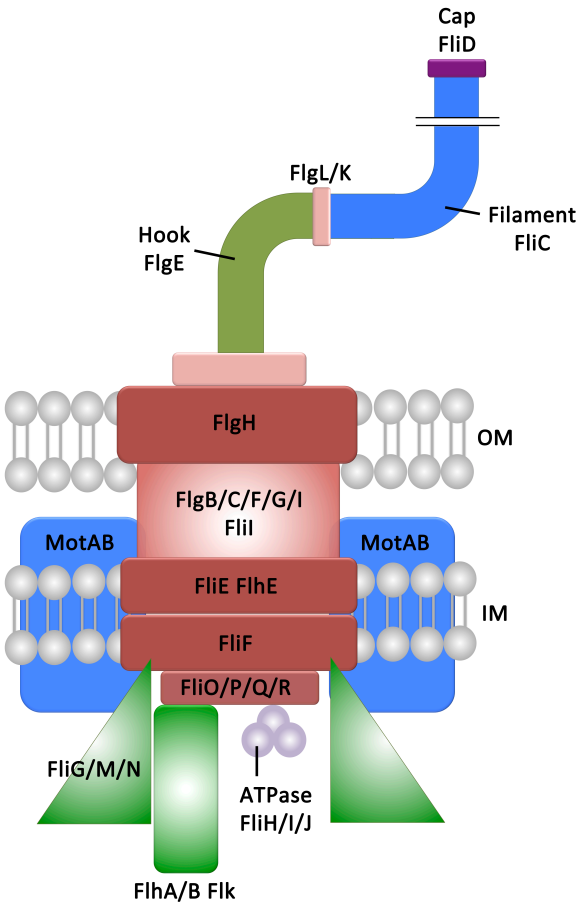

Figure 1-7 The *E. coli* flagellum. Schematic representation of the flagellar structure. Similarly to the T3SS, the basal body spans the bacterial envelope and is integrated into the IM and OM, illustrated in red. The flagellar motor (MotAB) powers the system. The cytoplasmic C-ring is illustrated in green (FliG/M/N) as well as the FlhA/B/Flk substrate recognition system. The flagellar propeller or hook (FlgE) extends from the basal body and propels the FliC filament for motility. This figure has been adapted from Büttner 2012.

Transcriptional context of the flagellum is very conserved among Gram-negative bacteria with peritrichous flagella (Chevance & Hughes 2008). There are over 60 genes responsible for the formation of a flagellum and these are not organised in a single locus (Frye *et al.,* 2006). Rather, they are expressed as a transcriptional hierarchy of three promoter classes that are sequentially activated. In short, class I genes encode the master regulators, FlhD/C, that form an $FlhD_4C_2$ heteromultimeric activator of class II expression. Class II genes are responsible for the motor complex are under the control of the housekeeping sigma factor, σ^{70}. Once the motor complex is formed, class III expression is driven through the flagellar sigma factor, σ^{28}, and the system is completed (Chevance & Hughes 2008).

The flagella provide motility to bacteria in two ways, swimming and swarming. Swimming is the movement of bacteria individually through a liquid media by flagellar rotation. Swarming, on the other hand, is a surface based movement of bacterial groups (Kearns 2010). To swim, to flagella must rotate anti-clockwise in coordination with one another forming a bundle that propels the cell forward. If the rotation of the bundle is reversed to a clockwise motion then the cell will 'tumble', disrupting its forward movement (Turner *et al.,* 2000). Directionality of the flagella is not random however and relies on a process called chemotaxis (Figure 1-8). Chemotaxis can be defined as a network of sensory systems that coordinate to direct a cell towards a region more highly concentrated in beneficial chemicals, or less concentrated in toxic chemicals. Chemotactic signals are sensed by dedicated chemoreceptors in the cell membrane known as methyl-accepting chemotaxis proteins. These receptors respond to the environment and transfer signals directly to cytoplasmic response regulators, by phosphorylation, which then transfer the signal further downstream to other chemotactic response proteins (Wadhams & Armitage 2004). One of these is called CheY, and once phosphorylated binds to FliM on the flagellar motor, inducing directional switch in rotation and ultimately tumbling (Welch *et al.,* 1993).

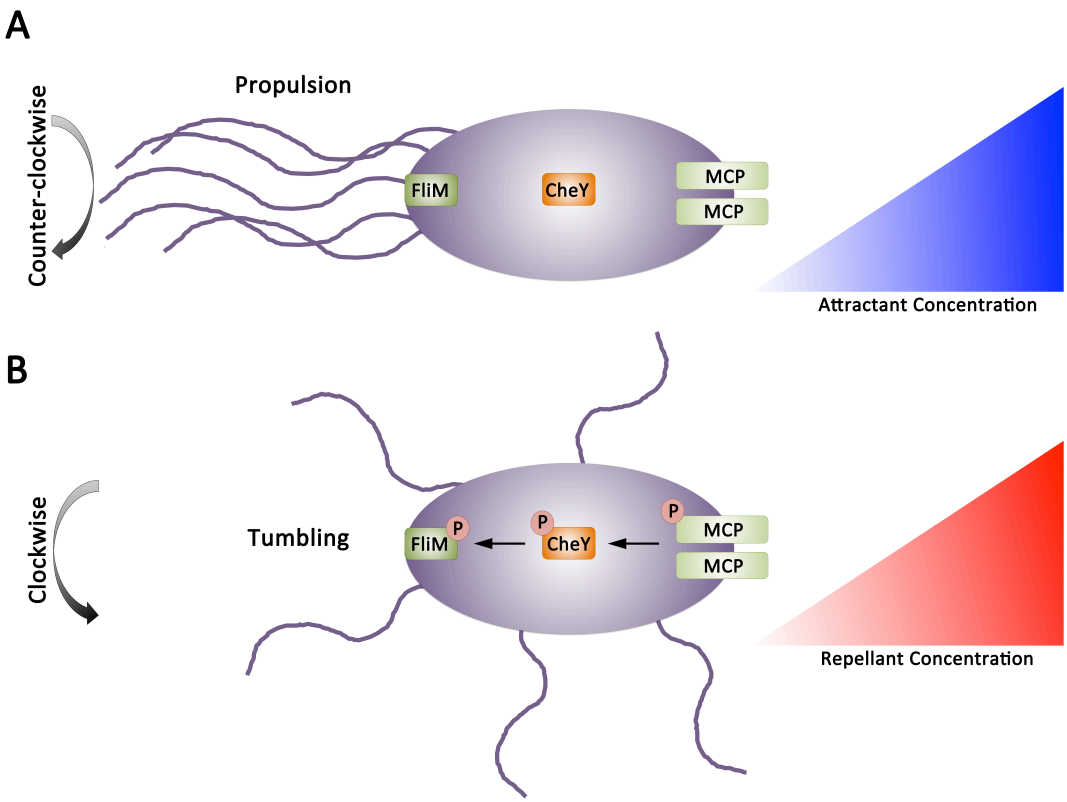

Figure 1-8 Bacterial propulsion, tumbling and chemotaxis. The directionality of flagellar rotation is determined by the process of chemotaxis. (A) When swimming in a favourable chemical concentration gradient, the flagella rotate counter-clockwise propelling the bacterial cell forward. (B) Conversely, if the chemical concentration of a repellent increases, dedicated surface receptors known as methyl-accepting chemotaxis proteins transmit signals via a phospho-relay system involving CheY to the flagellar motor system. This signal causes the flagella to rotate clockwise and the bacterial cell tumbles as a result, reorienting itself for propulsion towards a more favourable direction (Wadhams & Armitage 2004).

1.3.7 *Other O157:H7 virulence factors*

Horizontally acquired plasmids often contain virulence factors and various strains of O157:H7 have been shown to carry a number of plasmids. One of these, pO157, is a highly conserved plasmid known to be carried in all O157:H7 strains and a high percentage of STEC strains (Nataro & Kaper 1998). The ~90 kb plasmid sequence was shown to contain 100 ORFs with 19 of these predicted to be virulence associated genes (Burland 1998). A number of these factors have since been characterized and implicated in contribution to O157:H7 infections. EspP is a serine protease that belongs to the autotransporter family of proteins. Once secreted, EspP performs multiple functions including cleavage of human coagulation factor V, which implies exacerbation of

haemorrhagic disease (Brunder *et al.,* 1997). EHEC haemolysin is a repeats-in toxin protein that lyse erythrocytes although the exact role of haemolysin in infection is unknown (Schmidt *et al.,* 1996). Release of haem and haemoglobin from erythrocyte lysis may enhance proliferation of O157:H7 during infection by offering a prime source of iron for uptake (Law & Kelly 1995). StcE is a Type 2 secreted protease that cleaves the C1 esterase inhibitor of the complement pathway providing a potential protection mechanism for bacteria from the immune system. StcE has also been shown to be regulated by Ler (Lathem *et al.,* 2002). KatP is a pO157 encoded catalase-peroxidase that is employed to scavenge exogenous H_2O_2 for protection during infection of the host cell (Uhlich 2009). O-islands revealed during genome sequencing of O157:H7 also indicated many putative virulence factors including fimbrial and non-fimbrial like adhesins that may also play a role in host attachment (Perna *et al.,* 2001).

1.4 Extra Intestinal Pathogenic *E. coli* (ExPEC)

The ExPEC group of *E. coli* comprises an important but often overlooked branch of the pathogenic *E. coli* family, likely due to the lack of large scale epidemic style infections as is caused with STEC for instance (Johnson & Russo 2002). The two most commonly studied members of the ExPEC are UPEC (uropathogenic E. coli) and MNEC (meningitis associated *E. coli*). UPEC are the primary cause of UTIs being responsible for 70-95% of community-acquired UTIs and over 50% of nosocomial UTIs (Croxen & Finlay 2010, Pitout 2012). MNEC is the most common cause of Gram-negative associated meningitis in neonates with a case fatality rate as high as 40% (Croxen & Finlay 2010, Kim 2012). Both UPEC and MNEC also have the ability to further disseminate into the bloodstream and are among the most common causes of bacteraemia and sepsis (Johnson & Russo 2002). This section will briefly introduce both UPEC and MNEC giving an overview of their genetics, pathogenesis and associated virulence mechanisms.

1.4.1 *UPEC traits and pathogenesis*

UTIs are a great medical threat. UPEC is the most common cause of UTI and more than 40% of women will experience cystitis associated UTI in their lifetime but what is perhaps more alarming is the fact that UTI can also lead to pyelonephritis and in severe cases sepsis which results in over 35000 deaths in the US annually (Totsika *et al.,* 2012). Serologically, there is a high frequency of specific O-antigens associated with UPEC (O1, O2, O4, O6, O7, O8, O16, O18, O25 and O75) whereas the K- and H-antigens have a less defined pattern (Bidet *et al.,* 2007). From an evolutionary perspective, UPEC often group into the highly virulent B2 phylogroup, along with other ExPEC members (Johnson *et al.,* 2001a, Touchon *et al.,* 2009). Many isolates of UPEC have clonal characteristics but show little genomic specificity to UPEC as a stand-alone pathotype with most UPEC specific genes being of hypothetical function (Croxen & Finlay 2010). Indeed, the virulence factor repertoire of UPEC is incredibly heterogeneous with the pathotype having evolved on multiple occasions from independent clonal lineages (Brzuszkiewicz *et al.,* 2006, Welch *et al.,* 2002). The genomes of specific UPEC isolates (CFT073, 536 and UTI89) are up to 13% larger than the prototypical commensal *E.* coli K-12 with as much as 22% more ORFs (Welch *et al.,* 2002, Wiles *et al.,* 2008). The inflation is largely due to HGT of UTI-specific fitness traits such as various fimbrial adhesins, secreted toxins and iron-acquisition systems (Brzuszkiewicz *et al.,* 2006, Welch *et al.,* 2002). These characteristics are more specifically highlighted by the fact that the UPEC CFT073 genome contains 770 more genes than that of EPEC E2348/69, a dedicated InPEC, suggesting that the virulence repertoire required for a dissemination to and survival in the urinary tract is much broader that that of the intestine (Iguchi *et al.,* 2009). Indeed, InPEC virulence is often largely attributed to specific virulence factors such as the T3SS or Stx, whereas UPEC employs an array of virulence factors to thrive in its niche (Totsika *et al.,* 2012, Wiles *et al.,* 2008).

1.4.2 *An overview of UPEC virulence factors*

As mentioned in section 1.3.1, UPEC employ a diverse array of virulence factors seemingly specific for dissemination into the urinary tract that have been reviewed extensively elsewhere (Croxen & Finlay 2010, Wiles *et al.,* 2008). The infection process involves dissemination of UPEC from the gut to the urinary tract. The proximity of the anus to the

urinary tract facilitates this first step and adhesion to the uroepithelium soon follows. UPEC subsequently promotes epithelial invasion forming highly replicative intracellular bacterial communities. Bacterial proliferation and an influx of polymorphonuclear leukocytes result in exfoliation of the bladder and dissemination to further tissues. Untreated UPEC can then ascend to the kidney causing pyelonephritis and potentially being further trafficked to the blood supply leading to bacteraemia (Croxen & Finlay 2010, Wiles *et al.,* 2008). A schematic of UPEC virulence factors and infection is seen in Figure 1-9.

UPEC harbour multiple adhesion organelles including type 1, P, S and F fimbriae encoded on the *fim, pap, sfa* and *foc* operons respectively (Wiles *et al.,* 2008). Variation in fimbrial adhesins is UPEC strain specific with certain adhesins being specific for receptors on the uroepithelium or the kidney for more virulent isolates. For instance, Type 1 fimbriae are highly conserved between both UPEC and commensal *E. coli* and are the key factor in establishment of UTI whereas the P fimbriae tend to show preferential binding to cell receptors that are host tissue specific as well as being more commonly associated with pyelonephritis causing UPEC (Lane & Mobley 2007, Wiles *et al.,* 2008). FimH, a type 1 adhesin, was recently found to mediate internalisation by interacting with integrins on the host epithelium triggering a signalling cascade that results in actin reorganisation around the bacterium indicating more than simply an adhesive role for type 1 fimbriae (Eto *et al.,* 2007).

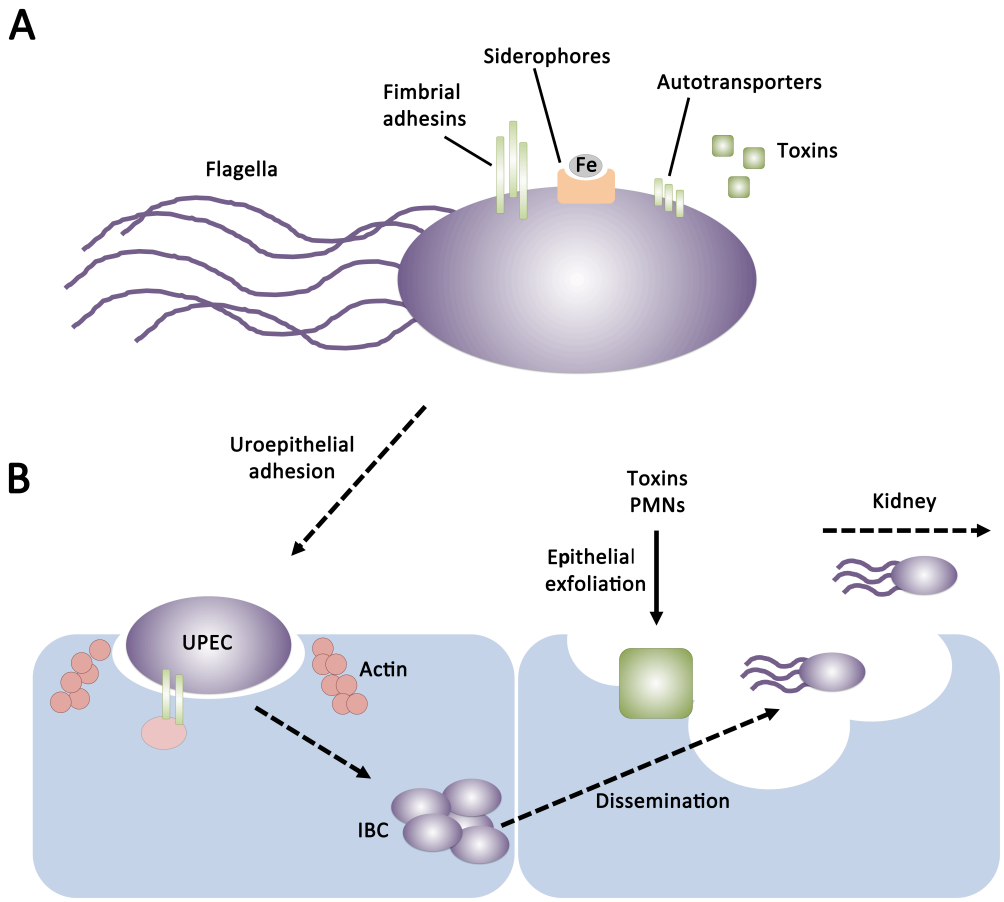

Figure 1-9 UPEC virulence factors and pathogenesis. (A) An overview of the multiple virulence factor employed by UPEC. These include diverse fimbrial adhesins, iron acquisition siderophores and autotransporter proteins. (B) Once UPEC has reached the uroepithelium, they adhere to the host cell triggering an invasion process involving modulation of host cell actin. Invasion allows the propagation of intracellular bacterial communities and toxin release as well as an influx of polymorphonuclear leukocytes results in exfoliation of the epithelium. UPEC is then free to disseminate past the immediate epithelium and ascend to the kidney. This figure was adapted from Croxen and Finlay 2010.

Although UPEC do not employ a T3SS for virulence, they utilise alternative secretion systems to secrete toxins. As mentioned above, epithelial exfoliation is important for UPEC dissemination. Another key virulence factor is the haemolysin HlyA, encoded by over 50% of clinically severe UPEC isolates, which can facilitate exfoliation (Bhakdi *et al.,* 1988, Wiles *et al.,* 2008). This process releases nutrients to a nutrient poor environment and allows an entry point for UPEC tissue dissemination. UPEC also regularly encode autotransporter proteins known to play roles in virulence. Autotransporters are secreted by the type 5 secretion system. Autotransporters encode the effector as well as

translocation signal peptide for transport across the IM by the Sec system. A third domain, the beta strand, exposes the folded effector at the cell surface (Henderson *et al.,* 2004). Autotransporters are known to play both cytotoxic and adhesive roles in UPEC infection (Allsopp *et al.,* 2012, Maroncle *et al.,* 2006).

Aside from adhesion and cytotoxicity, UPEC cannot thrive in the urinary tract without nutrients. In any infectious environment, the battle for iron between the host and the bacteria is key (Fischbach *et al.,* 2006). UPEC strains can express four characterised siderophore systems (enterobactin, salmochelin, yersiniabactin and aerobactin) in various combinations to scavenge iron from the environment (Totsika *et al.,* 2012). These systems are expressed under low iron conditions and contribute to UPEC virulence (Garcia *et al.,* 2011, Totsika *et al.,* 2012).

1.4.3 *MNEC traits and pathogenesis*

MNEC have long been recognised as the primary cause of Gram-negative associated meningitis in neonates with case fatality rates as high as 40% and severe neurological defects in surviving patients (de Louvois 1994, Kaper *et al.,* 2004). MNEC infections associated with adults should also not be overlooked. A small Japanese study highlighted that all patients monitored not receiving appropriate antibiotic treatment died, with the remaining patients experiencing a 27% mortality rate (Yang *et al.,* 2005). In contrast to UPEC, MNEC are not represented by a large number of O-antigen serotypes and ~80% of known isolates are of the K1 capsular antigen type (Croxen & Finlay 2010). The evolution of MNEC strains is an interesting point. Like other ExPEC, they comprise a group of *E. coli* that cannot establish a pathogenic infection in the GI tract but have acquired numerous virulence factors that permit dissemination to extraintestinal sites, given this opportunity. Phylogenetically, they largely occupy the highly virulent B2 group (Bingen *et al.,* 1998). Even though MNEC infections are serious, they are considered to be relatively rare and the brain seems to be a dead end for the pathogen offering little in the way of evolutionary pressure for mass dissemination from the gut (de Louvois 1994). It should also be pointed out that the quintessential MNEC serotype, O18:K1:H7, exhibits a stereotypical virulence factor profile when recovered from patients with acute cystitis, suggesting that UPEC and MNEC share similar virulence mechanisms, presumably acquired along their evolutionary timeline, that allow them to disseminate as opposed to

a single specific virulence factor that promotes dissemination unconditionally (Johnson *et al.,* 2001b). The MNEC genome has been subject to size inflation, a common theme among ExPEC. The MNEC strain RS218 was shown to contain ~500 kb of additional information over the *E. coli* K-12 strain and also carries a large ~114 kb plasmid which was recently shown to have high sequence similarity in many ExPEC strains and also to be essential for full MNEC virulence both *in vitro* and *in vivo* (Rode *et al.,* 1999, Wijetunge *et al.,* 2014).

1.4.4 *An overview of MNEC virulence factors*

The mechanisms of pathogenicity employed by MNEC are unique but not extensively understood. In neonates, the bacteria initially colonize the host perinatally from the mother followed by invasive transcytosis across the intestinal epithelium, allowing access to the bloodstream of the host, although the precise mechanism of this step are not largely understood (Figure 1-10A) (Burns *et al.,* 2001). The extent of disease progression by MNEC is largely dependent on the levels of bacteraemia experienced. That is, the higher the number of CFU per ml of blood, the higher the likelihood of meningitis progression (Dietzman *et al.,* 1974). Once in the bloodstream, the bacteria are protected from the immune system by multiple mechanisms - their antiphagocytic capsule, the presence of OmpA that has been shown to bind the complement-inhibiting factor C4bp and interactions with immune cells (Figure 1-10B). MNEC can invade and survive within macrophage and monocytes by inhibiting apoptosis and release of chemokines (Selvaraj & Prasadarao 2005, Sukumaran *et al.,* 2004, Wooster *et al.,* 2006).

The next step in MNEC pathogenesis involves crossing the blood brain barrier (BBB) to gain access to the central nervous system. Translocation of the BBB is not a damaging process however and instead occurs via transcytosis (Figure 1-10C). Invasion of the brain microvascular endothelium requires type 1 fimbrial adhesion FimH as well as OmpA which bind the host cell factors CD48 and ECGP96 respectively (Khan *et al.,* 2007, Prasadarao 2002). FimH and OmpA promote invasion by mediating host cell binding and subsequent actin modulation through cell surface signal transduction (Maruvada *et al.,* 2008). The K1 capsule, type S fimbriae and multiple Ibe (‘invasion of brain epithelium’) proteins are also known to have a role in invasion of the BBB (Croxen & Finlay 2010).

Once the central nervous system has been accessed, MNEC cause severe inflammation and neural damage.

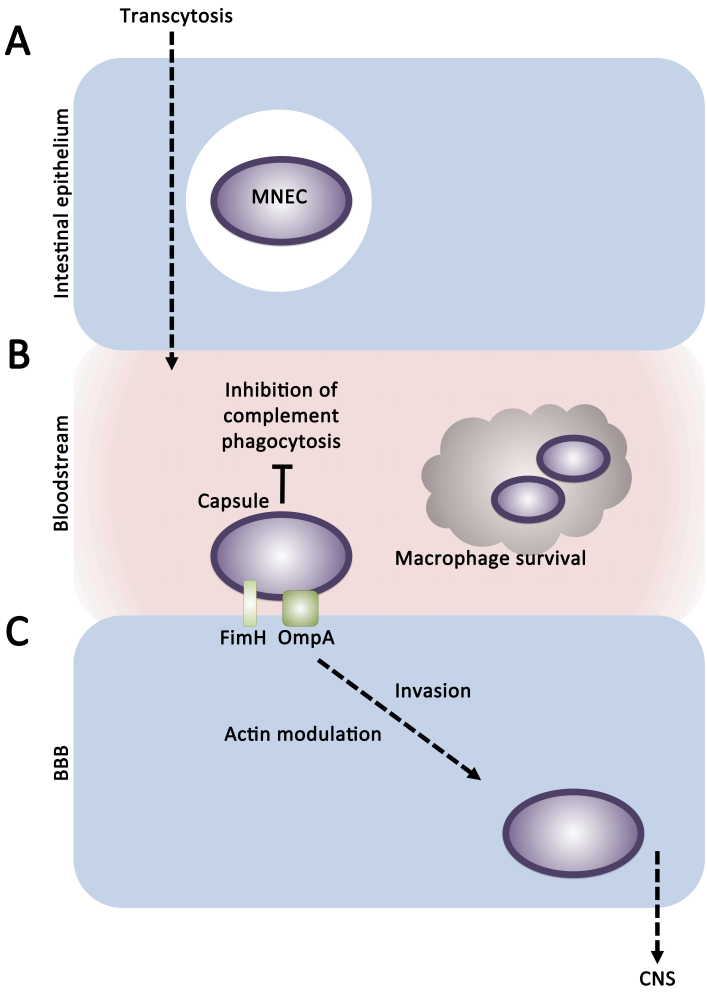

Figure 1-10 The pathogenesis of MNEC. (A) MNEC crosses the intestinal epithelium by an elusive transcytosis mechanism. (B) Once in the bloodstream, MNEC are protected from the immune system by virtue of the K1 capsule, which can inhibit complement activation and phagocytosis. (C) Transport of MNEC across the blood-brain barrier (BBB) involves the binding of surface factors FimH and OmpA to the microvascular epithelium. Binding of these factors promotes actin modulation and cellular invasion in a non-damaging manner. This figure was adapted from Croxen and Finlay 2010.

1.4.5 *The E. coli type 3 secretion system 2 (ETT2)*

The completion of the first O157:H7 genome sequences revealed detailed information about the carriage of virulence factors encoded on multiple islands within the genome (Hayashi *et al.,* 2001, Perna *et al.,* 2001). As described in section 1.2.3, the LEE encoded T3SS is conserved among a variety of InPEC as a key virulence factor but is absent from

ExPEC strains. Analysis of the O157:H7 EDL933 genome sequence however revealed the presence of a second cryptic T3SS that shares much sequence similarities to the Spi-1 (*Salmonella* pathogenicity island 1) T3SS of *Salmonella*. The system was designated the *E. coli* type 3 secretion system 2 (ETT2) and was originally thought to be fully present only in STEC strains with other enteric pathotypes containing truncated or mutated forms of the ETT2 locus (Hartleib *et al.,* 2003, Makino *et al.,* 2003, Perna *et al.,* 2001). Subsequently to this, Ren *et al.* published a comprehensive study detailing that the ETT2 locus is indeed present in majority of *E. coli* strains, be it pathogenic or commensal, but it is subject to widespread attrition rendering the system apparently non-functional. This came with the exception of the EAEC O42 strain that contained an 'intact' ETT2 system, according to genomic context at least (Ren *et al.,* 2004). Furthermore, a number of regulators of ETT2 expression were shown to control multiple ETT2 associated effectors in EAEC O42 and even affect expression of LEE encoded genes, suggesting the system has had certain functional capabilities whether they currently play a role *in vivo* or not (Sheikh *et al.,* 2006, Zhang *et al.,* 2004).

The role of a T3SS in ExPEC infection was somewhat obscure. ExPEC strains do not carry the LEE PAI but the ETT2 was demonstrated to be present in *E. coli* strains associated with septicaemia and it was this study that first demonstrated the contribution of the ETT2 system to pathogenesis by showing severe impairment of pathogenesis in an ETT2 mutant (Ideses *et al.,* 2005). Following on from this, the existence of the ETT2 system was confirmed in MNEC K1 strains. MNEC RS218 and E253 were shown to harbour mutated ETT2 loci but MNEC CE10 contained all the necessary genes for establishing a functional T3SS (Yao *et al.,* 2006). Furthermore, mutants in the CE10 ETT2 or one of its putative regulators, *eivA*, revealed a pathogen severely defective in invasion of human brain microvascular endothelial cells (Yao *et al.,* 2009). These findings suggest that ETT2 may play a crucial role in MNEC pathogenesis.

1.5 Regulation of the LEE in EHEC

As discussed in section 1.2.3 the LEE encodes a T3SS essential for the A/E phenotype associated with EHEC and other LEE positive pathogens such as EPEC and *C. rodentium.* Expression of the LEE is tightly regulated in response to multiple stimuli. Host temperature of 37°C provides optimal LEE expression as well as the exponential growth phase (Rosenshine *et al.,* 1996). Experimentally, the LEE can be induced by growth in tissue culture media (DMEM or MEM-HEPES) as the conditions mimic the physiological environment. In a comprehensive study, Kenny *et al.* attributed *in vitro* secretion to no single component but rather a combination of factors such as temperature, pH, osmolarity, calcium, iron and salt concentrations (Kenny *et al.,* 1997).

Transcriptional regulation of the LEE is extremely complex. It is controlled at the core level by two integral regulatory systems encoded on the LEE, Ler and GrlRA. Numerous other systems such as nucleoid regulators, stress response regulators and environmental sensing systems also feed into the LEE creating layers of specific control on the system, which will be discussed in the following sections (Mellies *et al.,* 2007).

1.5.1 *Master regulation of the LEE*

Transcription of the LEE is driven largely from five polycistronic operons (LEE1 through LEE5) in the form of a regulatory cascade with LEE1, LEE2 and LEE3 encoding the major structural components of the T3SS, LEE4 encoding the translocon pore and LEE5 encoding Tir and Intimin. Initial transcriptional mapping of the LEE revealed that transcription was driven from an ORF encoded at position one of the LEE1 operon. This ~15 kDa protein was named Ler (LEE1 encoded regulator) and was subsequently demonstrated to activate transcription at each of the subsequent LEE operons as well as non-operonic members of the LEE such as *map, espG* and *escD* (Elliott *et al.,* 2000, Haack *et al.,* 2003, Mellies *et al.,* 1999, Sánchez-SanMartín *et al.,* 2001). The importance of Ler to virulence was demonstrated by way of a deletion mutant in EHEC 86-24. This mutant was impaired in its ability to secrete effectors and form A/E lesions on host cells. Additionally, the *ler* mutant strain showed a decreased expression of non-LEE encoded virulence factors by Western blot (Elliott *et al.,* 2000). StcE, a non-T3SS secreted metalloprotease encoded on the

pO157 plasmid, was also shown in a separate study to be under the control of Ler, implying a global control of virulence gene regulation by Ler (Lathem *et al.,* 2002). Furthermore, a comprehensive study by Deng *et al.* that systematically analysed the virulence potential of each ORF by deletion within the LEE of *C. rodentium* also highlighted the absolute importance of Ler for virulence by the T3SS in a mouse model of infection, thus expanding the knowledge of LEE regulation to other pathogens (Deng *et al.,* 2004).

Transcriptional regulation of *ler* specifically is complex. Transcription has been demonstrated to be driven from not one but two promoters, designated P1 (distal) and P2 (proximal) (Islam *et al.,* 2011a). Recent studies have identified that P1 is likely to be the major promoter of the LEE1 operon but that the specific structure of the sequence is more detailed than previously thought, with the identification of a second cryptic P1 promoter that lies 10 bp upstream of the P1 transcriptional start site (Islam *et al.,* 2011b, Jeong *et al.,* 2012, Porter *et al.,* 2005). In two separate studies this phenomenon was identified in both EHEC and EPEC with minor differences in transcriptional start points highlighting the conservation of the LEE regulatory systems between pathotypes. Ler has also been demonstrated to elicit negative autoregulation on the LEE1 promoter as well as positively drive expression from operons LEE1 through LEE5, presumably as a mechanism of optimising LEE expression levels to steady-state during the infection process (Berdichevsky *et al.,* 2005). A schematic representation of LEE master regulation is seen in Figure 1-11.

1.5.2 *The GrlR/GrlA regulatory feedback loop*

The analysis of systematic LEE mutations by Deng *et al.* revealed a second LEE-encoded regulatory system controlling expression of this PAI. The researchers identified a positive regulator of *ler* expression, named GrlA (global regulator of *ler* activation), and a negative repressor of *ler*, named GrlR (global regulator of *ler* repression) (Deng *et al.,* 2004). GrlA forms a positive feedback loop with Ler to maintain steady activation of the LEE, whereas GrlR was postulated to inhibit GrlA mediated activation of *ler* under certain conditions (Barba *et al.,* 2005). Recently the structure of GrlR and a GrlR/GrlA complex has been solved and the mechanism of Ler regulation by this system elucidated. GrlR forms a dimeric structure in solution that binds GrlA, stabilises it and inhibits transcriptional

activation of the LEE1 operon (Padavannil *et al.,* 2013). This complements other findings that suggest under certain conditions favouring LEE expression GrlR is cleaved by the ClpXP serine protease complex, freeing GrlA and allowing *ler* transcription (Iyoda & Watanabe 2005). Despite this recent finding, the precise mechanism governing GrlR antirepression is unknown. A model of this system is seen in Figure 1-11.

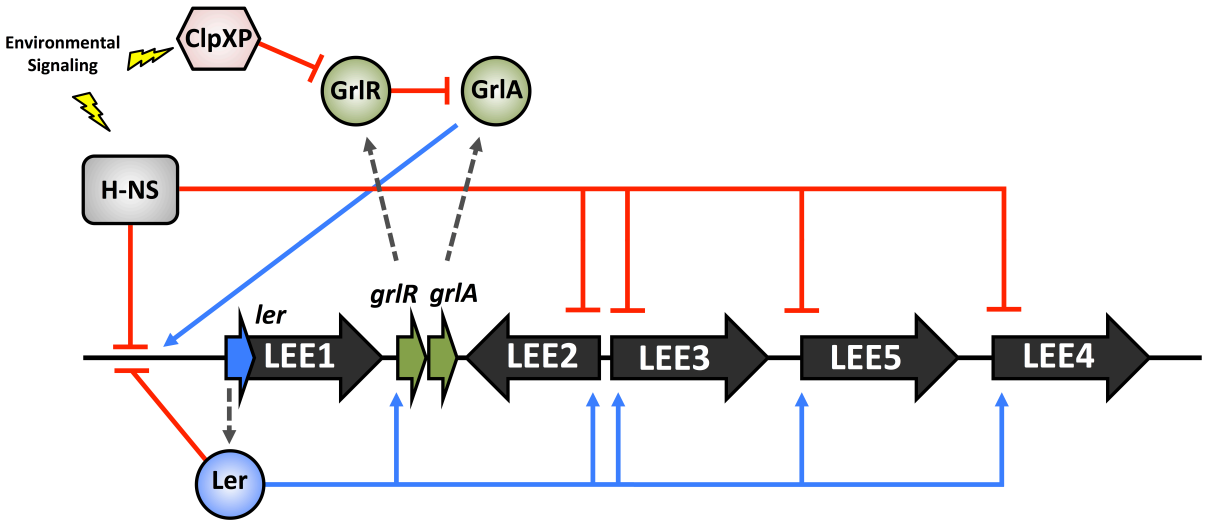

Figure 1-11 Master regulation of the LEE. The LEE is broadly divided into five polycistronic operons named LEE1 through LEE5. H-NS silences the LEE under non-inducing conditions. Ler is encoded by LEE1 first and subsequently activates LEE2-LEE5 by H-NS antirepression (discussed in section 1.4.3). Ler also autoregulates its own transcription to maintain steady state levels. Ler additionally drives transcription of *grlRA,* which codes for a second master regulatory input. GlrR binds GrlA inhibiting its action. Upon the correct stimuli GrlR repression is relieved by ClpXP, thus freeing GrlA, which further activates *ler* transcription. Ler and GrlA therefore form a positive feedback loop with one another. Transcriptional activation is indicated by a blue arrow; transcriptional repression is indicated by a red blunt arrow; translation is indicated by a grey broken arrow. This figure was adapted from (Padavannil *et al.,* 2013).

1.5.3 *Nucleoid regulation of the LEE*

Nuceloid-associated proteins (NAP) are classically defined as proteins involved in shaping and packaging DNA. They exist across all lineages of life (Eukarya, Bacteria and Archaea) and play global roles in nucleoid maintenance. The DNA-binding and shaping ability of NAPs allows them to affect more than just DNA architecture and they have emerged as key global regulators of transcription (Dillon & Dorman 2010). Perhaps of most relevance to the LEE is the histone-like nucleoid structuring protein (H-NS). This regulator can form

DNA-bridges at both intergenic and recently emerged intragenic sites on both sense and antisense strands of DNA, thus blocking or trapping DNA polymerase and 'silencing' the promoter (Dillon & Dorman 2010, Grainger *et al.,* 2006, Kahramanoglou *et al.,* 2011, Singh *et al.,* 2014). H-NS binding is biased towards AT rich segments of the genome, in other words foreign DNA acquired by HGT (Navarre *et al.,* 2006, Oshima *et al.,* 2006). LEE1 through LEE5 are all H-NS repressed under non-inducing conditions such as incorrect temperature for LEE expression (Umanski *et al.,* 2002). Ler acts as an H-NS antagonist by displacing H-NS from LEE1 to LEE5 operonic junctions thus relieving the LEE from H-NS mediated silencing (Figure 1-11) (Mellies *et al.,* 2007). Integration host factor (IHF) is a dimeric NAP capable of wrapping DNA and affecting transcription on a genome-wide scale. Indeed IHF was found in a global binding study to bind with sequence-specificity to ~30% of all operons in *E. coli* K-12 (Prieto *et al.,* 2012). With regards to LEE regulation, IHF positively promotes *ler* expression by binding upstream of the LEE1 promoter and is required for full LEE expression and A/E lesion formation (Friedberg *et al.,* 1999). In contrast, Hha has a negative effect on *ler* transcription and has a proposed mechanism of concentrating H-NS repression by DNA binding mimicry (Madrid *et al.,* 2007, Sharma & Zuerner 2004). The factor for inversion stimulation (Fis) is responsible for early exponential phase transcriptional regulation and has been demonstrated to affect *ler* and LEE4 expression in EPEC (Dillon & Dorman 2010, Goldberg *et al.,* 2001). Collectively, the NAP regulators elicit diverse transcriptional control over the LEE by feeding signals into its core regulatory circuit. Transcriptional effects of NAPs are summarised in Figure 1-12.

1.5.4 *The adaptable GAD acid stress response regulators*

Upon entry into stationary phase, *E. coli* employs specific mechanisms to counteract acidification of the environment. The glutamate-decarboxylase (GAD) acid stress response comprises two biochemically indistinguishable glutamate decarboxylases (GadA and GadB) located at different positions on the chromosome. These enzymes convert glutamate to ϒ-aminobutyrate and CO_2, which consumes intracellular H^+. ϒ-aminobutyrate is subsequently excluded from the cell via a dedicated antiporter, GadC (De Biase *et al.,* 1999). The GAD system is regulated by H-NS, the stationary phase sigma factor RpoS and multiple acid induced transcriptional regulators (Atlung & Ingmer 1997, De Biase *et al.,* 1999, Giangrossi *et al.,* 2005). The system is a complex signalling cascade involving an environmental response regulator EvgA, a non-acid dependent AraC/XylS

family regulator YdeO and two acid resistance regulators GadX and the LuxR-like GadE, that together positively affect transcription of acid resistance genes. A third regulator, YhiF, shares 23% homology with GadE and has been implicated in control of the acid resistance regulon but has not been extensively explored (Masuda & Church 2003).

Aside from acid resistance, transcriptional regulators of the GAD system have shown unique adaptability to LEE regulation. GadX has been demonstrated to bind directly to the *perA* (plasmid encoded regulator which positively affect *ler* expression) promoter in EPEC and inhibit its expression (Shin *et al.,* 2001). In O157:H7, GadE and YhiF mutants displayed increased LEE2 and LEE4 expression and improved adherence to host cells by a mechanism independent of *ler* (Tatsuno *et al.,* 2003). Additionally, Tree *et al.,* described the prophage-encoded secretion regulators (*psr*) found on numerous O-islands in O157:H7. In this study the researchers found that GadE and YhiF mediated Psr repression of the LEE, largely through repression of LEE2/3 and LEE5. Furthermore, they demonstrated binding of GadE to the LEE1 and LEE2/3 promoters indicating both direct and indirect control over LEE expression (Tree *et al.,* 2011). Acid in the stomach has long been considered as a colonisation barrier for *E. coli* and these studies suggest that GAD regulators may mediate adherence factor repression as the bacteria passage through the gut into more neutral intestinal environments as well as promoting timed expression of prophage encoded regulators and effectors (Tucker *et al.,* 2003, Tree *et al.,* 2011). More recently, the GAD regulators GadX and GadE were implicated in nitric oxide (NO) mediated repression of the LEE through a complex regulatory interplay. In a detailed study, the researchers showed that under LEE inducing conditions, the nitrite-sensitive repressor (NsrR) promotes LEE expression and represses GadE, whereas upon exposure to the abundance of NO in gastric juices, NsrR repression of GadE is lifted. GadE then represses LEE4/5 whilst also activating GadX, which indirectly represses LEE1 (Branchu *et al.,* 2014). These studies highlight the adaptive power of the conserved GAD system that has been usurped by environmental sensory systems and cryptic prophage elements to coordinate LEE expression appropriately. Transcriptional effects of the GAD regulators are summarised in Figure 1-12.

1.5.5 *Crosstalk between EHEC regulatory elements and the LEE*

As highlighted in Figure 1-11, the LEE elicits master regulation of its genes by internal activation of its intrinsic regulators Ler and GrlA. However, as mentioned briefly several external elements can feed into the LEE to regulate virulence, such as NAPs and the GAD regulators. LEE regulation by way of horizontally acquired elements is well documented. In EPEC, the *perABC* genes are encoded on the EAF virulence plasmid but this system is not found in EHEC (Mellies *et al.,* 1999). Interestingly, three functional homologues of the EPEC Ler activator PerC, *pchABC* encoded on lambdoid prophages, were identified in O157:H7 and found to be required for full virulence. Strikingly, the regulators could function in the opposite background by the same mechanism (Iyoda and Watanabe 2004; Porter *et al.,* 2005). Furthermore, Abe *et al.* comprehensively demonstrated the binding capacity of Pch to both the *ler* promoter and many NLE promoters and coding sequences (Abe *et al.,* 2008). Temporal regulation of the LEE through Pch and Psr has been implicated in facilitating NLE competition with LEE-encoded effectors for translocation via the T3SS (Tree *et al.,* 2011). Additionally to this idea of phage mediated control, Stx2 prophage have been shown to inhibit LEE1 expression by way of the prophage encoded regulator molecule cII, hypothesised to allow a subset of the population to induce Stx2 and create dependency on the Stx prophage (Xu *et al.,* 2012). Crosstalk between LEE and prophage regulatory systems can also be found in the SOS response. SOS induction is known to induce Stx2 but was also found to increase transcription of LEE2/3 operons as well as the CP-933P encoded *nleA* (Mellies *et al.,* 2007). The reasons for SOS regulation of the LEE are unclear but again, may imply a temporal regulation of phage-encoded elements throughout the course of infection.

Cross-regulation between the LEE and flagella is an interesting area of research. The H7 flagellum has been suggested to play a role in initial binding to host sites with down-regulation as intimate attachment proceeds (Mahajan *et al.,* 2009). A model was proposed whereby flagellated bacteria reach their niche, adhere and subsequently colonise via T3SS mediated attachment when the flagella no longer need to be expressed. This cross-regulation has been explored. As mentioned in section 1.4.2, GrlR repression of GrlA is cleaved by ClpXP protease, thus promoting LEE expression by GrlA (Padavannil *et al.,* 2013). Furthermore, ClpXP post-translationally modifies the flagellar master regulator, FlhDC, thus inactivating it. Additionally, free GrlA also transcriptionally represses *flhDC*

creating two independent mechanisms of motility inhibition when the LEE is induced (Iyoda *et al.,* 2006, Kitagawa *et al.,* 2011). IHF has also been shown mediate *flhDC* repression and is known to be required for full LEE expression, thus contributing to the theory of temporal attachment/motility (Friedberg *et al.,* 1999, Yona-Nadler *et al.,* 2003). LrhA is a known repressor of flagellar genes and has recently been demonstrated to induce *pchA/B* and therefore the LEE in O157:H7 by directly binding *pchA/B* promoter regions (Honda *et al.,* 2009, Lehnen *et al.,* 2002).

The stationary phase sigma factor RpoS plays a global role in gene expression particularly in the stationary phase of growth or under stressful conditions. RpoS levels are tightly controlled post-translationally in exponential phase by ClpXP and regulates ~10% of the *E. coli* genome (Battesti *et al.,* 2011). Conflicting accounts of the effect of RpoS on LEE expression have been reported in the past and the differences in phenotypes seem to be strain specific. Sperandio *et al.* first reported that LEE3/5 were induced by RpoS in the 86-24 strain whereas the opposite was seen in the Sakai strain, although this repression was via Pch (Iyoda & Watanabe 2005, Sperandio *et al.,* 1999). A more recent study on the RpoS regulon in EDL933 suggested differential expression of LEE encoded genes in response to RpoS but this was possibly conditions dependent (Dong & Schellhorn 2009a). In *C. rodentium* RpoS contributes to virulence in a mouse model of infection although RpoS is central to the general stress response and separating survival and virulence can be problematic (Dong *et al.,* 2009).

The RcsDCB phosphorelay is an environmental sensory system involved in regulation of capsule biosynthesis and cell envelope stress (Laubacher & Ades 2008). This system activates *ler* expression through the EHEC specific transcriptional regulator GrvA (Tobe *et al.,* 2005). The system also positively controls Pch activation of *ler*, and this has been proposed to act through LrhA (Honda *et al.,* 2009, Tobe *et al.,* 2005). Conversely, the RcsDCB system negatively regulates motility in *E. coli* (Francez-Charlot *et al.,* 2003).

Hfq, the small RNA (sRNA) chaperone, is a global post-transcriptional regulator of gene expression. Hfq has been characterised to mediate regulatory pathways involving H-NS and RpoS through the sRNA molecule DsrA (Laaberki *et al.,* 2006, Mikulecky *et al.,* 2004). Indeed, the LEE was found to be negatively regulated both directly and indirectly of Ler by Hfq in EDL933 but positively regulated in 86-24, similarly to the RpoS phenomenon

described above, suggesting pathogen specific regulation involving Hfq (Hansen & Kaper 2009, Kendall *et al.,* 2011, Shakhnovich *et al.,* 2009).

Finally, the ETT2 system constitutes a cryptic T3SS found in all *E. coli* with a large degree of degeneracy (see section 1.3.5). The precise role of ETT2 in pathogenesis is an obscure topic however regulators encoded by the system, EtrA and EivF, were found to be functional repressors of the LEE in two EHEC strains. This phenomenon which the authors named the "Cheshire Cat Effect" demonstrates the importance of functional crosstalk between differentially regulated virulence systems irrespective of the functionality of said systems (Zhang *et al.,* 2004).

Overall, regulatory crosstalk between the LEE and numerous global regulatory systems is a well established yet constantly evolving area of research highlighting the complexity and adaptability of *E. coli* transcriptional regulation. Figure 1-12 summarises the regulatory network as discussed in this section.

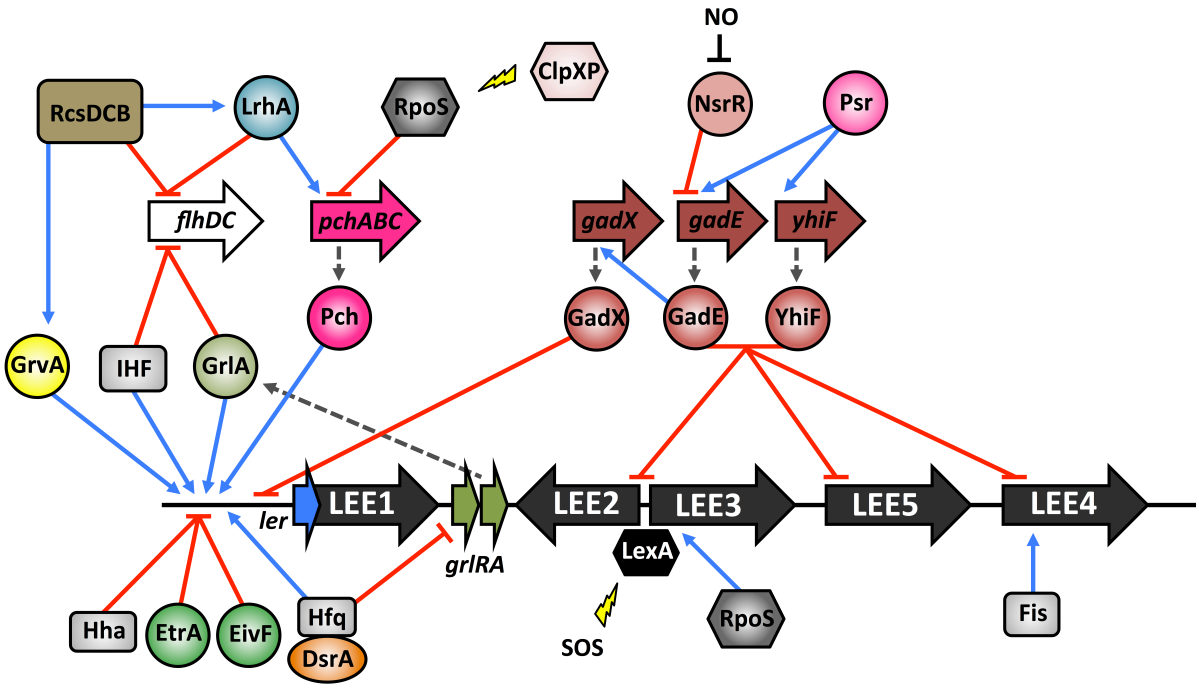

Figure 1-12 LEE regulation by known transcriptional regulators. Schematic regulatory map highlighting the crosstalk between EHEC regulatory elements and the LEE. Transcriptional activation is indicated by a blue arrow; transcriptional repression is indicated by a red blunt arrow; translation is indicated by a grey broken arrow; a physiological signal is indicated by a black blunt arrow; the yellow bolt indicates post-translational modification. See sections 1.4.3 to 1.4.5 for details.

1.5.6 *The impact of intestinal physiology on LEE expression*

Despite the staggering amount of knowledge that has been gained on regulatory systems feeding into the LEE the precise signals that stimulate these systems are often not fully understood. However, a rapidly emerging theme is the role that the intestinal physiology, that is metabolism and metabolic by-products of both the host and resident microbiome, plays on LEE regulation. Indeed, the LEE is largely associated with intestinal pathogens and characteristics of this environment can shape how these pathogens thrive in such a niche. The following section will discuss how bacterial and host metabolites can influence regulation of the LEE.

Bacteria have the fascinating ability to communicate between cells via a process known as quorum sensing. This involves the transmission and sensing of bacterially produced hormone-like signals by a system of sensor kinase surface receptors (Hughes & Sperandio 2008). Even more interesting is the fact that the system has evolved to sense not only bacterially produced signals but also host-derived hormones, namely epinephrine (Epi) and norepinephrine (NEpi). The QseBC and QseEF two-component regulatory systems and QseA transcriptional regulator have been previously defined. Briefly, QseA can both directly and indirectly activate *ler* expression in response to environmental signals (Russell *et al.,* 2007, Sharp & Sperandio 2007). Epi, NEpi and an undefined bacterial signalling molecule auto-inducer 3 feed in via the sensor kinase QseC, thus phosphorylating the response regulator QseB (Clarke *et al.,* 2006). QseB can positively regulate *flhDC* (Clarke & Sperandio 2005). The second system, QseEF, can sense Epi, phosphate and sulphate then signal to QseA and this is dependent on QseC (Reading *et al.,* 2007, 2009). QseF can also activate *ler* as well as transcription of $espF_U$. Furthermore, EHEC can sense hormone-like signals produced by other bacteria in the environment. The transcriptional regulator SdiA has the remarkable ability of responding to acyl-homoserine lactones produced by neighbouring bacteria but not EHEC (Hughes *et al.,* 2010, Nguyen *et al.,* 2013). SdiA acts to repress the LEE but also activate expression of the GAD system (described in section 1.4.4) thought to inhibit LEE expression as EHEC passage through the stomach (Hughes *et al.,* 2010, Kanamaru *et al.,* 2000). Thus, the Qse/SdiA systems can sense diverse hormones in the environment to modulate LEE expression under the appropriate conditions.

The intestine of the host offers an environment that is rich in diverse nutrients for resident microbiota to make use of. Researchers have made important progress in understanding how *E. coli* use sugars for niche adaptation. Chang *et al.* described by systematic mutagenesis of metabolic pathways that only mutation in sugar utilisation systems affected competitive colonisation in a mouse model. They also described the *in vivo* order of preference for seven key sugars involved in colonisation (Chang *et al.,* 2004). The researchers have gone on to propose "The Restaurant Hypothesis" in which they describe how differential nutrient availability within different intestinal mixed biofilms may explain how different *E. coli* strains occupy individual niches in the gut (Leatham-Jensen *et al.,* 2012, Meador *et al.,* 2014). In a fascinating study, Ng *et al.* demonstrated how *Salmonella typhimurium* and *Clostridium dificille* are capable of expansion by taking advantage of liberated carbohydrates in the host intestinal mucosa post antibiotic treatment (Ng *et al.,* 2013). O157:H7 utilises both glycolytic and gluconeogenic conditions in the gut to temporally regulate colonisation. Expression of *ler* is directly inhibited and promoted by glycolytic and gluconeogenic conditions respectively by the catabolite responsive Cra/KdpE system and this system is dependent on QseC (Njoroge *et al.,* 2012). The importance of gluconeogenic metabolism in O157:H7 intestinal persistence was investigated by transcriptome analysis in the bovine small intestine. The study found regulons involved in ethanolamine, urea and amino acid utilisation, key nitrogen sources, were upregulated in this environment (Bertin *et al.,* 2014). This agreed with an earlier observation that ethanolamine, a cell membrane component released into the gut lumen, provided a competitive advantage for O157:H7 over the resident microbiota that normally do not metabolise ethanolamine extensively (Bertin *et al.,* 2011). More recently, ethanolamine was confirmed as a signal for EutR, the transcriptional regulator, which drives expression of *ler* and the *eut* ethanolamine metabolic genes independently of one another (Kendall *et al.,* 2012, Luzader *et al.,* 2013). Finally, ethanolamine also activated expression of putative O157:H7 fimbrial-like loci, further promoting the theory that ethanolamine plays an important physiological role in O157:H7 colonisation (Gonyar & Kendall 2014).

Fucose is found in high concentrations in the intestine and is one of the favourable sugars used by *E. coli* as a carbon source (Chang *et al.,* 2004). To avoid unnecessary competition with resident *E. coli,* the fucose response regulator system FusKR of EHEC senses abundant fucose in the upper mucus layer of the lumen and in return directly represses

the LEE and expression of fucose utilisation genes. Once EHEC reaches the epithelium, QseCE signalling cascades promote LEE expression and repress *fusKR*, in turn allowing fucose utilisation to occur once EHEC has reached its niche (Pacheco *et al.,* 2012). This study is an elegant example of how metabolism and gene expression networks work together but not dependently on each other to carefully control niche-specific virulence in response to particular niche nutritional content.

As well as carbon and nitrogen sources, the intestine is abundant in short chain fatty acids (SCFAs) such as butyrate, acetate and propionate that can reach concentrations of up to 140 mM (Cummings & Macfarlane 1991). Nakanishi *et al.,* showed that butyrate could significantly increase LEE expression in EHEC at concentrations that did not inhibit growth. The researchers also described how this regulation was mediated via PchA and this was dependent on the leucine-responsive regulator (Lrp) (Nakanishi *et al.,* 2009). Lrp was found in K-12 to act at 138 different binding sites including that of *lrhA* (Cho *et al.,* 2008). Thus, butyrate activation of the LEE may be via an Lrp-LrhA-PchA signalling cascade. The researchers further characterised butyrate induced LEE expression by identifying the global LysR type transcriptional regulator LeuO as a mediator in the Lrp dependent system. They showed that butyrate induced LeuO via Lrp, which subsequently activated *ler* via PchA. Ler and PchA binding sites were identified upstream of *leuO* suggesting the systems forms a positive feedback loop upon induction (Takao *et al.,* 2014). Furthermore, the researchers also demonstrated more complex roles that SCFAs play in environmental signalling by demonstrating that SCFAs could activate *flhDC* under T3SS inducing conditions that would normally repress motility (Tobe *et al.,* 2011). Recently, Beckham *et al.* added to this area of research by demonstrating that an O157:H7 strain deleted for *adhE,* the gene encoding alcohol dehydrogenase, displayed a pleiotropic phenotype – overproduction of non-functional flagella and post-transcriptional repression of the T3SS. The researchers attributed this phenotype to an increase in intracellular acetate as a result of interference with AdhE mediated metabolic pathways (Beckham *et al.,* 2014).

The impact of intestinal physiology on virulence gene expression is at an early stage of understanding but exciting progress has been made. The findings provide important links between complex regulatory systems and how they respond to relevant environmental

characteristics that surround the bacteria that express them. A schematic regulatory map highlighting the signalling discussed is seen in Figure 1-13.

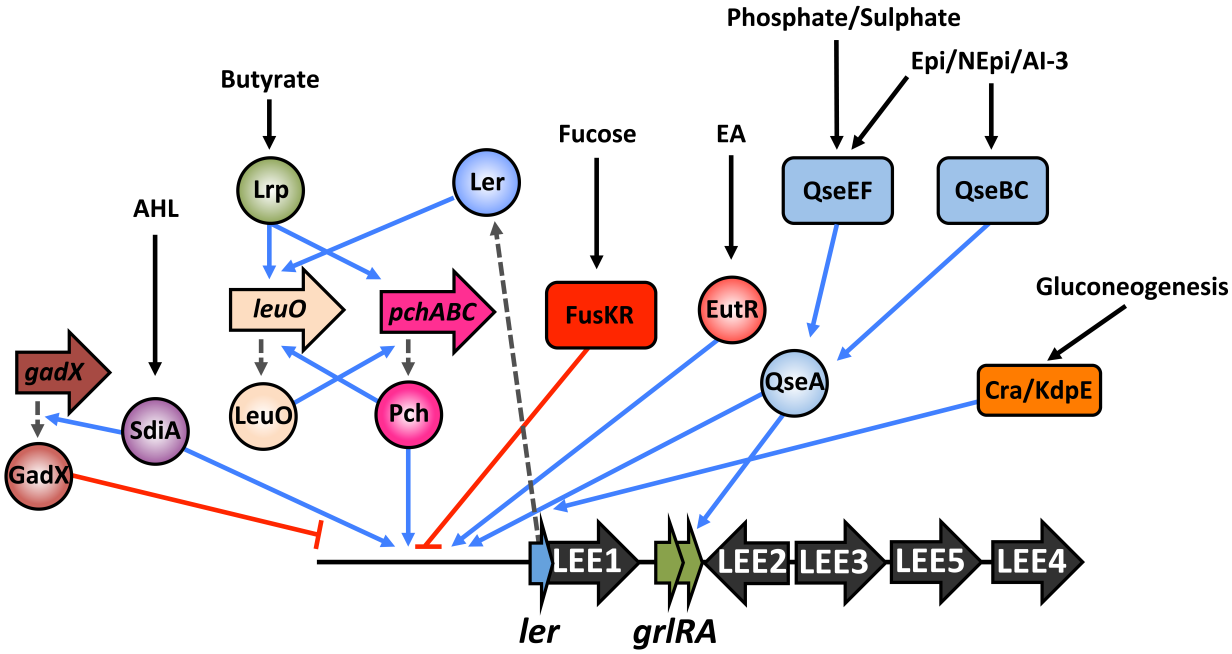

Figure 1-13 The effect of intestinal physiology on regulation of the LEE. Schematic regulatory map highlighting the several known physiological signals found in the intestinal niche that feed into the LEE regulatory network. Transcriptional activation is indicated by a blue arrow; transcriptional repression is indicated by a red blunt arrow; translation is indicated by a grey broken arrow; physiological signals are indicated by a black solid arrow. See section 1.4.6 for details.

1.6 Identification of YhaO as a novel virulence determinant

Previous work performed by Dr. Andrew Roe aimed to identify novel genes that may relate to LEE expression in O157:H7. Microarray analysis revealed that a previously uncharacterised gene, *yhaO*, was upregulated in conditions that promote high expression of the LEE (Roe *et al.,* 2007). Sequence analysis of *yhaO* revealed that the 1.3 kb gene coded for a putative inner membrane serine/threonine transporter protein that showed high sequence identity (>95%) in a number of bacterial species including *Shigella, Citrobacter* and *Salmonella.* Additionally, *yhaO* is positioned adjacent to *yhaM* that codes

for a putative serine deaminase that again showed high sequence identity (>80%) in other bacterial species. Early biochemical characterisation of YhaO and YhaM by Dr. Mads Gabrielsen confirmed that the YhaO indeed functioned as a serine transporter, indicating that functional aspects of YhaO agreed with sequence predictions (Figure 1-14; unpublished data). YhaM was also demonstrated to function as a serine deaminase with a preference for D-serine, suggesting that D-serine may be the substrate for the system. Furthermore, deletion of *yhaO* resulted in a marked decrease in the expression of LEE encoded effectors and the T3SS as analysed by SDS-PAGE and western blot, a phenotype that could be restored upon complementation of *yhaO* in trans (Figure 1-14D). This result raised a number of intriguing questions: Why would deletion of a D-serine transporter influence expression of the T3SS? Why is this transporter upregulated alongside the T3SS in the first instance? Why would transport of D-serine be important for regulation of the T3SS? This thesis aims to address these points in detail and the following sections provide an introduction to the relevance of D-serine and indeed D-amino acids (D-AAs) to *E. coli* and bacteria in general.

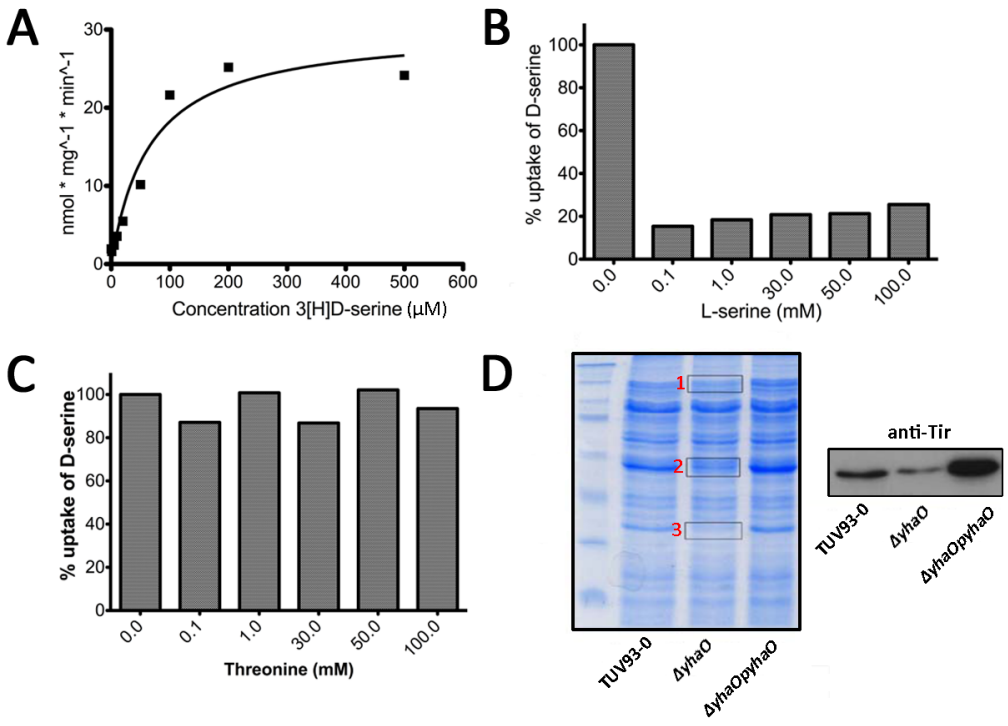

Figure 1-14 YhaO is an IM serine transporter involved in virulence. (A) Kinetic evaluation of 3[H]-D-serine transport. (B) Uptake of 3[H]-D-serine is outcompeted by an excess of L-serine. (C) Uptake of 3[H]-D-serine is not outcompeted by an excess of threonine. (D) SDS-PAGE analysis of secreted protein profiles of wild type O157:H7 (TUV93-0), *ΔyhaO* and *ΔyhaOpyhaO*. Bands labelled 1, 2 and 3 corresponded to Tir, EspD and EspA as identified by MS-MS. Western blot analysis of secreted fractions using anti-Tir antibody confirmed the SDS-PAGE results. Experiments were performed by Dr. Mads Gabrielsen and Dr. Andrew Roe.

1.6.1 *D-serine and E. coli physiology*

D-serine has emerged in recent years as playing an important role in the mammalian brain. It was found to be produced endogenously by way of a serine racemase enzyme that directly converts L-serine into the corresponding D-isomer (Wolosker *et al.,* 1999). This results in an unusually high concentration of D-serine in the brain (~0.3 µmol/g) where it acts as the main coagonist for *N*-methyl D-aspartate receptors, mediating processes such as neurotransmission and neurotoxicity (Wolosker *et al.,* 2008). Another mammalian organ has been demonstrated to contain high levels of D-serine, the urinary tract. D-serine was reported to be excreted in human and rat urine at concentrations of ~28 µM to ~1 mM (Anfora *et al.,* 2007).

This observation prompted much research in the lab of Professor Rodney Welch. They observed that the D-serine tolerance locus, *dsdCXA* that was previously characterised in K-12, was present in the UPEC CFT073 genome but disrupted in the O157:H7 genome of EDL933 which is an intestinal pathogen (Roesch *et al.,* 2003). The *dsdCXA* locus is comprised of three genes: *dsdX* which encodes an outer membrane D-serine transporter; *dsdA* which encodes a D-serine deaminase enzyme that converts D-serine to pyruvate and ammonia; and *dsdC* which is a LysR-type transcriptional regulator (LTTR) that controls expression of the system in response to D-serine (Anfora & Welch 2006, Federiuk *et al.,* 1983, Nørregaard-madsen & Fall 1995). The need for *E. coli* to detoxify itself from D-serine was highlighted by Cosloy and McFall who described how D-serine at ~0.475 µM was bacteriostatic to *E. coli* strains lacking deaminase activity in minimal medium by posing as an analogue of β-alanine and interfering with pantothenate (vitamin B5) and L-serine biosynthetic pathways (Cosloy & McFall 1973). Researchers in the Welch lab found that not only does the ability to tolerate D-serine as a carbon source via *dsdCXA* offer a fitness advantage to UPEC strains during a UTI (when they would likely be exposed to high concentrations of D-serine) but also that a *dsdA* mutant strain had an unusual pattern of gene expression displaying a hyper-motile/hyper-colonisation phenotype attributed to upregulation of virulence genes and also *dsdX*, hence an intracellular accumulation of D-serine (Anfora & Welch 2006, Anfora *et al.,* 2007, Haugen *et al.,* 2007, Roesch *et al.,* 2003). The authors proposed that D-serine played a role as an environmental signal modulating virulence gene expression but the exact mechanism or physiological relevance of this phenomenon was never understood.

The Welch lab offered further evidence that D-serine may play niche-specific roles in fitness or virulence again when they described a phenotypic study comparing patterns of D-serine utilisation by a selection of InPEC and ExPEC strains. The results showed that 97.5% of K1 isolates could use D-serine as a sole carbon source and this was in contrast to a mere 5% of InPEC strains that displayed this ability. It was also identified that a large number of K1 isolates used in the study harboured two copies of the *dsdCXA* locus (Moritz & Welch 2006). The brain is a D-serine rich environment and combined with the knowledge of these strains causing meningitis-associated disease, the theory of D-serine as a niche-specific virulence determinant was further bolstered. This study also reinforced an observation made previously by another group that *E. coli* strains lacking the ability to metabolise D-serine often have a truncation in *dsdX* resulting in loss of *dsdC* and a large portion of *dsdX,* rendering DsdA useless. The truncation in *dsdCXA* is due to the locus being present at one end of a hypervariable region of the genome, with the other end being the *argW* integration site for bacteriophage, making this region a hotspot for genetic recombination (Jahreis *et al.,* 2002, Moritz & Welch 2006). In InPEC strains, a reciprocal pattern of D-serine and sucrose utilisation was seen and this was attributed to the acquisition of the sucrose utilisation locus *cscRAKB* (encoding the CscR repressor, the CscA invertase, the CscK fructokinase and the CscB sucrose symporter) in place of *dsdCX,* presumably due to the lack of selective pressure for the need to utilise amino acids as a carbon source in a carbohydrate rich environment such as the intestine (Jahreis *et al.,* 2002, Moritz & Welch 2006). The genomic context of the *dsdCXA* locus is highlighted in Figure 1-15.

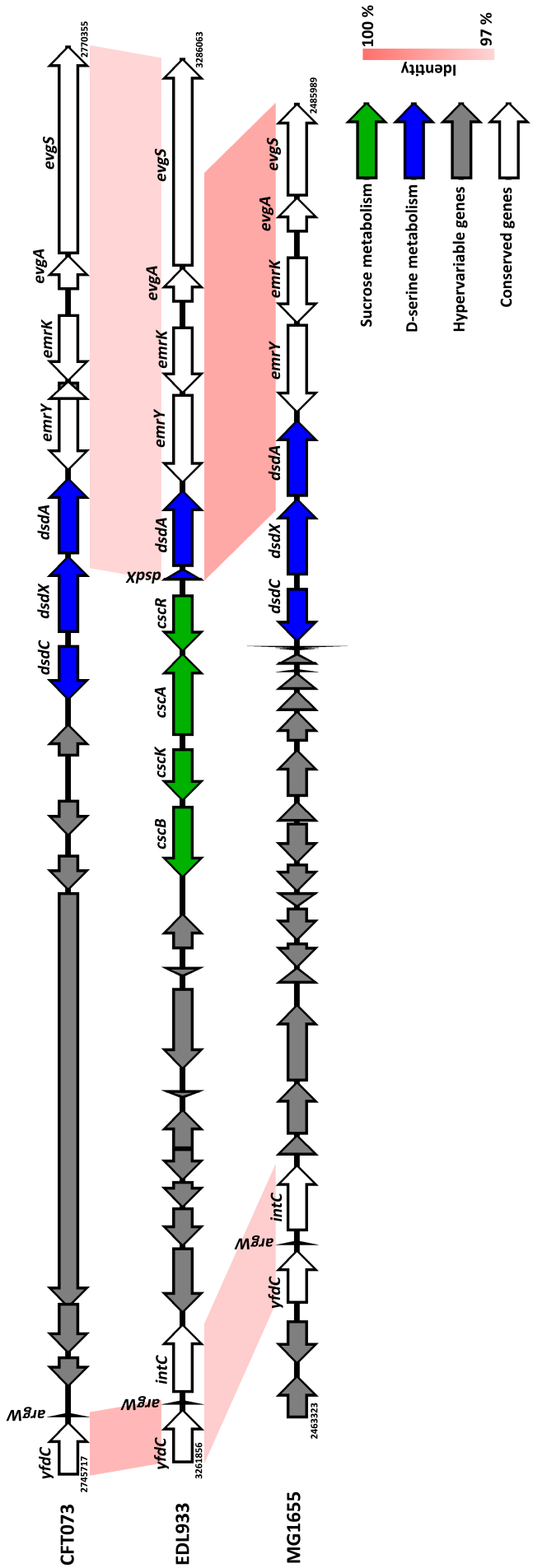

Figure 1-15 Genomic context of the *dsdCXA* locus. Diagram illustrating the genomic positioning of the *argW-dsdCXA* hypervariable region in the UPEC strain CFT073, the EHEC strain EDL933 and the K-12 strain MG1655 respectively. The regions have been aligned and sequence identity between conserved genes (white) is indicated by the legend. The *argW* integration site is labelled vertically on each region and hypervariable genes are coloured in grey. The *dsdCXA* locus has been truncated in EDL933 and replaced with the sucrose utilisation locus (*cscRAKB*) coloured in green. The genomic position of each region is labelled at either end of the diagram. This figure was generated using EasyFig (Sullivan et al., 2011).

The toxicity of D-serine is not limited to *E. coli.* Species of *Flavobacterium, Erwinia, Micrococcus* and *Mycobacterium* have all been reported to be inhibited by D-serine via interruption of metabolic pathways (Durham & Milligan 1962, Grula & Grula 1962, Whitney & Grula 1964, Yabu & Huempfner 1974). More recently, *Staphylococcus saprophyticus* was demonstrated to respond to D-serine by upregulating a known virulence factor, Ssp. This is interesting as it is the only species of uropathogenic *Staphylococcus* and carries the *dsdA* gene, which was essential *in vivo* (Korte-Berwanger *et al.,* 2013). Exposure to D-serine clearly can have detrimental effects on a variety of bacteria but also appears to be a positive fitness trait and virulence determinant to uropathogens however more work is needed to elucidate the specific mechanism behind this phenomenon.

1.6.2 *The role of D-amino acids in bacteria*

The French scientist Louis Pasteur (1822-1895) is most well known for his studies on vaccine development, fermentation and food preservation by way of pasteurisation however he was also renowned for making significant advances in the field of chemistry. Pasteur discovered that crystals of tartaric acid formed a mirror image of one another, despite being chemically identical (Flack 2009). The mirror images, known as isomers, are the basis of chirality (Bentley 2010). Chiral molecules of amino acids and sugars take on the nomenclature of D-isomers or L-isomers (as mentioned in section 1.5.1). In biological systems, L-amino acids (L-AAs) predominate in protein biosynthesis, as ribosomes cannot incorporate D-AAs into proteins. Conversely, D-sugars are more naturally occurring than their L-isomer counterpart (Cava *et al.,* 2011b). These principles again fall back to Pasteur's work on fermentation when he described how utilisation of tartaric acid displayed isomer-specificity (Gal 2008).

Despite this predominance for L-AAs in biology, D-AAs play many roles in biological systems such as mediating neuronal signalling and acting as alternative carbon sources (discussed in section 1.5.1). Peptides can also contain D-AAs and this occurs via two mechanisms: posttranslational modification of ribosomally synthesised proteins or non-ribosomal peptide synthesis (Cava *et al.,* 2011b). Probably the most well known example of D-AAs in bacterial systems however is in the synthesis of peptidoglycan (PG). PG is a mesh-like component of the bacterial cell wall composed of glycan sugar chains,

repeating disaccharides of *N*-acetyl Glucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), cross linked by short peptide bridges, collectively known as the murein sacculus (Typas *et al.,* 2012). Gram-negative bacteria have a thin PG layer that spans the periplasmic space between the IM and OM whereas Gram-positives have a much thicker PG layer that is directly exposed after the IM. PG is largely responsible for maintenance of cell shape, resistance to osmotic and physical pressure, as well as anchoring cell envelope components. PG synthesis is a complex process involving many enzymes and has been reviewed extensively elsewhere (Typas *et al.,* 2012). In summary, the process involves three stages. Firstly, nucleotide linked sugar precursors of GlcNAc and MurNAc are synthesised in the cytoplasm. Secondly, MurNAc precursors are linked to the IM acceptor undecaprenyl-P forming the lipid I intermediate. GlcNAc precursors are next incorporated yielding the lipid II intermediate. This hydrophilic precursor is then 'flipped' across the IM and attached to the extending sacculus by a DD-transpeptidase in the third stage of PG synthesis. PG turnover involves incorporation of precursors into pre-existing PG layers. Peptide side chains of glycan subunits are trimmed and modified for cross-linkage by a variety of enzymes known as carboxypeptidases, endopeptidases, amidases, lytic transglycosylases and transpeptidases. A detailed description of this process and the enzymes involved can be found in Typas *et al.,* 2012. The basic structure of PG is illustrated in Figure 1-16A.

Interestingly, the glycan peptide side chain of PG subunits typically contains the D-AAs D-alanine and D-glutamate. These isomers are formed by racemisation of their L-isomer counterpart (Cava *et al.,* 2011a). The classical structure of the peptide side chain is a pentapeptide of L-alanine, D-glutamate, meso-DAP (diaminopimelate) and a D-alanine dipeptide to finish. The position 4 D-alanine of one chain is covalently cross-linked to the position 3 meso-DAP of a neighbouring chain forming 4-3 cross bridging (Figure 1-16B). However, PG is a 3-dimensional structure and the formation of non-classical cross-linking structures such as peptide trimers or tetramers is also possible as detected by chromatographic analysis (Desmarais *et al.,* 2013). Once cross-linked, the peptide chains can be modified further to allow PG expansion (Typas *et al.,* 2012). The presence of D-AAs in bacterial PG provide resistance to proteases and thus are likely to have evolved as a protective adaptation to environmental stresses (Cava *et al.,* 2011b). Specific scenarios of bacteria incorporating alternative D-AAs into their PG have also been reported for example *Enterococcus gallinarium* strain BM4174 exhibited vancomycin resistance by

virtue of VanT, a membrane bound serine racemase that converted L-serine to D-serine for PG remodelling and subsequent resistance to vancomycin (Arias *et al.,* 1999).

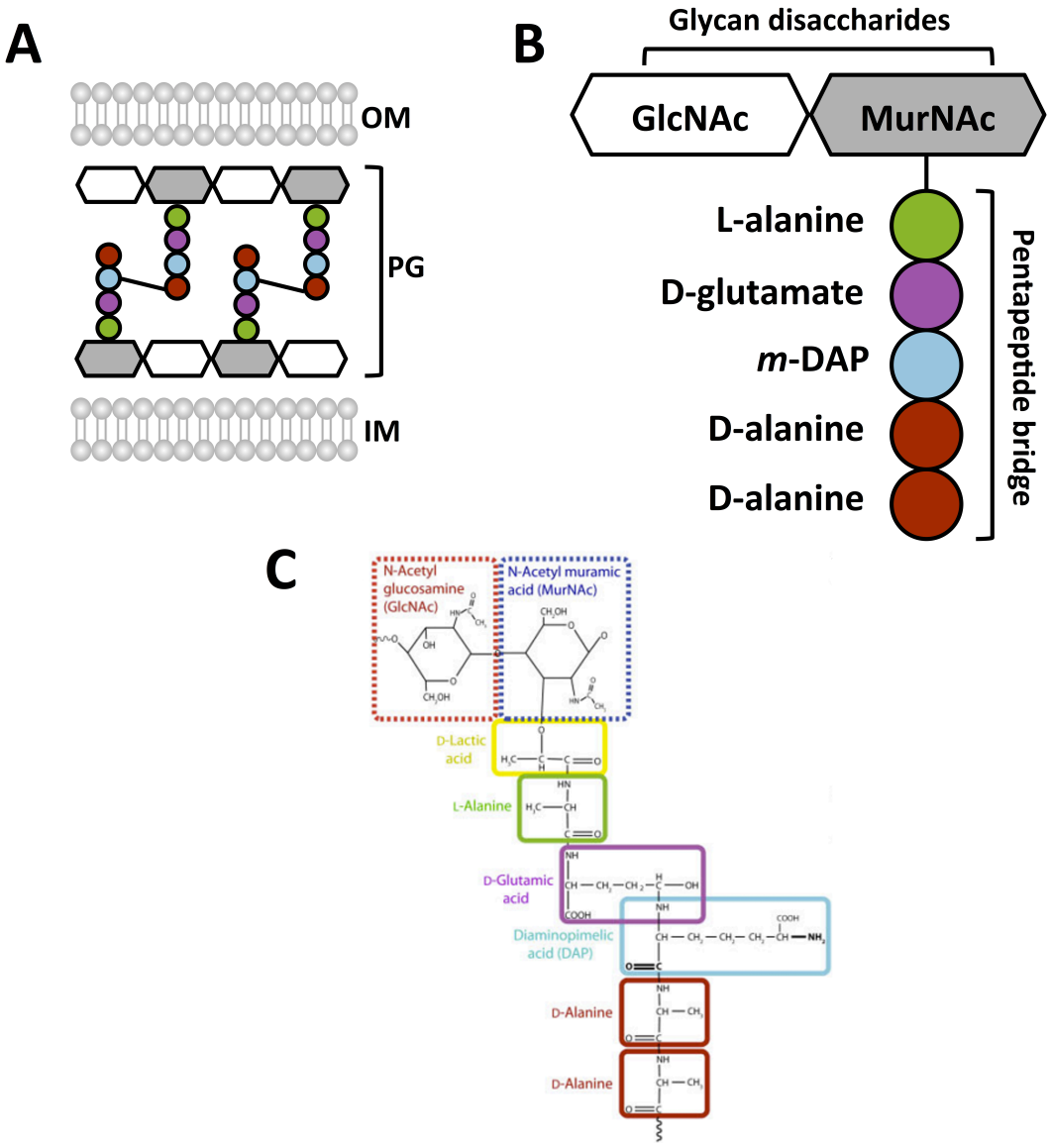

Figure 1-16 The basic structure of peptidoglycan. (A) Schematic representation of PG in Gram-negative bacteria. Layers of repeat structures of glycan disaccharides cross-linked by pentapeptide bridges. The structure spans the periplasm in between the IM and the OM. (B) Basic structure of a PG muropeptide. (C) Chemical structure of a PG muropeptide. This figure was taken directly from Desmarais *et al.,* 2013.

1.6.3 *Emerging roles of D-amino acids in cellular regulation*

As discussed in sections 1.5.1 and 1.5.2, D-AAs play a well conserved role in biosynthesis of PG and D-serine specifically has been implicated in vancomycin resistance for *Enterococci* and the regulation of virulence for uropathogens. Recently, the roles of non-

canonical D-AAs (NCDAAs) have been explored extensively in the lab of Professor Matt Waldor. While investigating growth phase dependent morphological shifts in *Vibrio cholerae,* the researchers found that the cell population accumulated millimolar concentrations of NCDAAs in the supernatant during stationary phase. The D-AAs identified were D-methionine, D-leucine, D-isoleucine and D-valine. A novel periplasmic racemase BsrV (broad-spectrum racemase in *Vibrio*), was found to be responsible for the conversion of L- to D-isomers and this enzyme was active in stationary phase (Lam *et al.,* 2009). Accumulation of D-AAs ultimately led to remodelling of the cell wall and the authors hypothesised that this was in order to regulate metabolic slowing and increase PG resistance to osmotic stress during stationary phase. Furthermore, they also investigated the mechanistic basis behind the phenomenon and proposed both periplasmic and cytoplasmic routes for D-AA remodelling of the PG that were seemingly dependent on the stationary phase sigma factor RpoS (Cava *et al.,* 2011a). Incorporation of high concentrations of exogenous D-AAs into bacterial PG has been established for some time however biological meaning of this was lacking (Cava *et al.,* 2011b). These studies have offered mechanistic relevance to the production and utilisation of NCDAAs. Production of D-AAs was documented in a diverse variety of bacteria including *Paeudomonas aeruginosa, Bacillus subtilis, Enterococcus faecalis* and *Staphylococcus aureus* and was capable of remodelling PG in non-producing bacteria such as *E. coli* and *Caulobacter crescentus* also (Cava *et al.,* 2011a).

1.7 Aims of this project

As introduced in section 1.5, our lab previously identified a novel D-serine transporter in *E. coli* O157:H7, YhaO, and an adjacent gene encoding a serine deaminase, YhaM. YhaO was found to be capable of transporting D- and L-serine however YhaM showed a preference for D-serine, although its activity was weak, consistent with the inability of O157:H7 to grow on D-serine as a sole carbon source (Moritz & Welch 2006). This lead to the hypothesis that YhaO was involved in transport of D-serine in O157:H7 however the physiological relevance of this system remained elusive. Metabolism of D-serine via the *dsdCXA* system is widespread among ExPEC but not a common trait of InPEC (discussed in

detail in section 1.5.1). D-serine was also implicated in previous studies to play a role in regulating UPEC virulence although the mechanisms behind this were not understood. An O157:H7 *yhaO* mutant was found to negatively affect expression of the T3SS. Together these findings imply a role for YhaO and therefore, D-serine transport, in regulation of virulence. Thus, the three specific aims of this work were:

1) To investigate the physiological role of YhaO in O157:H7
2) To investigate the effects of D-serine on O157:H7 virulence
3) To investigate the broad significance of these findings in relation to other *E. coli*

Using the *dsdCXA* system as a model, chapter 3 aims to address the physiological role of YhaO in O157:H7 including its function, context and role in virulence regulation. Chapter 4 will discuss the substrate of the system, D-serine, and investigate the effects that this molecule has on virulence and cellular morphology. Finally, chapter 5 will delve into the broader significance of the findings by using comparative genomics to investigate the prevalence of D-serine metabolic systems throughout the *E. coli* lineage.

2 Materials and Methods

2.1 Chemicals, growth media and buffers

2.1.1 *Chemicals and molecular reagents*

Chemicals used were purchased from Sigma-Aldrich, Invitrogen and Fisher Scientific unless otherwise stated in the text. Electrophoretic grade Agarose was purchased from Invitrogen. GoTaq DNA polymerase and nucleotides were purchased from Promega and New England Biolabs. Restriction enzymes were purchased from New England Biolabs and Promega. PCR primers were ordered from Invitrogen. 1kb Plus DNA ladder and SeeBlue Plus 2 protein standard were purchased from Invitrogen. Nuclease free water, DEPC treated water, TE buffer and TURBO DNase were purchased from AMBION. RNAprotect reagent was purchased from QIAGEN.

2.1.2 *Growth media*

All media and buffers were prepared using distilled deionised water (ddH_2O) and sterilized accordingly by filtration (0.2 μm) and/or autoclaving. Heat sensitive media components were filter sterilized and added post-autoclaving. Solid media was prepared by adding 15 g/l agar to liquid media before autoclaving. Detailed description of media and buffer recipes was obtained from Molecular Cloning (Sambrook and Russell). MEM-HEPES and DMEM were purchased from Sigma-Aldrich.

Table 2-1 LB recipe (1 litre; pH 7.5)

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

*LB components were added to 950 ml ddH_2O, pH adjusted and topped up to 1 litre before sterilization by autoclaving.

Table 2-2 TB recipe (1 litre; pH 7.5)

Component	Quantity
Tryptone	12 g
Yeast Extract	24 g
Glycerol	4 ml

*TB components were added to 950 ml ddH_2O, pH adjusted and topped up to 1 litre before sterilization by autoclaving.

Table 2-3 SOC recipe (1 litre; pH 7.0)

Component	Quantity
Tryptone	20 g
Yeast Extract	5 g
NaCl	0.5 g
KCl (250 mM)	10 ml

*SOC components were added to 950 ml ddH_2O, pH adjusted and topped up to 1 litre before sterilization by autoclaving. Just prior to use, 5 ml of sterile $MgCl_2$ (2 M) and 20 ml of glucose (1M) were added to the media.

Table 2-4 M9 recipe (1 litre; pH 7.5)

Component	Quantity
M9 Salts (5X)	200 ml
1M $CaCl_2$	2 ml
1M $MgSO_4$	100 µl

*M9 Minimal Media was made up to 1 litre with sterile filtered ddH_2O and supplemented with 0.2% filter sterilized Glucose (20%) post autoclaving.

Table 2-5 10x MOPS (1 litre; pH 7.4)

Component	Quantity
MOPS	83.72 g
Tricine	7.17 g
0.01 M $FeSO_4 \bullet 7H_2O$	10 ml
1.9 M NH_4Cl	50 ml
0.276 M K_2SO_4	10 ml
0.02 M $CaCl_2 \bullet 2H_2O$	0.25 ml
2.5 M $MgCl_2$	2.1 ml
5 M NaCl	100 ml
Micronutrient Stock	0.2 ml

*MOPS ingredients were prepared in the order listed above. MOPS and Tricine were first dissolved in 300 ml ddH_2O and brought to pH 7.4 with 10 M KOH (~20 ml). The volume was brought up to 440 ml with ddH_2O before adding fresh $FeSO_4$. The following ingredients were added in order and the mixture was made up to a final volume of 1 litre before filter sterilizing. The 10x MOPS was aliquoted and frozen at -20°C. The micronutrient stock and MOPS minimal media were prepared as follows:

Table 2-6 MOPS micronutrient stock (50 ml)

Component	Quantity
$(NH_4)_6Mo_7O_{24} \bullet 4H_2O$	0.009 g
H_3BO_3	0.062 g
$CoCl_2$	0.018 g
$CuSO_4$	0.006 g
$MnCl_2$	0.040 g
$ZnSO_4$	0.007 g

*All ingredients were mixed in 40 ml ddH_2O and brought up to a final volume of 50 ml.

Table 2-7 MOPS minimal media (1 litre; pH 7.2)

Component	Quantity
10x MOPS	100 ml
0.132 M K_2HPO_4	10 ml
100x Carbon Source	10 ml

*The mixture was made up to 900 ml and then adjusted to pH 7.2 with 10 M NaOH before being made to a final volume of 1 litre. The media was then filter sterilized and stored at 4°C for up to 1 month.

2.1.3 *Growth media supplements*

Antibiotics were prepared as concentrated stock solutions (Table 2-8) in either ddH_2O or ethanol. Antibiotics prepared in water were filter sterilized (0.2 μm). Stocks were aliquoted and stored at - 20°C.

Table 2-8 Antibiotic stock concentrations

Antibiotic	Stock (mg/ml)	Solvent	Final Conc. (μg/ml)
Ampicillin	100	ddH_2O	100
Kanamycin	50	ddH_2O	50
Chloramphenicol	25	Ethanol	25
Erythromycin	50	Ethanol	500

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Sigma-Aldrich and prepared as a 1 M stock in ddH_2O. IPTG was filter sterilized (0.2 μm) and aliquoted for short-term storage at -20°C. For protein overexpression IPTG was added at a final concentration of 1 mM.

2.1.4 *Buffers*

Buffers were made using ddH_2O, filter sterilized (0.2 μm), autoclaved where necessary and stored at room temperature. TAE, MOPS, TBE and NOVEX Transfer buffer were purchased as concentrated stocks from Invitrogen. TE buffer was purchased from Ambion. Maleic acid buffer, DIG wash buffer, DIG blocking solution and DIG detection buffer were all components of the DIG Wash and Block buffer kit (Roche) used for EMSAs.

Table 2-9 10x TEN solution (1 litre; pH 8.0)

Component	Quantity	Concentration
Tris-HCl	12.1 g	0.1 M
EDTA	2.92 g	0.01 M
NaCL	58.44 g	1 M

*This 10x solution was made in 800 ml ddH_2O, adjusted to pH 8.0 and made up to 1 litre before being filter sterilised and stored at room temperature. A 1x working stock was made by diluting the solution accordingly using sterile filtered ddH_2O.

Table 2-10 1x PBS solution (1 litre; pH 7.4)

Component	Quantity	Concentration
NaCl	8 g	137 mM
KCl	0.2 g	2.7 mM
Na_2HPO_4	1.44 g	10 mM
KH_2PO_4	0.24 g	2 mM

*The mixture was made up to 800 ml in ddH_2O and then adjusted to pH 7.4 with HCl before being made to a final volume of 1 litre. The buffer was then autoclaved and stored at room temperature. Where desired, Tween 20 (Sigma) was added to a final concentration of 0.1%.

Table 2-11 His-purification column buffer (1 litre; pH 7.4)

Component	Quantity	Concentration
NaCl	11.68 g	200 mM
Tris	6.05 g	50 mM
Imidazole	2.72 g	40 mM
Glycerol	100 ml	10 %

*This mixture describes standard His-column buffer. Concentrations were adjusted accordingly for optimization of purification. The mixture was made up to 800 ml in ddH_2O and then adjusted to pH 7.4 with HCl before being made to a final volume of 1 litre. The buffer was then filter sterilized and stored at room temperature.

Table 2-12 His-purification elution buffer (1 litre; pH 7.4)

Component	Quantity	Concentration
NaCl	11.68 g	200 mM
Tris	6.05 g	50 mM
Imidazole	34.03 g	500 mM
Glycerol	100 ml	10 %

*This mixture describes standard His-elution buffer. Concentrations were adjusted accordingly for optimization of purification. The mixture was made up to 800 ml in ddH_2O and then adjusted to pH 7.4 with HCl before being made to a final volume of 1 litre. The buffer was then filter sterilized and stored at room temperature.

Table 2-13 MBP-purification column buffer (1 litre; pH 7.4)

Component	Quantity	Concentration
NaCl	11.68 g	200 mM
Tris	2.42 g	20 mM
EDTA	0.29 g	1 mM
DTT	0.15 g	1 mM

*This mixture was made up to 800 ml in ddH_2O and then adjusted to pH 7.4 with HCl before being made to a final volume of 1 litre. The buffer was then filter sterilized and stored at room temperature.

2.2 Bacterial strains and plasmids used in these studies

Table 2-14 Bacterial strains used in these studies

Strain	Description	Source/Reference
EHEC TUV93-0	Wild type *E. coli* O157:H7 (Stx negative)	Roe lab inventory
Δ*yhaO*	TUV93-0 *yhaO* knockout (Chl^R)	Roe lab inventory
Δ*yhaJ*	TUV93-0 *yhaJ* knockout	Roe lab inventory
Δ*yhaM*	TUV93-0 *yhaM* knockout	Roe lab inventory
Δ*yhaK*	TUV93-0 *yhaK* knockout	Roe lab inventory
Δ*yhiF*	TUV93-0 *yhiF* knockout	Tree *et al.,* 2011
Δ*gadX*	TUV93-0 *gadX* knockout (Kan^R)	This study
Δ*mltD*	TUV93-0 *mltD* knockout (Kan^R)	This study
UPEC CFT073	Wild type UPEC O6:H1:K2	Prof Rod Welch
UPEC Δ*dsdA*	CFT073 *dsdA* knockout	Anfora *et al.,* 2006
UPEC Δ*dsdXcycA*	CFT073 *dsdX/cycA* knockout	Anfora *et al.,* 2006
EPEC E2348/69	Wild type EPEC O127:H6	Prof David Smith
MNEC CE10	Wild type MNEC O7:K1	Prof Kwang Sik Kim
ICC168	Wild type *C. rodentium*	Prof Brett Finlay
ICC168::lux	ICC168 marked with the luciferase cassette (lux) at the 16S locus	This study
C. rodentium Δ*ler*	ICC168 *ler* knockout	This study
C. rodentium Δ*yhaO*	ICC168 *yhaO* knockout	This study
C. rodentium Δ*yhaJ*	ICC168 *yhaJ* knockout	This study
C. rodentium Δ*ler*::lux	Δ*ler*::lux marked	This study
C. rodentium Δ*yhaO*::lux	Δ*yhaO*::lux marked	This study
C. rodentium Δ*yhaJ*::lux	Δ*yhaJ*::lux marked	This study
One Shot TOP10	Commercial *E. coli* storage strain	Invitrogen
StrataClone competent cells	Commercial *E. coli* storage strain	Agilent
BL21 (DE3)	Commercial *E. coli* expression strain	Invitrogen

Table 2-15 Plasmids used in these studies

Plasmid	Description	Source/Reference
p*yhaO*	pWSK29-*yhaO* complementation construct (Amp^R)	Roe lab inventory
p*yhaJ*	pWSK29-*yhaJ* (Amp^R)	Roe lab inventory
p*yhaM*	pWSK29-*yhaM* (Amp^R)	Roe lab inventory
p*yhaK*	pWSK29-*yhaK* (Amp^R)	Roe lab inventory
pLEE1:GFP	LEE1 promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al.,* 2003
pLEE2:GFP	LEE2 promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al.,* 2003
pLEE3:GFP	LEE3 promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al.,* 2003
p*tir*:GFP	*tir* promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al.,* 2003
p*rpsM*:GFP	*rpsM* promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe *et al.,* 2003
pAJR70	pACYC184 with *Bam*HI (5')/*Kpn*I (3') sites in frame with eGFP (Chl^R)	Roe *et al.,* 2003
p*yhaO*:GFP	Promoter region of *yhaO* from EDL933 fused in-frame to GFP in pUC57 (Kan^R)	Genscript
p*yhiF*	pBAD-*yhiF* (Amp^R) complementation construct (Arabinose inducible)	Professor David Gally
p*dsdA*	pACYC-*dsdA* complementation construct (Chl^R)	Roesch *et al.,* 2003
p*dsdX*	pACYC-*dsdX* (Chl^R)	Anfora *et al.,* 2006
p*ihfA*	Constitutively expressed *ihfA* (*rpsM* promoter driven) in pUC57 (Kan^R)	Genscript
pVS45	pBAD-*ler* (Arabinose inducible; Amp^R)	Sperandio *et al.,* 2000
pRFP	pACYC-RFP expression plasmid	Roe lab inventory
pET-28b	Plasmid for overexpression of N-terminal Histidine (6x) tagged recombinant proteins (Kan^R)	Addgene
pKD4	Template plasmid for Lambda Red Mutagenesis containing the Kanamycin cassette (Kan^R)	Datsenko and Wanner 2000
pKD46	Red recombinase expressing plasmid (Amp^R)	Datsenko and Wanner 2000
p16Slux	Integration plasmid for constitutive expression of the lux cassette (Ery^R)	Riedel *et al.,* 2007
pMAL-c5X	Plasmid for overexpression of N-terminal MBP tagged recombinant proteins (Amp^R)	NEB
StrataClone vector	Storage vector for cloning of PCR products	Agilent

2.3 Maintenance and growth of bacteria

2.3.1 *Storage of bacterial strains*

5 ml overnight culture of bacterial strains were started from a single colony on solid agar plates and grown overnight (~16 hours) in LB medium at 30°C, 37°C or 42°C, depending on the strain, with 220 rpm shaking. 0.5 ml of this culture was then added to a cryo-vial containing 1 ml sterile glycerol (40%) and peptone (2%). The stocks were frozen at -80°C for long-term storage.

2.3.2 *Bacterial working stocks*

Bacterial strains from -80°C storage were streaked onto LB agar plates (containing the appropriate antibiotics if required), grown overnight at 37°C and stored at 4°C for up to two weeks. Single colonies from these working plates were used to start subsequent overnight cultures.

2.3.3 *Bacterial growth and calculation of colony forming units (CFU)*

Overnight cultures were used to inoculate pre-warmed media containing any necessary antibiotics or supplements to a starting optical density at 600 nm (OD_{600}) of 0.06 and were cultured at the relevant temperature and 220 rpm. To obtain a growth curve, OD_{600} was measured every 30 minutes on a benchtop spectrophotometer. OD_{600} values were plotted against time using GraphPad Prism 5.

To calculate the CFU of growth experiments aliquots were taken from each bacterial culture and diluted serially in sterile PBS. Solid agar plates were divided into sections, one for each serial dilution, and 3 x 10 µl spots were pipetted in each section according to each serial dilution. Plates were allowed to dry and were incubated overnight at 37°C. Colonies were counted within each spot and multiplied by the relevant dilution factor to gain the number of CFU. For a given dilution, an average of the 3 spots was used. GraphPad Prism 5 was used to comparatively analyse the CFU between samples according to growth rate.

2.4 General molecular techniques

2.4.1 *Plasmid DNA purification*

5 ml overnight culture of a strains containing the desired plasmid vectors were grown at 37 °C and 200 rpm in LB medium. 3-5 ml of these cultures was pelleted in a benchtop microcentrifuge at 8000 rpm for 5 minutes. Plasmid DNA purification was performed using the QIAprep Spin Miniprep Kit according to the manufacturers specifications (Qiagen). Plasmid DNA was eluted in 25 – 50 µl nuclease free water (Ambion) and stored at - 20°C. Plasmid purification was confirmed by agarose gel electrophoresis.

2.4.2 *Agarose gel electrophoresis*

DNA analysis was typically performed on a 1.2% (w/v) agarose gel but was adjusted when necessary. Gels were prepared using 1x TAE buffer. Before allowing liquid agarose/TAE to set fully 10000x SYBR Safe DNA gel stain (Invitrogen) was added. DNA was mixed with 10x BlueJuice loading buffer (Invitrogen) and loaded into the wells alongside a 1Kb Plus DNA ladder (Invitrogen). Gels were run at 100 volts for approximately 40 minutes (this varied depending on the % of the gel) until adequate migration of samples had occurred according the loading buffer dye front. The DNA was visualized using a UV transilluminator.

2.4.3 *Gel purification of DNA*

DNA for purification from agarose gels was extracted using a QIAquick Gel extraction kit (QIAGEN) as per the manufacturers specifications. Gel purified DNA was eluted in 25 – 50 µl nuclease free water (Ambion) and stored at - 20°C. Gel purified DNA was confirmed by agarose gel electrophoresis.

2.4.4 *Polymerase chain reaction (PCR)*

PCR primers were designed using MacVector. For check primers, molecular cloning, transcript mapping and qRT-PCR annealing regions complimentary to template DNA were

typically 20 bp in length, approximately 50% in GC content and had a melting temperature of 55°C - 60°C. Primers for molecular cloning were designed to contain the relevant restriction flanks necessary for ligation reactions. For Lambda Red mediated recombination, primers were designed to have approximately 20 bp complimentary to the pKD4 vector kanamycin-cassette region and 5' flanking regions containing 50 bp complimentary to the regions of the chromosome immediately outside of the start/stop codons of the desired gene for removal. Primers were ordered from Invitrogen and Sigma. Lyophilised primer samples were reconstituted to a final concentration of 100 µM using nuclease free water (Ambion). From this, 10 µM working stocks were used in PCR reactions. Both stocks were stored at -20°C.

Table 2-16 Primers used in these studies

Name	Description	Sequence
yhaO_5_transcript	*yhaO* cDNA transcript validation forward primer	CTTCCTACCTGCATACCTTC
yhaO_3_transcript	*yhaO* cDNA transcript validation reverse primer	GACGCAGGATGTTCATTAC
yhaM_5_transcript	*yhaM* cDNA transcript validation forward primer	GTACTGAACCGATTTCACTG
yhaM_3_transcript	*yhaM* cDNA transcript validation reverse primer	AAGAGAGATCTTTCGCCAG
yhaJ_5_transcript	*yhaJ* cDNA transcript validation forward primer	CACTTAGCTACACCATGCAA
yhaJ_3_transcript	*yhaJ* cDNA transcript validation reverse primer	CTTTTTCGACCATCGGATAC
yhaK_5_transcript	*yhaK* cDNA transcript validation forward primer	GGCCCGGTATACTTTTTC
yhaK_3_transcript	*yhaK* cDNA transcript validation reverse primer	AGACAGGCAAATCTATCAGC
yhaO:yhaM_5_transcript	*yhaO*/*yhaM* co-transcript cDNA forward primer	GGCATCATGATTTTCGCC
yhaO:yhaM_3_transcript	*yhaO*/*yhaM* co-transcript cDNA reverse primer	GGATTTCATTGCAAGGTTCC
yhaM:yhaJ_5_transcript	*yhaM*/*yhaJ* co-transcript cDNA forward primer	GTGGAAAGCGGTGTTAATG
yhaM:yhaJ_3_transcript	*yhaM*/*yhaJ* co-transcript cDNA forward primer	GTAGTGCTTCCAGCGTTAAT
yhaJ_BamHI_5	*yhaJ* amplification forward primer with 5' in-frame *Bam*HI flank for cloning into pET28b	GGATCCATGGCCAAAGAAAGG GCATT
yhaJ_HindIII_3	*yhaJ* amplification reverse primer with 5' in-frame *Hind*III flank for cloning into pET28b	AAGCTTTTATTTTCCGTTAAAAA GTT
yhiF_MBP_5	*yhiF* amplification forward primer with 5' in-frame *Nde*I flank for cloning into pMALc5X	CATATGTTTCTTATAATTACCAG
yhiF_MBP_3	*yhiF* amplification reverse primer with 5' in-frame *Nde*I flank for cloning into pMALc5X	GGATCCTCACACCAGATAATCA ATATG
pMALc5X_check5	Forward primer for pMALc5X insert check	AAGGTGAAATCATGCCGAAC
pMALc5X	Reverse primer for pMALc5X insert check	CTTCGCAACGTTCAAATCC
eilA_BamHI_5	*eilA* amplification forward primer with 5' in-frame *Bam*HI flank for cloning into pAJR70	GGATCCCCTCAGATTATTGACA GTATG
eilA_KpnI_3	*eilA* amplification forward primer with 3' in-frame *Bam*HI flank for cloning into pAJR70	GGTACCCAGTTGCATAGAACG CTC
yhaJ_citro_LR_5	Forward primer for amplification of the Kan^R cassette from pKD4 plus 50 bp of *yhaJ* 5' genomic flank	TTAGCCAGATATGCCCGGTATG TTCAAATTTCCTGAATGAGAAC GATGTGTAGGCTGGAGCTGCT TC
yhaJ_citro_LR_3	Reverse primer for amplification of the Kan^R cassette from pKD4 plus 50 bp of *yhaJ* 3' genomic flank	GTGGGTTCGGGGCAAGCATAT TGCCCCGACTTTTATTGCAGCA GCAAAAACATATGAATATCCTC CTTAG
yhaJ_citro_LR_check_5	Lambda Red forward check primer (*yhaJ*)	TTATGTTAAGGGGCTGGACG
yhaJ_citro_LR_check_3	Lambda Red reverse check primer (*yhaJ*)	GCATATTGCCCCGACTTTTA
yhaO_citro_LR_5	Lambda Red *yhaO* forward primer	CGAACAGTAGTGCGATCCTCG ACGCTTCCACTCCGGCGCGTCG TGCGGGAGTGTAGGCTGGAGC TGCTTC
yhaO_citro_LR_3	Lambda Red *yhaO* reverse primer	CACAACGGATTTTCTGTAGTCT CAAACATAACAACAACCTTAAG

		GGTAGACATATGAATATCCTCC TTAG
yhaO_citro_LR_check_5	Lambda Red forward check primer (*yhaO*)	GGAAACGGCAACGAACAGTA
yhaO_citro_LR_check_3	Lambda Red reverse check primer (*yhaO*)	AGGATAAAACGCTGCCACAA
ler_citro_LR_5	Lambda Red *ler* forward primer	TGACCCATCCATGTAAGGATGA GCTTGTTAATATCTTAATATATA AAAGTGTGTAGGCTGGAGCTG CTTC
ler_citro_LR_3	Lambda Red *ler* reverse primer	TATTATTTCATCTTCCAGTTCAG TTATCGTTATCATTTAATTATTT CATGCATATGAATATCCTCCTT AG
ler_citro_LR_check_5	Lambda Red forward check primer (*ler*)	CCATCCATGTAAGGATGAGC
ler_citro_LR_check_3	Lambda Red reverse check primer (*ler*)	TGCAATGAGCAGTTCCTTTG
gadX_TUV_LR_5	Lambda Red *gadX* forward primer	GCGTGCTACATTAATAAACAGT AATATGTTTATGTAATATTAAG TCAACTGTGTAGGCTGGAGCT GCTTC
gadX_TUV_LR_3	Lambda Red *gadX* reverse primer	TCCTCTTCCCGGTCCCCTATGCC GGGTTTTTTTTATGTCTGAGTA AAACTCATATGAATATCCTCCTT AG
gadX_TUV_LR_check_5	Lambda Red forward check primer (*gadX*)	ATTTGACTTAAGAGGGCGGC
gadX_TUV_LR_check_3	Lambda Red reverse check primer (*gadX*)	TGGAGACGGCAGACTATCCT
mltD_TUV_LR_5	Lambda Red *mltD* forward primer	TCCGTTCGCCGTTATGATCGGT CGTCTTTTAAGCAACTATTGAC GCACACGTGTAGGCTGGAGCT GCTTC
mltD_TUV_LR_3	Lambda Red *mltD* reverse primer	TAAATAAAAAAGGCACCGGGG GAATCGGTGCCTTTCTATTATCT GGTTTGCATATGAATATCCTCC TTAG
mltD_TUV_LR_check_5	Lambda Red forward check primer (*mltD*)	CGTTATGATCGGTCGTCTT
mltD_TUV_LR_check_3	Lambda Red reverse check primer (*mltD*)	GGGAATCGGTGCCTTTCTAT
ihfA_promoter_5	*ihfA* +53 to -255 forward (EMSA)	GTGGTCGCAGAAAACGTTCC
ihfA_promoter_3	*ihfA* +53 to -255 reverse (EMSA)	GCCCAAGCTTATCAAACAGA
ler_promoter_5	*ler* +92 to -359 forward (EMSA)	GAAACGGTTCAGCTTGGTTT
ler_promoter_3	*ler* +92 to -359 reverse (EMSA)	GCTGTAGAACTGCAATTTGCTC
yhiF_promoter_5	*yhiF* -7 to -420 forward (EMSA)	TCCTTGATCCAGTGAACTATCG
yhiF_promoter_3	*yhiF* -7 to -420 reverse (EMSA)	CCTGGTGCGGACATTAACTT
yhaO_promoter_5	*yhaO* -6 to -306 forward (EMSA)	GCGGACTGGCCATTAAAGT
yhaO_promoter_3	*yhaO* -6 to -306 reverse (EMSA)	TCGCTCAATATTTTGTAGGGCT
yhaJ_promoter_5	*yhaJ* +12 to -291 forward (EMSA)	TCGACTTTGGGATAGGTTCG
yhaJ_promoter_3	*yhaJ* +12 to -291 reverse (EMSA)	TCTTTGGCCATTTCGTTCTC
yhaJ_promA_5	*yhaJ* +12 to -84 forward (EMSA)	TTTGGGATAGGTTCGCGG
yhaJ_promA_3	*yhaJ* +12 to -84 reverse (EMSA)	TGGACACTACTTCGACCCGA
yhaJ_promB_5	*yhaJ* -85 to -208 forward (EMSA)	AACAATTTCGGGTCGAAG
yhaJ_promB_3	*yhaJ* -85 to -208 reverse (EMSA)	TTACTACCCGAACTGCCAGG
yhaJ_promC_5	*yhaJ* -182 to -287 forward (EMSA)	TTCCTTTTATGGTAAGGGGC
yhaJ_promC_3	*yhaJ* -182 to -287 reverse (EMSA)	TCTTTGGCCATTTCGTTCTC
nleA_promoter_5	*nleA* +3 to -369 forward (EMSA)	ACAGAGCATTAGCGCAAGGT
nleA_promoter_3	*nleA* +3 to -369 reverse (EMSA)	TCACATATCCGATGTGGACAG
nleB_promoter_5	*nleB* -22 to -392 forward (EMSA)	CATTTTCCTGCAAAAGCTCA
nleB_promoter_3	*nleB* -22 to -392 reverse (EMSA)	ATGTTAATATCGCCCCGTCA
nleG1_promoter_5	*nleG1* -15 to -399 forward (EMSA)	AATCTTTGGGGCAATGGAAC
nleG1_promoter_3	*nleG1* -15 to -399 reverse (EMSA)	TCATTCCAAACACAGGGAGTC
nleG6-2_promoter_5	*nleG6-2* -25 to -399 forward (EMSA)	ACTCAGGTGAAGCACAATCG
nleG6-2_promoter_3	*nleG6-2* -25 to -399 reverse (EMSA)	TCAGTTTGCAGGATGTTTCA
groEL_5_QPCR	qRT-PCR forward *groEL*	AACCGTACTGGCTCAGGCTA
groEL_3_QPCR	qRT-PCR reverse *groEL*	CTTCAACTGCAGCGGTAACA

gapA_5_QPCR	qRT-PCR forward *gapA*	GCTAACCTGAAATGGGACGA
gapA_3_QPCR	qRT-PCR reverse *gapA*	CACCAGCGGTGATGTGTTTA
espB_5_QPCR	qRT-PCR forward *espB*	GGCGCAAAGCTATCAGATTC
espB_3_QPCR	qRT-PCR reverse *espB*	GATGAAATAATCCCGCCAAC
espD_5_QPCR	qRT-PCR forward *espD*	GCGCACAAGCTATCCCTATC
espD_3_QPCR	qRT-PCR reverse *espD*	ACTTTCTGCGCGGAAGTATC
tir_5_QPCR	qRT-PCR forward *tir*	GGGATCTCGTGCGCTATTTA
tir_3_QPCR	qRT-PCR reverse *tir*	AGTCCAGGAACATCACTGGC
ler_5_QPCR	qRT-PCR forward *ler*	GACTGCGAGAGCAGGAAGTT
ler_3_QPCR	qRT-PCR reverse *ler*	ATCCCAGCTCTTGTAAGGTT
nleA_5_QPCR	qRT-PCR forward *nleA*	TAGGATGCCAAGCTGGATTT
nleA_3_QPCR	qRT-PCR reverse *nleA*	AGCTGTTGTTTCACCGCATT
nleG_5_QPCR	qRT-PCR forward *nleG*	ATAGTGAAACGGATGGTCGC
nleG_3_QPCR	qRT-PCR reverse *nleG*	TTCAGGACCACCATTAAGCC
lexA_5_QPCR	qRT-PCR forward *lexA*	GAAGAGGAAGAAGGGTTGCC
lexA_3_QPCR	qRT-PCR reverse *lexA*	ACAAGGAAGGATCGACCTGA
recA_5_QPCR	qRT-PCR forward *recA*	GGTAAAACCACGCTGACGTT
recA_3_QPCR	qRT-PCR reverse *recA*	TACGTGCGTAGATTGGGTCC
sulA_5_QPCR	qRT-PCR forward *sulA*	TGCAGATTAGCCAGCTCTCC
sulA_3_QPCR	qRT-PCR reverse *sulA*	TCAAATCATCTGCCAACCAA
ihfA_5_QPCR	qRT-PCR forward *ihfA*	GGCGCTTACAAAAGCTGAAA
ihfA_3_QPCR	qRT-PCR reverse *ihfA*	CAACCAGTTCTTTGGCATCC
yhiF_5_QPCR	qRT-PCR forward *yhiF*	CCGCGATGAAAAACATTCTG
yhiF_3_QPCR	qRT-PCR reverse *yhiF*	TCTGATGCAACATTACGTCG
yhaO_5_QPCR	qRT-PCR forward *yhaO*	AACCAATCACCGATGAGAGC
yhaO_3_QPCR	qRT-PCR reverse *yhaO*	TTATGAGTATCGGGATGGCG

* Sequences underlined represent the indicated restriction site used for cloning

2.4.5 *PCR reaction setup and conditions*

The DNA polymerase generally used was GoTaq and its relative reaction buffers (Promega). Nuclease free PCR tubes were chilled on ice and reactions were set up on ice prior to amplification unless using a hot start polymerase, which can be prepared at room temperature. PCR reactions were generally set up in 25 µl volume and had a dNTP final concentration of 0.2 mM. When using purified plasmid or genomic DNA as a template, DNA samples were diluted to a 1/100 working stock before addition to the reaction mix. If using a bacterial colony as a template, a single colony was picked from a fresh working plate using a pipette tip and mixed into 50 µl sterile water. This sample was boiled at 95°C for 10 minutes and used as template for the PCR reaction. Generally 0.5 µl of template DNA was sufficient. Thermocycler (Techne TC-512) blocks were cooled to 4°C and lids heated to 100°C prior to starting the PCR reaction cycles to prevent condensation of samples in PCR tubes. PCR products were analysed by agarose gel electrophoresis and stored at -20°C for subsequent use. Reaction mixtures were set up as follows.

Table 2-17 PCR using GoTaq DNA polymerase

Component	Volume
5X GoTag Green Buffer	5 µl
10 mM dNTPs	0.5 µl
10 µM Forward Primer	0.5 µl
10 µM Reverse Primer	0.5 µl
Template DNA	x µl
GoTag DNA Polymerase	0.125 µl
Nuclease Free H_2O	Make up to 25 µl

Table 2-18 PCR using GoTaq Green Master Mix

GoTaq Green Master Mix	12.5 µl
10 µM Forward Primer	0.5 µl
10 µM Reverse Primer	0.5 µl
Template DNA	x µl
Nuclease Free H_2O	Make up to 25 µl

Table 2-19 Standard PCR cycle conditions for GoTaq DNA polymerase (30 cycles)

Incubation Step	Temperature	Time
Initial Denaturation	94°C	3 min
Denaturation	94°C	30 sec
Annealing*	50 – 60°C	30 sec
Extension	72°C	1 min/Kb DNA product
Final Extension	72°C	10 min
Hold	4°C	Hold

* Annealing temperature is typically 5°C below the primer Tm.

2.4.6 *Purification of PCR products*

PCR products were purified using a QIAquick PCR purification kit (QIAGEN) as per the manufacturers specifications. Purified PCR product was eluted in 25 – 50 µl nuclease free water (Ambion) and stored at - 20°C. Purified PCR product was confirmed by agarose gel electrophoresis.

2.4.7 *Preparation of genomic DNA*

Overnight cultures were prepared and 1 ml was used for genomic DNA (gDNA) extraction using the ChargeSwitch™ gDNA Mini Bacteria kit (Invitrogen) according to the manufacturers specifications. DNA was resuspended in nuclease free H_2O (Ambion) and aliquoted for storage at -20°C.

2.5 Molecular cloning and transformation of DNA

2.5.1 *Restriction enzyme digest*

Restriction enzymes were purchased from Promega and NEB (New England Biolabs). For cloning reactions, 1 µg of plasmid DNA (either containing inserted genes or for linearization) was incubated with 1 µl of the desired restriction enzyme, 2 µl of relevant buffer and made up to 15 µl with nuclease free water (Ambion). Reactions were incubated for 2 hours at 37°C (or 30°C for temperature sensitive plasmids). Reactions

were analysed by agarose gel electrophoresis using non-digested controls. Inserts and linearized plasmid DNA of interest were gel purified and either used for ligation straight away or stored at -20°C.

2.5.2 *DNA ligation*

Linearized plasmid and insert DNA of interest were cleaned using a QIAquick PCR purification kit (QIAGEN) and mixed (ratio adjusted empirically) to a volume of 10 µl. 1 µl T4 ligase (NEB) and 1.5 µl T4 buffer (NEB) was then added and the reaction was made up to 15 µl with nuclease free water (Ambion). Ligation reactions were performed either at room temperature for 1 hour or overnight at 16°C. 2-5 µl of the reaction mixture was used for transformation into the desired *E. coli* strain and plated on the correct antibiotic marker for selection screening. Positive colonies were then confirmed by PCR and used for plasmid purification and digestion with the desired restriction enzymes to determine if the insert was ligated successfully. Positively cloned plasmids were sent for DNA sequencing to confirm (Eurofins).

2.5.3 *Preparation of electro-competent cells and transformation*

Electrically competent cells were prepared as follows. 2 ml of an overnight culture was used to inoculate 200 ml of fresh LB media and was incubated at 37°C, 250rpm until an OD_{600} of 0.6 was reached. Culture flasks were placed on ice for 30 minutes post inoculation and then harvested by centrifugation at 4000 g (4°C) for 15 minutes. The supernatant was removed and bacterial pellet was resuspended in 200 ml ice-cold sterile ddH_2O. The cells were harvested again by centrifugation and resuspended in 200 ml ice-cold sterile glycerol (10%). This step was repeated twice before a final harvest and resuspension in 1ml ice cold sterile Glycerol (10%). The cells were aliquoted (50 µl) into sterile PCR tubes and quick frozen for 5-10 minutes in liquid nitrogen or ethanol/dry ice. Aliquots were stored at -80°C for future use.

Aliquots of electrically competent cells were thawed on ice prior to use. Typically 1-5 µl of plasmid DNA was mixed with 50 µl competent cells and left on ice for 5 minutes (Normally 1-2 µl of mini-prepped plasmid DNA was sufficient for successful transformation). The DNA/cell mix was added to a pre chilled electroporation cuvette and

shocked at 2500 volts in an Eporator electroporator (Eppendorf). 500 µl of pre warmed SOC recovery media was added to the cuvette instantly after electroporation and the cuvettes were allowed to static incubate for 2 hours at 37°C. Cell suspensions (both 100 µl neat and 100 µl of a 1:1 cells and SOC) were then plated out on LB agar with the appropriate selective marker.

2.5.4 *Heat-shock transformation of DNA*

Aliquots of commercially purchased chemically competent cells were thawed on ice prior to use. Typically 1-5 µl of plasmid DNA was mixed 50 µl of competent cells and left on ice for 20 minutes. Samples were heat shocked at 42°C for 45 seconds then placed back on ice for 1 minute exactly. 500 µl of pre warmed SOC media (Invitrogen) was added to each reaction and they were incubated shaking at 37°C for 2 hours. As with electro-transformation, cell suspensions were then plated out on LB agar with the appropriate selective marker.

2.6 Transcriptome analysis

2.6.1 *Total RNA extraction*

Total RNA extraction was carried out using an RNeasy Mini Kit (Qiagen) with minor modifications. 10 ml of bacterial culture grown under the desired conditions was mixed 1:2 with RNA protect reagent (Qiagen), vortexed briefly and incubated at room temperature for 5 minutes before harvesting the bacterial pellet by centrifugation at 5000 g for 15 minutes. The bacterial pellet was resuspended gently by pipetting in 200 µl TE buffer containing lysozyme (15 mg/ml) and 20 µl proteinase K. The samples were incubated on an orbital shaker at room temperature for 10 minutes, vortexing for 10 seconds every 2 minutes. 700 µl buffer RLT (containing β-mercaptoethanol 10 µl/ml) was added to each sample and vortexed vigorously. 500 µl of ethanol (100%) was then added and mixed. 700 µl of this sample was loaded on to an RNeasy Mini spin column in a 2 ml collection tube and centrifuged for 15-20 seconds at 12000 g. A further 700 µl was loaded

on to the same column and centrifuged again. The flow through was discarded. An on-column DNase step was performed next using DNase I (Qiagen). 350 µl buffer RW1 was added to the column and centrifuged for 15 seconds. 80 µl of DNase solution (10 µl DNase I stock solution and 70 µl buffer RDD) was added directly to the column membrane and incubated at room temperature for 15 minutes. A final 350 µl of buffer RW1 was added to the column and centrifuged again, discarding the flow through. 500 µl buffer RPE was added to the column and centrifuged again for 15 seconds. This was repeated with a 2 minute centrifugation. The column was added to a fresh collection tube and centrifuged for 1 minute to dry the membrane. The column was finally placed in a 1.5 ml collection tube and 50 µl of RNase-free TE buffer (Ambion) was added to the membrane for 1 minute before centrifugation.

The samples were then analysed on a NanoDrop 2000 (Thermo Scientific). Samples should have a 260/280 absorbance reading of ~1.8-2.0 and a 260/230 reading of ~2.0-2.2, indicating high quality, contaminant free RNA. Samples were treated to a second DNase step using DNase TURBO (Ambion). 10x TURBO buffer was added to each sample and DNase TURBO was added to the appropriate concentration (1 µl for 10 µg in a 50 µl reaction). Samples were incubated for 30 minutes at 37°C and inactivated at 75°C for 10 minutes. RNA samples were quick frozen to avoid degradation and stored at -80°C in appropriate aliquots to avoid repeated freeze thawing of samples.

2.6.2 *mRNA enrichment from total RNA*

Enriching for mRNA was carried out using a MICROBexpress kit (Ambion) with minor modifications. 20 µl of total RNA was used (approximately 10 µg in concentration) for the enrichment process. The remainder of the total RNA was frozen and stored at -80°C. For precipitation of total RNA, 0.1 volumes of 3 M sodium acetate and 3 volumes of ice cold 100% ethanol were added to each sample, mixed well and allowed to precipitate at -20°C overnight or -80°C for 2 hours. Samples were then centrifuged in a table-top centrifuge at 13000 g for 30 minutes (4°C) and supernatant was carefully removed. The pellets were resuspended in 1 ml ice-cold 70% ethanol. Samples were centrifuged again for 10 minutes and the supernatant removed. The 70% ethanol wash step was carried out twice. After the final spin and the pellet was resuspended in 15 µl TE buffer. The samples were mixed with 200 µl binding buffer in a 1.5 ml RNase free eppendorf tube and mixed gently. 5 µl of

Capture Oligo mix was added and vortexed gently. Samples were heated at 70°C for 10 minutes to denature secondary structures then incubated at 37°C for 1 hour to allow annealing of rRNA to Capture Oligo beads.

During the above incubation, the Oligo MagBeads were prepared. 65 µl of the bead solution (per sample) was added to a fresh 1.5 ml RNase free eppendorf tube. Tubes were placed on a MagnaRack (Invitrogen) and left for 3 minutes to allow beads to pellet fully. The supernatant was carefully removed and discarded. The tubes were removed from the rack and 65 µl of nuclease free water was added to each tube and gently vortexed to resuspend the beads. The beads were once again pelleted via the MagnaRack and the supernatant removed. The pellet was resuspended in 65 µl binding buffer to equilibrate the beads and pelleted once again. The pellet was resuspended in a final 65 µl binding buffer and heated to 37°C. The wash solution was also heated to 37°C at this stage. Heated bead mixture was added to the annealed mixture containing the RNA. The combined mixture was gently vortexed and incubated at 37°C for 15 minutes. The tubes were placed again on the MagnaRack and the beads allowed to pellet out of the mix. The supernatant (which contains enriched mRNA) was carefully removed and added to a fresh 1.5 ml RNase free tube on ice. The beads were washed as above and the supernatant was added to the RNA pool on ice. Final precipitation of the enriched mRNA was carried out by adding $1/10^{th}$ volume of 3 M Sodium Acetate and $1/50^{th}$ volume of Glycogen, mixing gently then adding 3 volumes of ice cold 100% ethanol. Samples were precipitated at -20°C for 2 hours then centrifuged at 13000 g (4°C) before carefully removing the supernatant. mRNA pellets were resuspended in 750 µl ice cold 70% ethanol and briefly vortexed before centrifuging for 5 minutes. This step was performed twice. The supernatant was removed completely and the pellet was air dried for 5 minutes before being resuspended in 25 µl TE buffer. Enriched mRNA samples were split into aliquots and stored at -80°C. Samples were sent to the University of Glasgow Polyomics Facility for analysis of quality on the Agilent 2100 Bioanalyzer and subsequently used for cDNA synthesis.

2.6.3 *RNA-seq transcriptome analysis*

The transcriptome of an organism is defined as the entire set of mRNA (transcripts), small RNA (sRNA) and non-coding RNA fragments in a cell at a specific point in time. The transcriptome is used primarily to gain detailed information on gene expression levels and patterns, under various conditions, on a global scale. Classically, transcriptomes were analysed using hybridization-based arrays and fluorescently labelled cDNA. Array based transcriptomics were favoured for being high-throughput and inexpensive, however, suffered from major disadvantages such as requirements on pre-existing genome sequence information and high levels of background signal from the hybridization arrays. They also suffered from a lack of dynamic range from the hybridization signals. Next-generation sequencing has facilitated a shift from array-based transcriptomics to a direct sequencing approach known as RNA-seq. RNA-seq involves stabilizing and purifying the total RNA from a sample (generally a population but the technology has been used for single-cells and meta-transcriptomic analysis), enriching for the mRNA or sRNA/non-coding RNA fractions and converting these specific RNAs to cDNA, which is then directly sequenced via a desired sequencing platform. Sequenced reads are then mapped *in silico* to a reference genome allowing a global analysis of transcript abundance under specific sets of conditions. This high-throughput approach allows high resolution, dynamic transcriptome analysis without the drawbacks of array-based technologies (Ozsolak & Milos 2011).

Enriched mRNA samples for RNA-seq analysis were sent to the University of Glasgow Polyomics facility for cDNA synthesis and sequencing on an Illumina Genome Analyser IIx and the University of Edinburgh GenePool facility for sequencing on an Illumina HiSeq 2000. Sequencing was performed with single ended reads, 75 bp (Genome Analyser) or 100 bp (HiSeq) read length respectively and 6 samples per lane. Raw reads were QC checked for degradation in read quality using FastQC (Babraham Bioinformatics). Raw reads were then imported into CLC Genomics Workbench 7 (CLC Bio), trimmed accordingly and aligned to the EDL933 reference genome (NCBI accession number: NC_002655.2). Transcriptome data from chapter 3 was analysed for differential gene expression using CLC Genomics Workbench also. These experiments consisted of triplicate biological replicates for both the wild type (TUV93-0) and the Δ*yhaO* mutant. Total mapped reads for each annotation were normalized using the Empirical analysis of

DGE function, which is an implementation of the 'exact test' used in the EdgeR Bioconductor package (Robinson *et al.,* 2010). Transcriptome data from chapter 4 was normalised and analysed for differentially expressed genes using the Bioconductor packages DEseq and EdgeR (Anders & Huber 2010, Robinson *et al.,* 2010). This experiment consisted of triplicate biological replicates for the wild type (TUV93-0) and duplicate biological replicates for wild type supplemented with D-serine. DEseq and EdgeR analyses were performed by Dr. Scott Beatson (University of Queensland) and Dr. Zofia Jones (University of Glasgow). Once normalised, differentially expressed genes were identified by a fold change in expression with a significance cutoff of $p \leq 0.05$ and a false discovery rate (FDR) of 5%. Specific genes of interest were validated for differential expression using qRT-PCR.

2.6.4 *Visualisation of RNA-seq data*

Volcano plots illustrating distribution of data were generated using CLC Genomics Workbench. Heatmaps were generated using the online tool CIMminer (Genomics and Bioinformatics Group, NCI). Custom RNA-seq coverage graphs were generated using EasyFig as follows (Sullivan *et al.,* 2011). Annotated sequence files were exported from CLC Genomics Workbench in ace format and converted to a custom coverage format using a python script (getCov454.py) provided by Mitchell Sullivan (University of Queensland). Coverage files were subsequently trimmed using the following terminal command to obtain coverage files for desired regions of the genome:

```
cat infilename | head –n$end | tail –n$length > outfilename
```

Infilename refers to the file name of the custom coverage file generated from the getCov454.py script whereas outfilename refers to the name of the newly trimmed coverage file. n$end refers to the end position of the trimmed region on the genome and n$length refers to the length of the region of interest. Corresponding regions were downloaded from the NCBI database in genBank format to be used as the annotation reference for figure generation in EasyFig.

2.6.5 *Quantitative real time PCR (qRT-PCR)*

Validation of differential gene expression by qRT-PCR was carried out using KAPA SYBR® FAST Universal qRT-PCR master mix (KAPA Biosystems) and M-MLV Reverse Transcriptase (Promega). Total RNA was extracted as above and was quantified using a NanoDrop 2000. Samples for comparison were normalized to a concentration of 50 ng/µl using TE buffer (Ambion). qRT-PCR analysis was performed using a one-step reaction; cDNA synthesis first followed by qRT-PCR. Individual reactions were performed in triplicate to account for technical variance and each gene to be analysed was performed in biological triplicate. Two controls were used in the qRT-PCR analysis; a no-template control to screen for nucleic acid contamination in the reagents and primer-dimer formation, and a no Reverse-Transcriptase (RT) reaction to assess if any contaminating DNA was present in RNA samples. qRT-PCR reactions were carried out using the ECO™ Real-Time PCR System (illumina) according to the manufacturers specifications and the data were analysed according to the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen 2001).

Table 2-20 qRT-PCR using KAPA SYBR® FAST reaction mix

Component	Volume
2X KAPA SYBR® FAST Universal Master Mix	10 µl
ROX High	0.4 µl
M-MLV RT	0.3 µl
5’ Primer	0.5 µl
3’ Primer	0.5 µl
Template RNA	1 µl
Nuclease Free H2O	7.2 µl

2.7 Genetic techniques

2.7.1 *Lambda Red-mediated recombination*

Lambda Red-mediated recombination is a technique used to make single step gene knockouts in bacteria via PCR products. The technique relies on the exploitation of the bacteriophage Lambda Red genes (*exo, bet and gam*), which encode a 5’-3’ dsDNA exonuclease, an ssDNA annealing protein and an anti-exonuclease protein, respectively

(Datsenko & Wanner 2000, Murphy & Campellone 2003). PCR products of known antibiotic resistance markers used in Lambda Red mutagenesis are generated using 50 bp overhangs that directly flank the gene of interest to be knocked out. Once, transformed into cells expressing the Lambda Red genes, the Gam protein prevents host degradation of the foreign DNA by RecBCD and SbcCD nucleases. Exo functions to convert the 5' ends of the PCR product into 3' ssDNA overhangs and Bet then uses these overhangs to pair them with the complimentary target, ultimately resulting in recombination of the gene of interest with a known resistance marker (Datta *et al.,* 2006). The Lambda Red protocol utilizes pKD4 (Kan^R) to generate an ~1.6 kb kanamycin cassette flanked with 50 bp regions of homology to the gene of interest and pKD46 (Amp^R), the arabinose inducible Red recombinase expression plasmid (Datsenko & Wanner 2000).

Prior to the recombination reaction, concentrated PCR product of the gene of interest plus 50 bp overhangs must be obtained. 10 PCR reactions (100 µl each) were set up and checked for single band amplification by agarose gel electrophoresis. The PCR reactions were pooled and mixed with an equal volume of Phenol:Chloroform:Isoamilic alcohol (P:C:I) (Sigma), vortexed gently and incubated shaking at room temperature for 10 minutes. Samples were centrifuged at 10000 rpm for 10 minutes and the upper aqueous phase was carefully removed and mixed with equal volume P:C:I in a fresh tube. After this second extraction, the aqueous phase obtained was washed in an equal volume of Chloroform:Isoamilic alcohol (C:I). The aqueous phase obtained at this point was mixed with 2.5 volumes of ice-cold ethanol (100%) and 0.3 M potassium acetate and precipitated overnight at -20°C. The DNA pellet was obtained by centrifugation at 13000 rpm and washed with 1 ml of room temperature ethanol (70%). The pellet was air dried and resuspended in 12 µl of nuclease free water.

Strains of interest for gene knockouts were transformed with pKD46 prior to the recombination protocol. Strains containing pKD46 were grown as overnight cultures in LB at 30°C and 250 µl was inoculated into 25 ml of SOB containing 2.5 ml L-arabinose (1 M) and 100 µg/ml ampicillin. Cultures were grown at 30°C to an OD_{600} of 0.6. Arabinose acts as the inducer for lambda Red induction. Cultures were harvested at 5000 rpm, 4°C for 10 minutes before removing the supernatant and resuspending the pellet in the same volume of ice-cold sterile water. This step was repeated twice, the second time resuspending in 1 ml of ice-cold water in an eppendorf tube. Cells were washed a further

3 times in 1 ml of ice-cold water with 2 minute centrifugations in between at 10000 rpm in a table top micro centrifuge (4°C). Pellets were resuspended in a final volume of 100 µl. 50 µl of cells was mixed with 4 µl concentrated PCR product of the desired gene product and added to pre-cooled electroporation cuvettes. These samples were kept on ice for 10 minutes before being electroporated. 1 ml of pre-warmed SOC media (Invitrogen) was immediately added and samples were recovered at 30°C static for 3 hours. 200 µl of each transformation reaction was finally plated out on LB agar (containing 50 µg/ml kanamycin) and incubated overnight at 42°C. The remainder of the transformation reaction was left at room temperature overnight and plated on LB agar (containing 50 µg/ml kanamycin) the following day. Colonies were used as PCR template to screen for successful recombination (Table 2-16). Colonies were re-streaked on both LB-Kan and LB-Amp plates for 2-3 days at 41°C after the reaction to allow curing of the mutant strain of temperature sensitive pKD46. Successful mutants have a 1.6 Kb PCR product in place of the gene of interest and only grow on kanamycin.

2.7.2 *Luciferase (lux) marking of strains*

Strains generated using Lambda Red recombination for use in the *C. rodentium* infection model were luciferase (*lux*) marked according the method described by previously (Riedel *et al.,* 2007). This method describes the site-specific integration of the bacterial *lux* operon into the 16S locus of the chromosome. This allows constitutive expression of luciferase without any exogenous substrate addition.

p16S*lux* (Ery^R) was transformed into the relevant strains via electroporation. Transformed bacteria were plated on LB agar containing erythromycin and incubated for 24-48 hours at 30°C. Colonies were checked for bioluminescence using the IVIS system. Positive clones were inoculated into 5 ml LB and grown overnight at 30°C. 5 ml of fresh LB with erythromycin was inoculated with these overnight cultures at a concentration of 1:1000 and incubated overnight at 42°C. 100 µl from these cultures were plated onto LB agar containing erythromycin and incubated overnight at 42°C. p16S*lux* is non permissive at 42°C and subculturing at this temperature in the presence of erythromycin forces a low frequency homologous recombination event resulting in integration into the 16S locus of the genome. p16S*lux* integration was PCR confirmed and colonies were checked for luminescence using the IVIS system or by inoculation into LB containing erythromycin,

growth overnight at 37°C and subsequent luminescence analysis using a FLUOstar Optima Fluoresence Plate Reader (BMG Labtech, UK).

2.8 Biochemical techniques

2.8.1 *SDS-PAGE*

Secreted, whole cell and purified protein samples were analysed and prepared for western blotting by SDS-PAGE using the NuPAGE® system (Invitrogen). Samples were mixed with 4x NuPAGE® LDS loading buffer and boiled for 10 minutes at 95°C before loading onto NuPAGE® 4-12% Bis-Tris pre-cast gels (Invitrogen). Samples were generally loaded in 10 µl aliquots (concentrations estimated using a NanoDrop 2000) and SeeBlue® plus2 pre stained protein standard (Invitrogen) was using as a molecular weight marker. SDS-PAGE gels were run using NuPAGE® MOPS running buffer (Invitrogen) for 50 minutes at 180 volts or until the dye front was just before the end of the gel. SDS-PAGE gels were stained for 1 hour on an orbital shaker using Coomassie blue stain (Table 2-21) and destained using 2-3 changes of destain solution (Table 2-22) for 3 hours or dH2O overnight.

Table 2-21 Coomassie Blue recipe (1 litre)

Component	Volume/Amount
ddH2O	500 ml
Methanol	400 ml
Acetic Acid	100 ml
Coomassie blue R250	0.5 g

Table 2-22 Destain recipe

Component	Volume/Amount
ddH2O	500 ml
Methanol	400 ml
Acetic Acid	100 ml

Protein bands to be analysed using tandem mass-spectrometry (MS/MS) were carefully excised fro SDS-PAGE gels using a clean, sterile scalpel blade and placed into individual sterile 1.5 ml microcentrifuge tubes. These bands were prepared and analysed by Dr Richard Burchmore in the University of Glasgow Polyomics division and Mr Dougie Lamont in the University of Dundee Proteomics division.

2.8.2 *Western blotting*

Proteins for western blot analysis were transferred from SDS-PAGE gels to an Amersham Hybond ECL nitrocellulose membrane (GE Healthcare) using the XCell SureLock® Transfer system (Invitrogen). Proteins were transferred for 1 hour at 30 volts followed by blocking in 5% skimmed milk (Marvel) made with PBST for at least 1 hour. Primary antibody was made up to the specified concentration (Table 2-23) using 1% skimmed milk in PBST and incubated on the membrane for 1 hour at room temperature on an orbital shaker. Membranes were then washed 3 times in 50 ml PBST for 10 minutes each before incubation with secondary horseradish-peroxidase (HRP) conjugated antibody for 1 hour. Secondary antibody was generally added at a concentration that was half of what was used for the corresponding primary antibody. Membranes were washed a further 3 times before development with SuperSignal West Pico Chemiluminescent ECL Substrate (Thermo Scientific).

Table 2-23 Antibodies and relevant concentrations used in these studies

Antibody	Source	Primary Concentration	Secondary Concentration
GroEL	Rabbit	1/25000	1/30000
YtfM	Rabbit	1/2000	1/4000
Tir	Mouse	1/1000	1/2000
EspD	Mouse	1/3000	1/6000
RecA	Rabbit	1/5000	1/7000
FliC H7	Rabbit	1/1000 or 1/500 for Immunofluorescence	1/2000 or 1/1000 for Immunofluorescence
O157	Rabbit	1/500 for Immunofluorescence	1/1000 for Immunofluorescence
His-HRP	Rabbit	1/5000	N/A

2.8.3 *Cellular fractionation by ultracentrifugation*

Bacterial cells containing His-tagged proteins of interest (prepared by Dr. Mads Gabrielsen) were cultured to an OD_{600} of 0.8 and harvested by centrifugation at 5000 g for 20 minutes and resuspended in 15 ml per litre column buffer. Samples were homogenized with a cOmplete protease inhibitor tablet (Promega), $MgCl_2$ and DNAse I before being lysed. Lysis was carried out using a French Press at 950 bar of pressure. Whole cell lysate was centrifuged at 8000 g for 30 minutes to remove unbroken cells. The supernatant was transferred to an ultracentrifuge tube and was subjected to 45000 g for 45 minutes to break open lysed membrane particles. The supernatant from this step was retained for identification of cytosolic proteins. The membrane particle pellet was resuspended in a high salt buffer (3M NaCl) to separate membrane-associated proteins from membrane-integrated proteins. The solution was centrifuged again at 45000 g for 45 minutes. The supernatant was retained for identification of membrane-associated proteins whereas the pellet was resuspended for identification of membrane-integrated proteins.

2.8.4 *Sample preparation for metabolomics analysis*

Preparation of metabolites was adapted from Creek *et al.,* 2011. Bacterial cultures were grown to an OD_{600} of 0.6 in MEM-HEPES and were subsequently diluted to 0.1 in fresh, warm MEM-HEPES to a final volume of 40 ml. Cultures were quenched by rapid submersion in ethanol and dry ice to cool samples to 4°C. Care was taken to shake samples vigorously whilst cooling to avoid freezing and possible cell lysis. Cells were then centrifuged at 3500 rpm (4°C) for 10 minutes. Bacterial cell pellets were resuspended in 1 ml of supernatant and transferred to an ice-cold 1.5 ml microcentrifuge tube before obtaining the final cell pellet at 3500 rpm (4°C) for 5 minutes. The supernatant was completely removed at this stage and metabolites were extracted from the cell pellet by addition of 200 µl of chloroform:methanol:water (1:3:1) at 4°C. Pellets were resuspended by pipetting and tubes were vortexed vigorously for 1 hour at 4°C. Cellular debris and precipitate was removed by a final centrifugation at 13000 rpm (4°C) for 5 minutes. 180 µl of the supernatant was transferred to a fresh tube and stored at -80°C until LC-MS analysis. Dr Karl Burgess and Dr Stefan Weidt of the University of Glasgow Polyomics division carried out the targeted chiral-metabolite analysis. Samples were separated on a

ChiroBiotic T2 column (2 mm x 25 cm) using an isocratic flow at 85% ethanol and 15% water at 200 µl/min. Amino acids were detected using atmospheric pressure chemical ionisation in an OrbiTrap elite (Thermo UK) at 70000 resolution. Source settings were configured as described previously (Desai & Armstrong 2004). Analysis was performed using ExactFinder software. Samples of cell-free medium, spent medium removed at the cell harvest step and extraction solvent were used as controls for analysis. Biological samples were performed in triplicate.

2.8.5 *Sample preparation for peptidoglycan analysis*

Desired strains were cultured in the appropriate conditions in 400 ml of MEM-HEPES to late exponential phase (OD_{600} of 0.6). Cultures were then cooled rapidly in an ice water bath (4°C) for 10 minutes at which stage cells were harvested at 3750 rpm in a pre-cooled centrifuge (4°C). Supernatant was removed and bacterial cell pellets were resuspended in 6 ml of ice cold PBS. This cell suspension was dropped slowly in a drop-wise fashion to a 100 ml Erlenmeyer flask containing 6 ml of boiling 8% SDS solution (pre-heated) and a magnetic stirrer. Drop-wise addition of cells is essential for rapid inactivation of endogenous autolysins. Samples were boiled for 30 minutes with stirring. Drops of PBS were added if the volume of the solution decreased too much due to evaporation. After 30 minutes samples are cooled to room temperature and sent for analysis in 15 ml falcon tubes. Professor Waldemar Vollmer's group in the University of Newcastle performed HPLC-MS analysis of the peptidoglycan.

2.8.6 *Overexpression and purification of recombinant proteins*

Proteins cloned into pET28b (N-terminal His-tagged) were used for overexpression in *E. coli* BL21 (DE3) cells. Overnight cultures were used to inoculate LB media and cultures were grown to an OD_{600} of 0.6 at 37°C, 220 rpm. At this point, expression cells were induced with IPTG at a final concentration of 1 mM and were left at room temperature, 220 rpm overnight. Bacterial cells were harvested at 5000 rpm for 20 minutes and resuspended in 15 ml per litre His-purification column buffer. Cell harvest fractions were either used straight away for purification or stored at -20°C. Samples for purification were homogenized with a cOmplete protease inhibitor tablet (Promega), $MgCl_2$ and DNAse I before being lysed. Lysis was carried out using a French Press at 950 bar of pressure. After

lysis, cell fractions were centrifuged at 5000 rpm for 20 minutes and the supernatant was removed whilst retaining the cell debris also. The supernatant was then used for immobilized metal affinity ion chromatography (IMAC) using an AKTA-prime. The supernatant was loaded onto a 5 ml HisTrap HP Ni column (GE Healthcare), which was pre washed with ddH2O and equilibrated with His-purification column buffer (flow through was retained). Loaded samples were washed with His-purification column buffer before a gradient elution was used to elute His-tagged protein from the Ni column using His-purification elution buffer and each fraction was retained. All fractions (pellet, flow through, wash, elution fractions) were analysed via SDS-PAGE to identify the quality of the purification. Further purification of desired elution samples was carried out using size-exclusion chromatography (SEC). Elution fractions from the IMAC stage were dialyzed overnight in SEC buffer before being re-purified according to size using an AKTA-prime using a Superdex S200 column. Elution fractions corresponding to the desired size of the protein to be purified were retained.

Proteins cloned into pMALc5X (N-terminal MBP-tagged) were used for overexpression in NEBexpress cells (New England Biolabs). Overnight cultures were used to inoculate LB media containing 0.2 % glucose and cultures were grown to an OD_{600} of 0.6 at 37°C, 220 rpm. At this point, expression cells were induced with IPTG at a final concentration of 0.3 mM and were left at room temperature, 220 rpm overnight. Bacterial cells were harvested at 5000 rpm for 20 minutes and resuspended in 15 ml per litre MBP- column buffer. Cell harvest fractions were either used straight away for purification or stored at -20°C. Samples for purification were homogenized with a cOmplete protease inhibitor tablet (Promega), $MgCl_2$ and DNAse I before being lysed. Lysis was carried out using a French Press at 950 bar of pressure. After lysis, cell fractions were centrifuged at 5000 rpm for 20 minutes and the supernatant was removed whilst retaining the cell debris also. Small-scale purification of MBP-tagged protein was carried out using Amylose magnetic beads (New England Biolabs) as per the manufacturers specifications with minor changes to the protocol. 500 µl of bead solution was pelleted using a magnetic stand and washed twice in MBP-purification column buffer. Beads were subsequently mixed with 5 ml of expression culture supernatant and incubated with agitation at 4°C for 1 hour. The beads were then pelleted using a magnetic rack and washed 3 times as above. 100 µl of MBP-purification column buffer containing 10 mM maltose was added to the beads, vortexed and incubated for 10 minutes with agitation. The beads were

pelleted again and the supernatant retained. A second elution step was performed and the fractions were pooled. All fractions were then analysed by SDS-PAGE for identification of overexpressed protein.

2.8.7 *Electrophoretic mobility shift assay (EMSA)*

EMSA assays were performed using the DIG Gel Shift Kit, 2nd Generation and the DIG Wash and Block buffer set (Roche). Buffer compositions are described in section 2.1.4 and all other components were supplied with the kit. Prof. Steve Goodman kindly supplied purified IHF and IhfA used in EMSA experiments. Purified His- and MBP- tagged proteins were purified as above and used for EMSA assays. Promoter regions (~300-400 bp) containing putative DNA binding sites were amplified by PCR, confirmed by agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit (Qiagen) for use as binding probes in EMSA experiments. For DIG labelling of EMSA probes, 100 ng of purified promoter regions were adjusted to a final volume of 10 µl using sterile TEN buffer (~4 pmol/µl). 4 µl labelling buffer (5x), 4 µl CoCl2-solution, 1 µl DIG-ddUTP solution and 1 µl terminal transferase (400 U) were added to each probe on ice and then incubated for 15 minutes at 37°C before returning to ice. The reaction was terminated by addition of 2 µl 0.2 M EDTA (pH 8.0) and 3 µl of ddH2O was added to a final volume of 25 µl in which the concentration of DIG labelled probe is 0.155 pmol/µl (4 ng/µl). The efficiency of probe labelling was assessed by making serial dilutions of DIG labelled probes and spotting 1µl onto a piece of positively charged nylon membrane (Roche) before UV cross-linking and direct detection by luminescence, as described below. Probes were cross-linked to the membrane by UV trans-illumination (20 minutes, tested empirically). For EMSA experiments 4 µl binding buffer, 1 µl poly [d(I-C)] (1 µg/µl), 1 µl poly L-lysine (0.1 µg/µl), 2 µl DIG labelled probe (30 fmol/µl) and 11 µl ddH2O were added per reaction. Increasing concentrations of purified protein were added to each reaction and the mixture was incubated at room temperature for 15 minutes before placing back on ice and addition of 5 µl loading buffer to each reaction. 20 µl samples were loaded immediately to a pre-electrophoresed NOVEX 6% DNA retardation gel (Invitrogen) and ran at 100 volts for 2 hours in 0.5x TBE buffer (Invitrogen). Electro-transfer of samples to a positively charged nylon membrane (Roche) was performed using the NOVEX X-cell blot II module (Invitrogen) for 1 hour at 30 volts in 0.5x TBE buffer. Once complete, membranes were UV cross-linked for 20 minutes then rinsed in DIG wash buffer. Membranes were blocked

using DIG blocking solution for 30 minutes, followed by a 30 minute incubation in DIG antibody solution. Membranes were washed twice in DIG wash buffer for 15 minutes each then equilibrated for 2 minutes in DIG detection buffer. Membranes were placed DNA side up on cling film and 750 µl of DIG CSPD working solution was added directly to each membrane. Cling film was lightly placed on top and excess liquid was squeezed out before sealing the edges. Sealed membranes were incubated at 37°C for 10 minutes before exposure to x-ray film. Exposure was typically performed for 1-3 minutes. EMSA experiments were performed in triplicate.

2.9 Phenotypic characterisation techniques

2.9.1 *E. coli O157:H7 secreted protein profiling*

Extraction of secreted proteins from *E. coli* O157:H7 was carried out as previously described (Tree *et al.,* 2011). Overnight cultures were used to inoculate 50 ml of pre-warmed MEM-HEPES to a starting OD_{600} of 0.06 and were cultured at 37°C, 220 rpm until at an OD_{600} of 0.7. Cultures were harvested at 5000 rpm for 20 minutes at which point the supernatant only was carefully removed. Supernatants were filtered through a 0.2 µm filter (Fisher) before ice-cold trichloroacetic acid 6.1 N (TCA) was added to a final concentration of 10% (v/v) in a fume hood and samples were mixed and left at 4°C overnight to allow secreted proteins to precipitate out of solution. BSA can be added to a final concentration of 4 mg/ml to act as a co-precipitant and maximize recovery of the secreted protein fraction. After overnight precipitation, secreted protein fractions were harvested by centrifugation at 6000 g, 4°C for 60 minutes. The supernatant was carefully removed and the inside of the tubes were blotted dry without disruption of the pellet. The pellet was allowed to air dry for 30 minutes. Secreted protein pellets were resuspended in 100 µl of Tris-HCl 1.5 M, pH 8.0 and analysed via SDS-PAGE as described above.

2.9.2 *Bacterial motility assay*

Bacterial motility assays were adapted from the swarm method described previously (Wolfe & Berg 1989). Bacterial cells were grown to an OD_{600} of 0.6 in MEM-HEPES at 37°C, 220 rpm containing either no supplement or 1 mM D-serine. 5 µl of this culture was transferred to the centre of a 0.25% Tryptone agar plate and plates were incubated, agar side up, for 8 hours at 31°C. Diameter of the bacterial swarm was measured. Samples were tested in triplicate.

2.9.3 *Chemical 'in-plug' and 'in-capillary' assays*

To assess if D-serine was acting as a chemorepellant, the chemical 'in-plug' and 'in-capillary' assays was used (Tso & Adler 1974). Chemical plugs were prepared by adding 4% agar made with ddH_2O to an equal volume of chemotaxis media (2x) containing twice the desired concentration of D-serine at 70°C. This solution was poured to a petri dish and allowed to set before 0.5 cm plug were cut out using a sterile cork-borer. TUV93-0 cells were grown to an OD_{600} of 0.6 in Tryptone broth and harvested at 3000 rpm. The pellet was resuspended once in equal volume warm (50°C) chemotaxis medium and suspended to a final OD_{600} concentration of 0.1 in 30 ml 0.3% agar before immediately pouring this mixture to a 90 mm petri dish. Plates were cooled for 5 minutes at room temperature before the plugs of desired concentration were added to the still molten 0.3% agar. Once set, plates were transferred to a 30°C static incubator and incubated for 8 hours, with regular observations. Plugs containing L-serine were used as a control for positive chemotaxis. For the 'in-capillary' method 1 mm glass capillaries were loaded with 1 mM L or D-serine and inserted into a chamber containing TUV93-0 growing in TB at 30°C for one hour. Capillaries were sealed at one end to prevent overflow. After incubation the capillaries were rinsed with a gentle flow of sterile PBS and the contents were emptied into sterile PBS and serially diluted. Dilutions were plated out on LB agar and CFU were calculated after growth at 37°C.

2.9.4 *Serine migration swarm assay*

DeLoney-Marino *et al.* described a method of using an excess of L-serine to disrupt the migration of bacteria in a motility assay (Deloney-marino *et al.,* 2003). Motility assays

were set up for TUV93-0, as described previously, using Tryptone broth soft agar (0.25%) plates. Cells were allowed to migrate for 4 hours before a 1 M excess of both L-serine and D-serine was spotted (3 µl) just before the outermost ring of the swarming bacteria. Plates were incubated for a further 4 hours with regular observations.

2.9.5 *Transcriptional reporter fusion assay*

Population based gene expression analysis was performed using WT TUV93-0 transformed with various GFP-promoter transcriptional fusion constructs (Table 2-15) that have been described elsewhere (Roe *et al.,* 2003). Strains were grown overnight in LB and subcultured into MEM-HEPES or DMEM at 37°C. 200 µl samples were transferred to a black flat-bottomed 96-well microtitre plate and analysed for GFP expression using a FLUOstar Optima Fluorescence Plate Reader (BMG Labtech, UK). Readings for GFP fluorescence and OD_{600} were taken hourly. WT TUV93-0 containing empty pAJR70 was used as a negative control to measure for background fluorescence. All samples were performed in triplicate from separate growth cultures. Readings for fluorescence were plotted against OD_{600} using GraphPad Prism 5 (GraphPad Software, USA) and fluorescence at OD_{600} was used as a statistical comparator.

2.10 Microscopy techniques

2.10.1 *Immunofluorescence microscopy*

E. coli strains were cultured to an OD_{600} of 0.6. 200 µl of culture was mixed 1:1 with 4% (w/v) paraformaldehyde (PFA) and allowed to incubate at room temperature for 20 minutes. 20 µl of each sample was dotted onto a single well each of a multi-well microscope slide and allowed to dry. 20 µl of PBS + 0.2% BSA with the appropriate concentration of primary antibody (Table 2-23) was pipetted onto each spot and incubated for 1 hour, shaking gently at room temperature. Each spot was rinsed 3 times with PBS + 0.1% BSA. α-rabbit AlexaFluor conjugate (Invitrogen) antibodies were used as secondary at 1:1000 dilution and incubated for 1 hour at room temperature with gentle

shaking. For AlexaFluor incubation, slides must be incubated in the dark. Wells were rinsed a further 3 times with PBS + 0.1% BSA before being dried and a cover slip mounted using Dako fluorescent mounting fluid (Invitrogen). Slides were left in the dark overnight to set and imaged on a Zeiss M1 Axioskopp microscope. Cells were imaged at either x40 or x100 magnification and were captured and analysed using Volocity software (PerkinElmer) or ZEN Pro (Zeiss).

2.10.2 *Transmission electron microscopy*

Margaret Mullin, head of the University of Glasgow electron microscopy service, prepared bacterial and tissue samples for imaging. Specimens were fixed in a 2% glutaraldehyde, 2% paraformaldehyde, 0.1 M sodium cacodylate solution with added 0.1% magnesium and 0.05% calcium chloride for 15 minutes during dissection at RT and then placed on ice for a further 45 minutes. The specimens were washed three times for 5 minutes each in sodium cacodylate buffer rinse with added chlorides on ice before post fixing with 1% osmium tetroxide/0.1 M sodium cacodylate buffer for 1 hour at room temperature then given three 10 min buffer rinses. Specimens were treated with 1% tannic acid for one hour followed by a 10 minutes in 1% sodium sulphate before dehydrating through an ethanol series 20, 30, 50, 70, 90% staining with 2% uranyl acetate at the 30% stage for one hour then given three changes in 100% ethanol and three changes in dried 100% ethanol (3 A molecular sieve). The specimens were treated with three x 10 minutes propylene oxide then left overnight in a 1:1 PO:Araldite 502/812 resin mix (TAAB Laboratories). The specimens were given several changes of pure resin during the next day before embedding in flat mould trays and polymerised for 48 hours at 60˚C. Semi-thin sections of 0.35 µm were cut using a Leica Ultracut E using a diamond histo-ultramicrotome knife. Sections were stained with 1% toluidine blue to find appropriate areas on the light microscope then ultra thin sections of 60-70 nm were cut using a ultramicrotome diamond knife and collected on 100 mesh formvar coated copper grids. Ultra thin sections were contrast stained with 2% methanolic uranyl acetate for five minutes followed by Reynolds lead citrate for five minutes. Images were captured on a FEI 200 kV Tecnai T20 using Gatan Digital Microscopy software.

2.10.3 *Scanning electron microscopy*

The samples were processed as for transmission electron microscopy until the osmium tetroxide stage where the protocol for OTOTO (Malick and Wilson, 1975) was then followed. Samples were dehydrated through the ethanol series up to dried 100% ethanol then critical point dried using a POLARON E3000. Samples were mounted onto aluminium pin stubs with silver dag conductive paint and then Gold Palladium coated using a POLARON SC515 sputter coater. Samples were viewed on a JEOL6400 SEM and tiff Images captured using Olympus Scandium imaging software.

2.11 *In vitro* and *in vivo* infection models

2.11.1 *E. coli cell adhesion assay*

HeLa cells were cultured in MEM-HEPES supplemented with 1 mM L-glutamine and 10% foetal calf serum (prepared by Claire McQuitty). Glass coverslips were seeded with $4x10^4$ HeLa cells in a multi-well plate and incubated overnight in MEM-HEPES at 37°C with 5% CO_2. TUV93-0 (transformed with a pRFP plasmid if not carrying any other plasmids) and CFT073 were used for adhesion assays. Bacterial cultures for infection were grown in MEM-HEPES at 37°C until at an OD_{600} of 0.6. Seeded cells were washed with fresh media and infected with 100 µl bacterial culture (adjusted to OD_{600} of 0.1) in 500 µl fresh MEM-HEPES with or without 1 mM D-serine. Plates were centrifuged at 400 rpm for 3 minutes and incubated at 37 °C with 5% CO_2 for 2 hours. Wells were washed with fresh media to remove unbound bacteria and incubated as above for a further 3 hours. Wells were washed 3 times with sterile PBS before fixing for 20 minutes with 250 µl PFA (2%). Wells were washed again with PBS. Wells were incubated for 5 minutes with 250 µl of Triton X-100 (0.5%) and washed. Host cell actin was stained with Phalloidin-488 (Invitrogen) for 1 hour. Wells were washed as described and if *E. coli* strains were not carrying the pRFP plasmid bacterial cells were stained with anti-O157 antibody (MAST) for 1 hour followed by a further wash stage and incubation with alexa-fluor 555 (Sigma) for 1 hour. Wells were washed a final time before mounting on microscope slides with fluorescent

mounting medium (Dako). Slides were imaged using a Zeiss M1 Axioskopp microscope and data was acquired and deconvoluted using Zen Pro software. Adhesion assays were performed in triplicate. Bacterial counts for adhesion assays were analysed in GraphPad Prism 5.

2.11.2 *Infection of BALB/c mice with C. rodentium*

C. rodentium is routinely used as a surrogate model for EHEC *in vivo* infection modelling as it is a natural mouse pathogen and infects the host via a LEE encoded T3SS (Deng *et al.,* 2001, Schauer & Falkow 1993). Lambda Red recombination was used to generate relevant knockout strains for *in vivo* analysis and these strains were subsequently LUX marked for IVIS imaging (Riedel *et al.,* 2007). Strains used for challenge were cultured overnight in LB at 37°C then subcultured into DMEM and grown to an OD_{600} of 0.8 (corresponding to a CFU/ml count of approximately 5 x 10^9). Bacterial cells were harvested via centrifugation and resuspended in sterile PBS to a concentration of 1 x 10^{10}. 100 µl of this was used to orally challenge Balb-C mice in replicates of 5 per strain. Faecal samples were taken from each mouse on specified days, weighed, resuspended in sterile PBS (10 µl per mg of faeces) and serially diluted. Serial dilutions were plated out and the CFU/mg of faeces was calculated. Mice were also IVIS imagined for bioluminescence to monitor the extent and location of colonization in the mouse. Handling of mice was performed by Dr. Gillian Douce (University of Glasgow). Data was analysed in GraphPad Prism 5.

2.12 Bioinformatics and statistical analysis

2.12.1 *Bioinformatic and database tools*

Nucleotide and amino acid sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov) and Colibase (Chaudhuri *et al.,* 2004). Alignments were performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Nucleotide sequences were translated and reverse complemented using ExPASy (http://web.expasy.org/translate/). Cloning strategies were designed using MacVector. Determination of gene names from non-

annotated references, gene products and protein functions were carried out using a combination of NCBI, RegulonDB (Salgado *et al.,* 2013) and PortEco (http://www.porteco.org). Prediction of bacterial promoter sequences were determined using PhiSITE Promoter Hunter online software (http://www.phisite.org). Genomic context figures were generated using EasyFig, according to the author's guidelines (Sullivan *et al.,* 2011).

2.12.2 *Bioinformatic analysis of locus carriage in E. coli*

Dr. Robert Goldstone performed Bioinformatic analyses described in this section based upon an experimental design strategy devised in collaboration with Prof. David Smith (University of Glasgow/Moredun Research Institute). The nucleotide sequences for 159 core-genes in *E. coli* were elaborated as described in a recent publication (Goldstone *et al.,* 2014). Briefly, at each iteration, the core gene set (CGS), initialised as the nucleotide sequence for genes present in MG1655, was aligned to the next *E. coli* genome sequence using BLASTn (Altschul *et al.,* 1990). Genes aligning at >70% identity and >80% of the length of the coding sequence were retained in the CGS for use in the next iteration. This analysis resulted in the identification of 159 genes. The nucleotide sequences of each core gene were extracted from the *E. coli* genomes, aligned by Muscle (Edgar 2004), concatenated, and a maximum likelihood tree calculated using PhyML (Guindon & Gascuel 2003) under the general time reversible model of nucleotide substitution. Dendrograms were visualised using the APE package (Paradis *et al.,* 2004) and implemented in R (R Development Core Team 2012).

The amino acid sequences for genes present within O-island 148, encoding the LEE T3SS system, the *cscRAKB* locus and the *yhaOKJ* locus were collected from the sequence for EDL933 (NCBI accession number: NC_002655.2) and amino acid sequences for DsdX, DsdC and DsdA were collected from the sequence for CFT073 (NCBI accession number: NC_004431.1). These sequences were iteratively aligned against the *E. coli* genomes using tBLASTn and called as present in a genome if the protein sequence aligned at >70% identity over >80% to the translated nucleotide sequence of the genome in question. The distribution of the genes across the *E. coli* core-genome dendrogram was visualised using the Diversitree package (Fitzjohn 2012) implemented in R (R Development Core Team 2012).

For statistical analysis to take account of genes incorrectly labelled as 'absent' due to the splitting of coding sequences across contigs, the LEE was called as 'present' if an isolate possessed 21 or more of the following: *espF, Z5102, escF, Z5014, espB, espD, espA, sepL, escD, eae, Z5111, tir, Z5113, Z5114, Z5115, sepQ, Z5117, Z5118, escN, escN, Z5121, sepZ, Z5123, escJ, Z5125, escC, cesD, Z5128, Z5129, Z5131, escU, escT, escS, escR, Z5136, Z5137, Z5138, Z5139, Z5140, Z5142* and *Z5143*. *dsdCXA* was called as present only if all three of *dsdX, dsdC and dsdA* were identified. This was similarly applied to the *yhaOKJ* and *cscRAKB* loci.

The 'Odds Ratio' of LEE/*dsdCXA*/*cscBKAR* carriage was calculated by totalling the number of strains in a group (either locus-present or locus-absent) and dividing the query total of one scenario by the query total of the comparator to gain a numerical value of the likelihood of a particular scenario. Statistical significance for the association between carriage of *dsdCXA, cscBKAR* and the LEE was evaluated by Fisher's exact test.

2.12.3 *Statistical analysis*

Statistical analysis for comparison of two samples was carried out using a student's unpaired T-test in GraphPad Prism unless otherwise stated. Error bars were calculated using the standard error of the mean. RNA-seq analysis was performed using CLC Genomics Workbench (CLC Bio). qRT-PCR analysis was carried out in Microsoft Excel using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001). Significance levels are depicted in figures as NS, *, ** or *** denoting no significance, $p \leq 0.05$, $p \leq 0.01$ or $p \leq 0.001$ respectively.

3 Investigating the physiological role of YhaO in O157:H7

3.1 Introduction

The ability to sense and respond to niche-specific physiological signals is emerging as a key factor in *E. coli* virulence regulation. The LEE PAI encodes a T3SS that is temporally regulated by a complex network of mechanisms. This allows careful and controlled expression of this colonisation factor only when it is needed. Furthermore, how environmental signals feed into this network is beginning to be understood. Researchers have begun to characterise how sugars, SCFAs and hormone-like signals are sensed and modulate expression of the LEE (discussed in section 1.4.6).

Recent work in the Roe lab identified YhaO as a D-serine transporter in O157:H7 involved in regulation of virulence. The ability to sense and respond to environmental D-serine has been characterised in UPEC and more recently *S. saprophyticus* and the metabolite has emerged as an important fitness trait and regulator of virulence (Anfora *et al.,* 2007, Korte-Berwanger *et al.,* 2013). UPEC thrive in a nutrient poor environment that is abundant in D-serine, however, the intestinal niche is a more carbohydrate rich environment and thus InPEC have evolved to utilise sucrose instead of D-serine (Moritz & Welch 2006). The identification of YhaO as a potential virulence determinant was therefore quite surprising and begged many questions in regard to its physiological role. Why does O157:H7 maintain and express the ability to transport toxic D-serine, a trait that was previously thought to have been lost through evolution? Why can O157:H7 transport or sense D-serine but not metabolise it? How is the expression of the LEE affected by YhaO function?

This chapter aims to address these points through characterisation of YhaO and its associated genes. The *dsdCXA* system offers an interesting comparator as this locus encodes a D-serine transporter (DsdX) as well as a D-serine deaminase (DsdA) and an LTTR controlling the system (DsdC) (Anfora & Welch 2006). *yhaO* was found to be adjacent to *yhaM*, a gene encoding a D-serine deaminase, and a putative LTTR gene (*yhaJ*) and predicted cofactor (*yhaK*) were found to be located in proximity to these genes. Thus, this system will be referred to as the *yhaOMKJ* locus form here on.

3.2 Genomic, transcriptional and cellular context of *yhaOMKJ*

Genomic context of the *yhaOMKJ* locus was initially assessed by sequence comparison with *dsdCXA* from UPEC CFT073 using BLASTp (Figure 1-15). YhaO (NCBI accession NP_289683.1) is a D-serine transporter and its coding sequence shares 38% identity over 19% of the protein coding sequence with DsdX (NCBI accession NP_754782.1). YhaO however is also found on the CFT073 genome (c3869; NCBI accession NC_004431.1) suggesting these are two independent systems but is annotated as a pseudogene (sharing 97% identity over 89% of the protein coding sequence). Adjacent to *yhaO* is the D-serine deaminase *yhaM.* YhaM (z4462; NCBI accession NP_289682.1) shares 30% identity with DsdA (NCBI accession NP_754783.1) over 22% of the protein coding sequence but interestingly shares 99% sequence identity over 100% of the protein coding sequence with YhaM from CFT073 (c3867; NCBI accession NC_004431.1). The LTTR DsdC (NCBI accession NP_754781.1) is absolutely required for expression of *dsdX* and *dsdA* in CFT073 (Nørregaard-madsen & Fall 1995) and is positioned immediately upstream of these two genes (Figure 1-15). Similarly a putative LTTR, *yhaJ* (NCBI accession NP_289679.1) is positioned upstream of *yhaM* and *yhaO*, although in the opposite genomic orientation. YhaJ shares 42% sequence identity with DsdC over 39% of the protein coding sequence but is also found intact on the CFT073 genome (99% sequence identity over 100% of the protein coding sequence). Interestingly, *yhaJ* lies back to back with a characterised protein, YhaK (NCBI accession NP_289680.1) in divergent directions. The crystal structure of YhaK has been solved and it was identified as a novel bicupin capable of redox activity. Interestingly, the expression of YhaK was found to be upregulated 12-fold in a UPEC strain VR50 when grown in human urine. The authors proposed that YhaK may function as a DNA binding cofactor of YhaJ although this was never explored (Gurmu *et al.,* 2009). A fourth gene implicated in serine metabolism, z4464, is labelled as a putative L-serine deaminase in EDL933 but shares low sequence similarity with DsdA (41% over 6% of the coding sequence). Due to this and its position it was not studied further. Other genes in the locality encode functions not related to D-serine sensing or metabolism and these include putative cytochromes (*yhaH* and *yhaI*), an endoribonuclease (*yhaR*) and a hypothetical gene (*yhaL*). The genomic context of *yhaOMKJ* in EDL933, CFT073, MG1655 (K-12) and E2348/69 (EPEC) is illustrated in Figure 3-1.

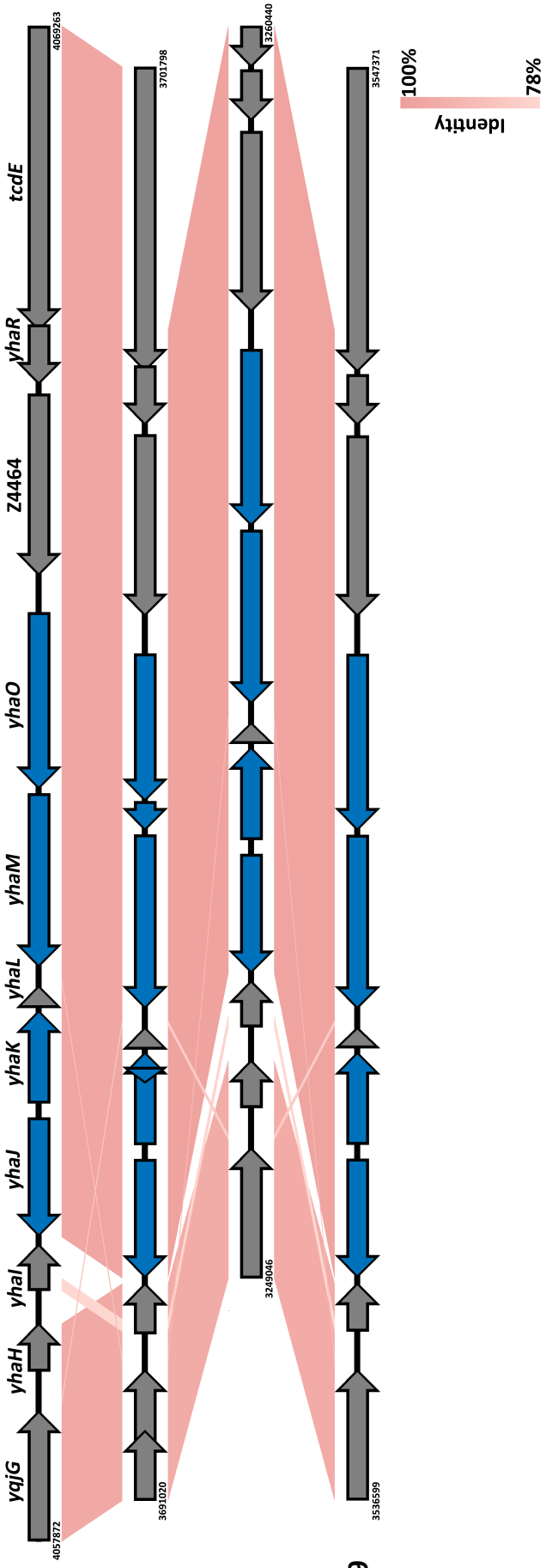

Figure 3-1 Genomic context of the *yhaOMKJ* locus. Diagram illustrating the genomic positioning of the putative D-serine sensory locus *yhaOMKJ* in four contrasting *E. coli* strains – EHEC EDL933, UPEC CFT073, K-12 MG1655 and EPEC E2348/69. The regions have been aligned and sequence identity is indicated by the legend. *yhaOMKJ* are labelled in blue and surrounding genes labelled in grey. The genomic position of each region is labelled at either end of the diagram. This figure was generated using EasyFig (Sullivan et al., 2011).

The transcriptional context of *yhaOMKJ* was assessed by isolation of total RNA, cDNA generation by reverse transcriptase PCR (RT-PCR) and specific PCR amplification using overlapping primers of each ORF (Figure 3-2). As expected, *yhaJ* and *yhaK* were expressed as individual transcripts by virtue of their location on opposing DNA strands. *yhaM* and *yhaJ* did not overlap in their transcription but interestingly *yhaO* and *yhaM* were found to be co-transcribed. This is similar to *dsdX* and *dsdA* however the location of the system LTTR, *dsdC*, is not identical to *yhaJ*, suggesting a unique transcriptional mechanism to *dsdCXA*.

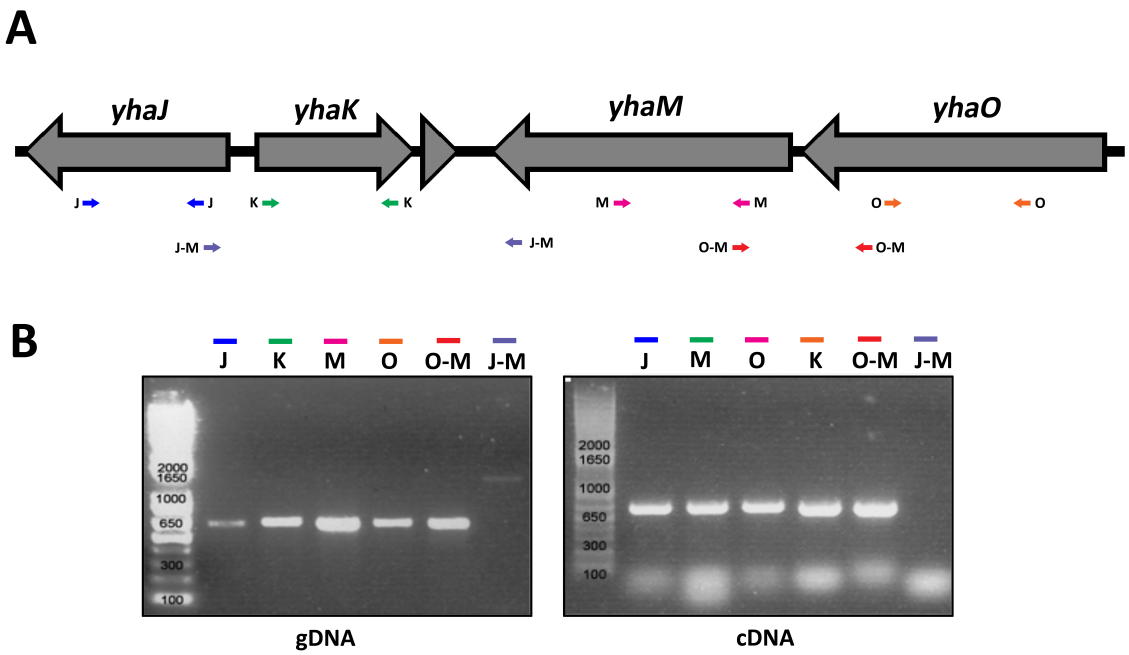

Figure 3-2 Transcriptional context of *yhaOMKJ*. Transcript map generated for the *yhaOMKJ* locus. (A) Primers used to test each possible transcript are colour coded and labelled according to their amplicon. J = *yhaJ;* K = *yhaK*; M = *yhaM*; O = *yhaO*; J-M = *yhaJ*/*yhaM* co-transcript; O-M = *yhaO*/*yhaM* co-transcript. (B) RT-PCR amplification of gDNA controls and cDNA template generated from transcribed mRNA.

Finally, the cellular context of each member of this locus was addressed. Dr. Mads Gabrielsen generated His-tagged recombinant proteins of YhaO, YhaM, YhaK and YhaJ. Cellular fractionation was performed on BL21 DE3 cells containing an expression plasmid of each protein. Cellular fractions were separated by SDS-PAGE and analysed by western blot using anti-His antibody (Figure 3-3). As expected, YhaM, YhaK and YhaJ were found in the cytoplasm whereas YhaO was a membrane bound protein. This agreed with biochemical characterisation of YhaO functioning as a serine transporter protein and

YhaM as a deaminase enzyme. Furthermore, an article highlighting global protein-protein interactions in K-12 identified YhaK as an interacting partner of YhaJ in pull-down assays. This finding was confirmed by Yeast-2 Hybrid screening and Far-western analysis (Dr. Andrew Roe, unpublished data) providing convincing evidence that YhaJ and YhaK interaction was of physiological relevance (Arifuzzaman *et al.,* 2006).

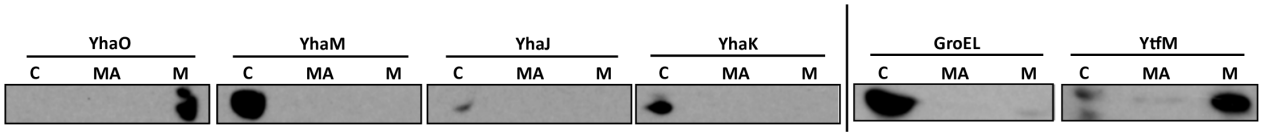

Figure 3-3 Cellular localisation of YhaOMKJ. Western blot analysis of His-tagged YhaOMKJ expressed in cells that were subsequently subject to fractionation by ultracentrifugation. The cellular compartments of each identified protein are labelled C, MA or M corresponding to cytoplasmic, membrane-associated or membrane-bound fractions respectively. Control blots for cytoplasmic GroEL and the OM protein YtfM are indicated also.

3.3 Characterisation of YhaO physiological function

Sequence and biochemical analysis of YhaO indicated that it functioned as an IM D-serine transporter. To assess its function in a physiologically relevant setting a CFT073 double mutant strain (*ΔdsdXΔcycA*) deleted for the D-serine transporter DsdX and a second transporter capable of uptake, CycA, was used as a tool. This strain was generously supplied as a gift by Professor Rodney Welch (Anfora & Welch 2006). As discussed, wild type CFT073 is capable of growing on MOPS minimal media agar plates supplemented with D-serine as a sole carbon source via the *dsdCXA* system, an ability that O157:H7 lacks. Similarly the *ΔdsdXΔcycA* mutant cannot grow on D-serine as a sole carbon source because it can no longer take up D-serine from the environment. A plasmid carrying a constitutively expressed copy of *yhaO* (p*yhaO*) was introduced into the mutant strain and tested for its ability to grow on D-serine. YhaO allowed growth on D-serine indicating that it indeed functions as a D-serine transporter in a physiologically relevant setting. A plasmid carrying a trans-acting copy of *dsdX* (p*dsdX*) was also introduced into ΔdsdXΔcycA as a control experiment and this also permitted growth on D-serine as a sole carbon source (Figure 3-4). The experiment was repeated on MOPS plates supplemented with glucose as a positive control, which allowed growth of all strains tested (data not shown).

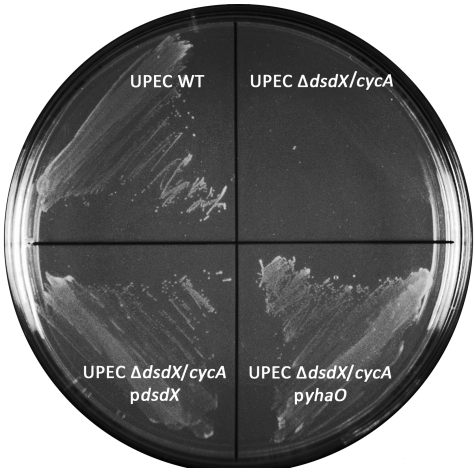

Figure 3-4 Analysis of YhaO function in CFT073. Growth on MOPS minimal media agar plates containing D-serine as a sole carbon source was assessed using UPEC wild type (CFT073) and a *ΔdsdXΔcycA* double transporter mutant. *ΔdsdXΔcycA* was also complemented with either p*dsdX* or p*yhaO* in trans, as labelled.

3.3.1 *Expression of YhaO and CFT073*

The *yhaO* gene in CFT073 is annotated as being a pseudogene however, resequencing of the *yhaO* ORF revealed this was in fact an annotation error. Given that a constitutively expressed YhaO from O157:H7 strain EDL933 is capable of transporting D-serine in a CFT073 *ΔdsdXΔcycA* background the expression levels of *yhaO* relative to each other in CFT073 and TUV93-0 (an EDL933 Stx negative strain) were assessed by qRT-PCR. Total RNA was isolated from bacteria cultured in MEM-HEPES as these conditions were previously known to induce *yhaO* expression. Expression of *yhaO* in CFT073 was > 3 fold lower than in TUV93-0 ($p \leq 0.05$) indicating that irrespective of its functionality it is less expressed in CFT073 rather than in TUV93-0 under the conditions tested (Figure 3-5).

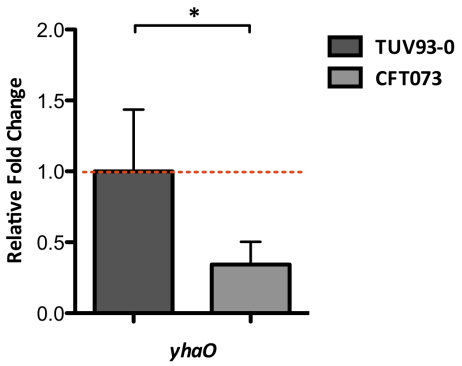

Figure 3-5 Relative expression of *yhaO*. qRT-PCR analysis of *yhaO* expression levels in TUV93-0 and CFT073. The red line indicates baseline expression of *yhaO* in TUV93-0 for which expression in CFT073 was compared. A * indicates $p \leq 0.05$.

Using the experiment described in Figure 3-4 as a model, the function of YhaM was also assessed. Biochemical analysis performed by Dr. Mads Gabrielsen using purified YhaM indicated that it can function as a D-serine deaminase although its activity is low (personal communication). To determine if YhaM functions in a physiological environment, a plasmid carrying a constitutively expressed *yhaM* (p*yhaM*) was transformed into a CFT073 *ΔdsdA* mutant strain (supplied by Professor Rodney Welch) and tested for its ability to allow growth on MOPS minimal agar plates supplemented with D-serine as a sole carbon source. Interestingly, YhaM complemented in trans to *ΔdsdA* did not allow growth on D-serine but complementation with *dsdA* in trans (p*dsdA*) did. Furthermore, introducing p*dsdA* to the TUV93-0 background allowed growth on D-serine (Figure 3-6). These results suggest that YhaM is not active at a sufficient level to allow growth on D-serine in either CFT073 or TUV93-0. This data is consistent with O157:H7 being unable to grow on D-serine as a carbon source and all together suggest redundancy or an alternative role for YhaM.

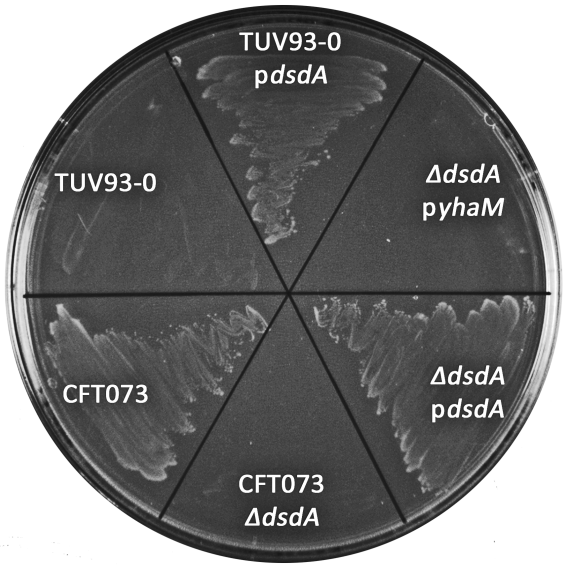

Figure 3-6 Analysis of DsdA and YhaM function in TUV93-0 and CFT073. Growth on MOPS minimal media agar plates containing D-serine as a sole carbon source was assessed using TUV93-0, CFT073 and a CFT073 *ΔdsdA* mutant. TUV93-0 was complemented with p*dsdA* in trans. *ΔdsdA* was complemented with either p*dsdA* or p*yhaM* in trans also.

3.4 The role of YhaO in LEE regulation

With functionality of YhaO in relation to D-serine established the next aim was to investigate the role of YhaO in LEE regulation. A mutant strain of TUV93-0 deleted for *yhaO* (*ΔyhaO*) was generated using Lambda Red mutagenesis by Dr. Dai Wang prior to this project and was tested for growth/motility defects and a role in LEE expression (Datsenko & Wanner 2000).

3.4.1 *Analysis of growth and motility in the ΔyhaO mutant*

The *yhaO* mutant was cultured in MEM-HEPES and DMEM tissue culture media at 37°C (LEE inducing conditions) and compared to growth of TUV93-0 under the same conditions to assess if deletion of *yhaO* had any detrimental effects on overall fitness under conditions that promote LEE expression. The growth behaviour of *yhaO* was virtually identical to that of the wild type with no observable detrimental effects (Figure 3-7A). Growth experiments were also performed in rich media (LB) but this data is not shown. As MEM-HEPES and DMEM do not promote expression of flagella in *E. coli* (Tobe *et al.,* 2011) motility was assessed by way of an adapted swarm method (Wolfe and Berg, 1989). Bacteria were cultured in MEM-HEPES to an OD_{600} of 0.6 before a 5 µl aliquot was taken and inoculated into the centre of a 0.25% Tryptone agar plate. Motility plates were incubated overnight at 31°C and the diameter of migration was measured. Deletion of *yhaO* had no effect on motility when compared to TUV93-0 (Figure 3-7B).

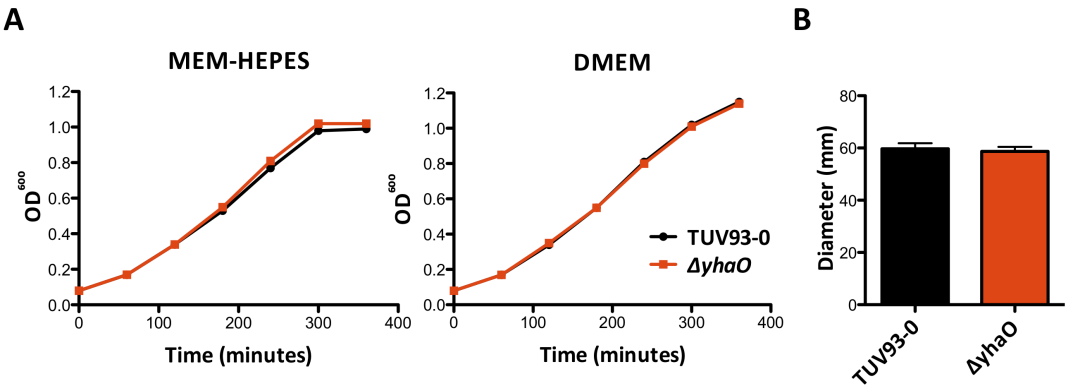

Figure 3-7 Deletion of *yhaO* does not affect growth or motility. (A) Growth of *ΔyhaO* was assessed in MEM-HEPES and DMEM by comparison with TUV93-0. Growth curves are representative graphs of triplicate experiments. (B) Diameter of migration was measured from swimming motility assays comparing TUV93-0 and *ΔyhaO* (indicated in black and red respectively). Motility assays were performed in triplicate.

3.4.2 *Deletion of yhaO negatively affects LEE associated type 3 secretion*

Under LEE inducing conditions *yhaO* expression is increased. In order to address a potential role for YhaO in virulence and confirm the previous observation that *ΔyhaO* had an impaired type 3 secretion (T3S) profile, analysis of LEE encoded effector secretion was carried out. TUV93-0 and *ΔyhaO* were cultured in MEM-HEPES at 37°C to an OD_{600} of 0.7. Secreted proteins were precipitated from the growth culture supernatant and separated via SDS-PAGE for comparison of secretion profiles. *ΔyhaO* displayed reduced secretion of known LEE encoded effectors including EspD, EspA and Tir when compared with the TUV93-0 profile (Figure 3-8A). This result was confirmed by western blot analysis using antibodies specific for the detection of EspD and Tir. To address the question of whether these proteins are being expressed in the cytoplasm or simply whether T3S is physically impaired western blot analysis of whole cell lysates were performed using anti-EspD antibody. This experiment confirmed that EspD expression was indeed reduced when compared with the wild type. As a control, anti-GroEL antibody was used to confirm normal housekeeping expression in the whole cell lysate thus eliminating the possibility that differential loading of samples occurred (Figure 3-8B). The negative effects on T3S observed in the *ΔyhaO* mutant could also be reversed by complementation with p*yhaO,* as demonstrated in Figure 3-8.

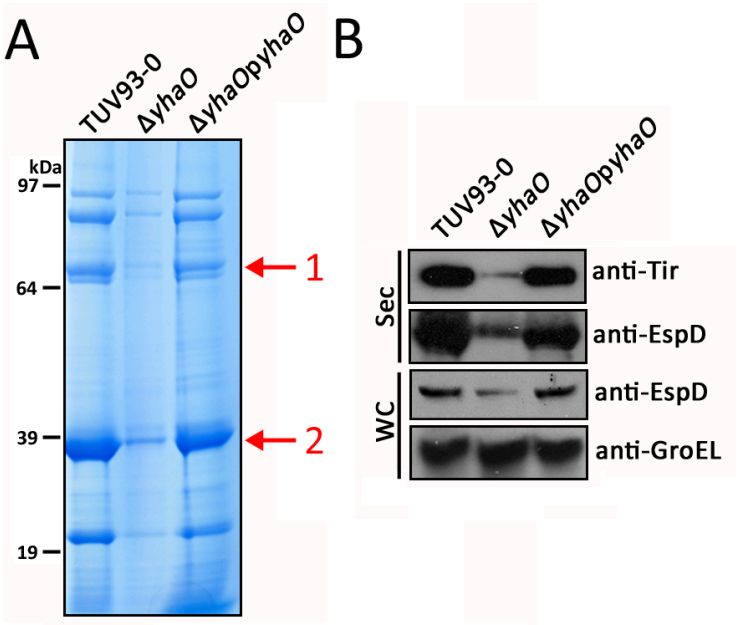

Figure 3-8 SDS-PAGE and western blot analysis of T3S in *ΔyhaO.* (A) SDS-PAGE analysis of the T3S profile from TUV93-0, *ΔyhaO* and *ΔyhaO*p*yhaO*. Protein bands corresponding to Tir (1) and EspD (2) are indicated by red arrows. Molecular weight markers are indicated on the left. (B) Western blot analysis of Tir, EspD and GroEL from secreted proteins (Sec) and whole cell lysate (WC) as indicated in panel A.

3.4.3 *YhaO is essential for normal binding to host cells*

As demonstrated in Figure 3-8, expression of YhaO is necessary for wild type levels of T3S under LEE inducing conditions in liquid culture. However, this experiment does not provide any information on the importance of YhaO in an infection setting. To test this, cell-adhesion assays were performed on HeLa cells to assess the ability of *ΔyhaO* to form A/E lesions comparable to the wild type TUV93-0 *in vitro*. A/E lesions are characterised by a condensed patch of host cell actin underneath the attached bacterium when infected cells are fixed, Phalloidin stained and analysed using immunofluorescence microscopy. This indicates the formation of a bacterial pedestal and thus tight attachment to the host cell (Figure 3-9A). TUV93-0 successfully colonised >80% of cells visualised in a count of more than 30 random fields of view. This was significantly ($p \leq 0.001$) more than *ΔyhaO* that colonised approximately 50% of cells visualised (Figure 3-9B). Of these colonised cells, the number of bacteria per host cell was ~19 for TUV93-0 and lower for *ΔyhaO* (~14), although this result was not significant. Most interestingly, the percentage of bacteria forming A/E lesions as indicated by actin condensation was significantly higher ($p \leq 0.01$) in TUV93-0 (~80%) as opposed to *ΔyhaO* (<40%). These results indicate that *ΔyhaO* is less capable of colonising host cells and that attached bacteria are significantly less likely to form A/E lesions associated with attachment via the T3SS. These data agree with findings described in section 3.4.2 and highlight the importance of YhaO for LEE expression and A/E lesion formation *in vitro*.

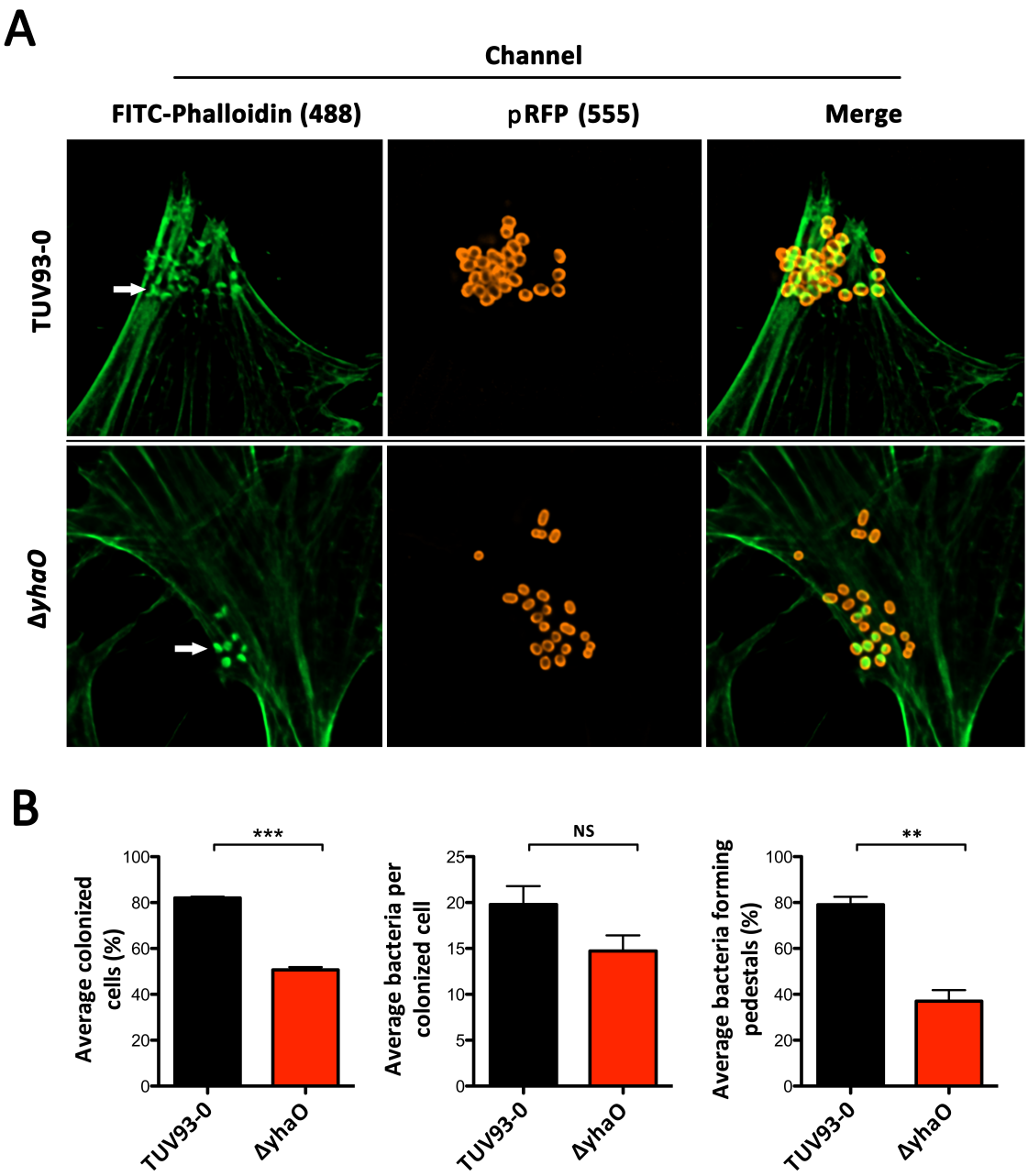

Figure 3-9 *in vitro* cell-adhesion assay of TUV93-0 and *ΔyhaO*. (A) Representative immunofluorescence images of HeLa cells infected with TUV93-0 or *ΔyhaO* captured by wide-field fluorescence microscopy. Cells were fixed and stained with Phalloidin, which fluoresces at 488 nm (green). Bacteria were transformed with the pRFP plasmid, which allowed constitutive expression of RFP and therefore visualisation in at 555 nm (red). White arrows indicate areas of condensed actin characteristic of A/E lesions. (B) Quantification of cell-adhesion data. TUV93-0 and *ΔyhaO* were compared for the average number of colonised host cells, the average number of bacteria per colonised cell and the average number of bacteria forming A/E lesions on host cells. NS, *** and ** denote no significance, $p \leq 0.001$ and $p \leq 0.01$ respectively calculated from three biological replicates (students T-test).

3.4.4 *Transcriptome analysis of ΔyhaO by RNA-seq*

Previous sections have confirmed a role for YhaO in regulation of LEE associated virulence although the mechanism of how such a signal may be transmitted to the genome was open to interpretation. The first step in investigating this was to identify genes belonging to the *yhaO* regulon by RNA-seq. Global transcriptome profiling is a powerful tool that can identify networks of genes that are co-regulated and therefore may uncover novel regulatory networks that have global effects on processes such as virulence. The enriched mRNA pool from three biological replicates of TUV93-0 and *ΔyhaO* under LEE inducing conditions (MEM-HEPES, 37°C, OD_{600} of 0.6) was isolated, reverse-transcribed and sequenced at the University of Edinburgh GenePool facility. Raw sequencing data was processed and analysed using CLC Genomics Workbench. Data was analysed by normalising total read counts between samples and calculating the genes differentially expressed according to an FDR of $p \leq 0.05$ by the Edge test implemented in CLC (Robinson *et al.,* 2010). For the purposes of identifying expression patterns without creating a bias for highly expressed or more variably expressed genes, analysis of the data was filtered by FDR alone and not by fold change, as this point is arbitrary.

RNA-seq analysis revealed 114 differentially expressed genes (DEGs) between the wild type and *ΔyhaO.* Genes were functionally annotated and were broadly assigned to functional groups of the Cluster of Orthologous Groups (NCBI) according to records in coliBase, RegulonDB and EcoCyc databases with a particular emphasis on genes related to virulence and regulation, as exemplified in Shah 2014. Functional patterns of DEGs were largely non-specific with 86 of these having known functions, 22 having putative functions and six being hypothetical. 29 genes were upregulated with functions in cellular metabolism (8), cell envelope biogenesis or transport (6), amino acid transport and metabolism (5), bacteriophage (2), nucleotide transport and metabolism (2), post-transcriptional regulation (1) and also hypothetical genes (2). Interestingly, *yhaM* was found to be upregulated 5.61 fold in the *ΔyhaO* mutant. Fold changes of significantly upregulated genes ranged from 1.38 to 29.06 fold. In contrast, 85 genes were significantly downregulated with fold changes ranging between -1.46 and -35.2 fold. Functions including cellular metabolism (4), cell envelope biogenesis or transport (4), bacteriophage (4), transcriptional (3) and post-transcriptional regulation (1), amino acid transport and metabolism (3) and also hypothetical genes (2) were identified in downregulated genes.

Strikingly, majority of the changes identified were involved directly in virulence. The entire LEE PAI was significantly downregulated, which is comprised of 41 ORFs and includes the master regulators *ler* and *grlA*. 20 NLEs were also significantly decreased as well as three putative adhesins. These data demonstrate a very specific response to deletion of *yhaO* with over half of the genes affected being involved in T3S virulence and suggest a key role for the expression of *yhaO* in virulence regulation. Differentially expressed genes not discussed in detail are summarised in Table 8-1.

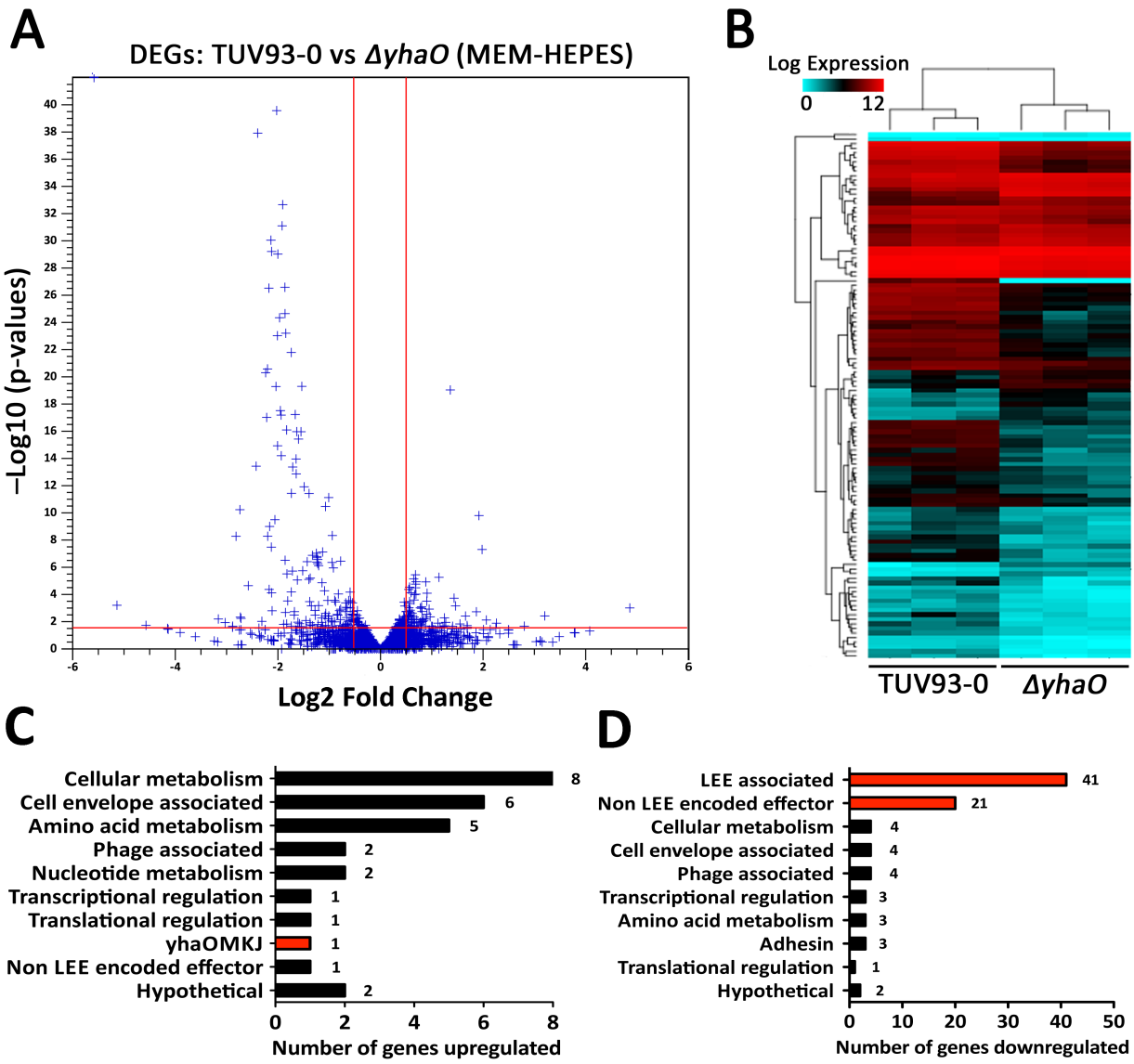

Figure 3-10 RNA-seq analysis of *ΔyhaO*. Overview of the *ΔyhaO* transcriptome. (A) Volcano plot identifying DEGs. The red lines on the x and y axes represent a $p \leq 0.05$ and minimum significant fold change respectively. (B) Heatmap of the 114 DEGs. Hierarchical clustering of normalised RNA-seq data from three biological replicates of TUV93-0 and *ΔyhaO.* (C) Functional grouping of upregulated genes in the *ΔyhaO* dataset. (D) Functional grouping of downregulated genes in the *ΔyhaO* dataset. Key genes relating to *yhaOMKJ*, the LEE and NLEs are highlighted in red.

3.4.5 *YhaO and the LEE/NLE regulon*

As mentioned above, the entire LEE PAI was significantly downregulated ($p \leq 0.001$) in the *ΔyhaO* mutant (Figure 3-11). The LEE encodes five major operons that are all under the control of the master regulator Ler (Haack *et al.,* 2003, Mellies *et al.,* 1999, Sánchez-SanMartín *et al.,* 2001). This suggests that *yhaO* may be a key member of the LEE regulon and its expression may be acting through Ler by way of an unknown regulator.

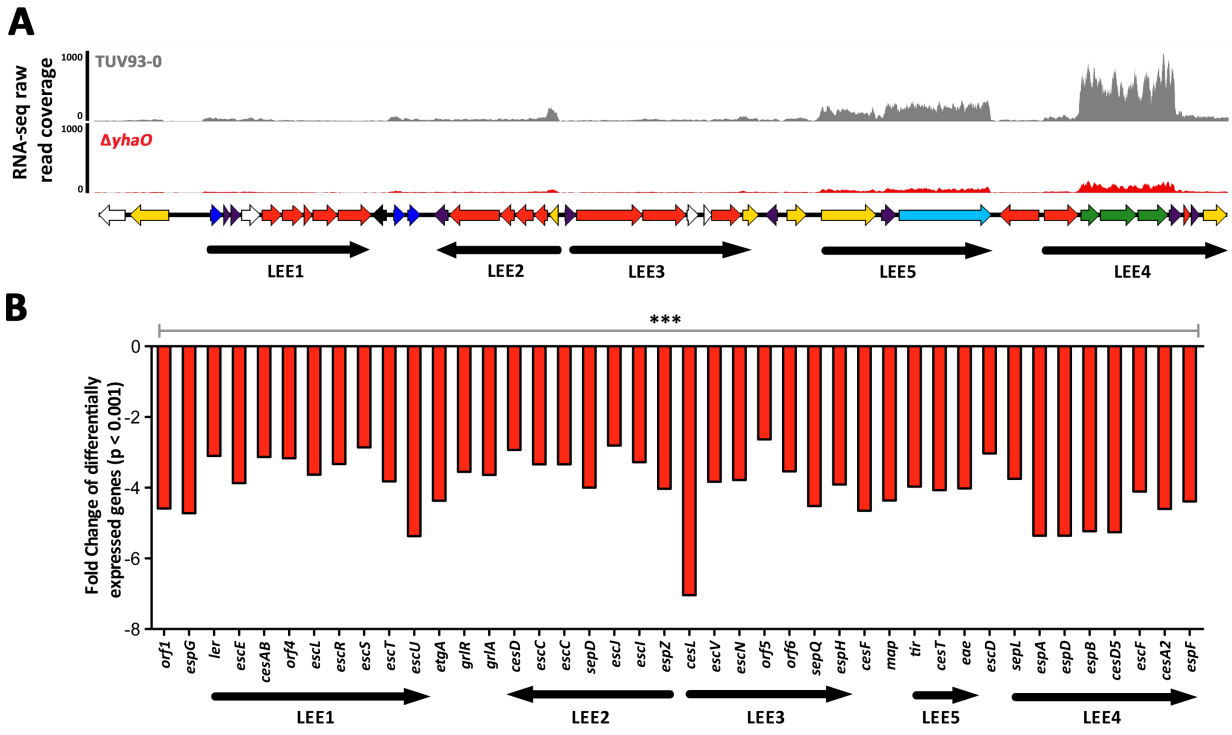

Figure 3-11 Downregulation of the LEE in the *ΔyhaO* mutant. (A) Visualisation of raw reads from RNA-seq analysis of TUV93-0 (grey) and *ΔyhaO* (red) mapped to the LEE region of the EDL933 genome. LEE genes are colour coded as per Figure 1-3. The coverage graph is representative of a single biological replicate and was generated in EasyFig (Sullivan *et al.,* 2011). (B) RNA-seq data indicating the fold change of mean counts mapped per transcript for each ORF encoded on the LEE as calculated by the Edge test from three biological replicates. *** denotes $p \leq 0.001$ for every ORF.

To investigate whether this repression of the LEE was via direct inhibition of the Ler protein or suggestive of transcriptional repression, a plasmid carrying an inducible *ler* gene under the control of the arabinose promoter (pVS45) was introduced into the *ΔyhaO* background and screened for effects of T3S. pVS45 has been demonstrated previously to produce a functional Ler that is constitutively expressed from this plasmid

(Sperandio *et al.,* 2000). Addition of the plasmid restored T3S to levels of the wild type TUV93-0 indicating that the negative effects on the LEE in *ΔyhaO* are likely as a result of direct transcriptional inhibition of the LEE by an unknown factor and not through inhibition of Ler itself (Figure 3-12).

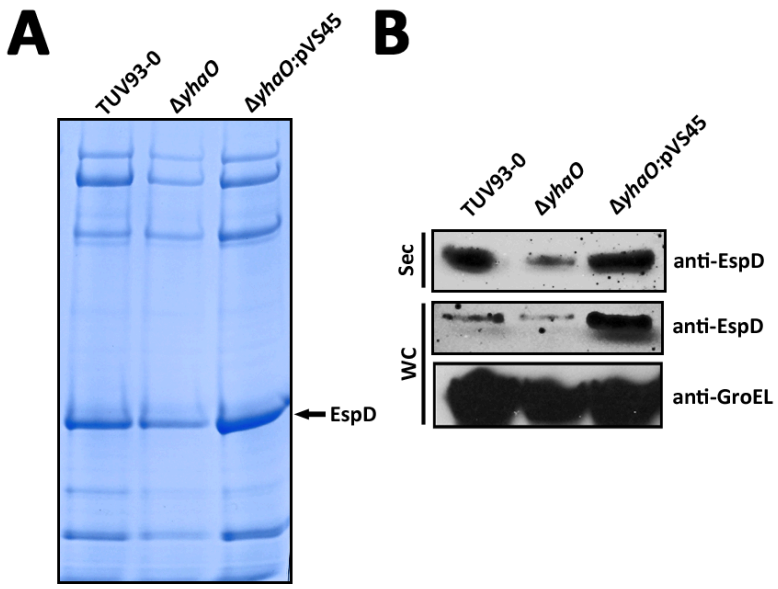

Figure 3-12 Repression of T3S in *ΔyhaO* is not through inhibition of Ler. (A) Secreted protein profile of TUV93-0, *ΔyhaO* and *ΔyhaO:*pVS45. Location of EspD on the SDS-PAGE is indicated by an arrow. (B) Western blot analysis of the corresponding secreted (Sec) and whole cell (WC) fractions using anti-EspD antibody. Detection of GroEL in the WC fraction was used as a loading control. Figure is representative of three biological replicates.

In addition to transcriptional inactivation of the LEE, the *ΔyhaO* RNA-seq data shows downregulation of 20 NLEs (Figure 3-13). NLEs are encoded throughout the genome on cryptic prophage elements and are subject to relatively unexplored regulatory control (Tobe *et al.,* 2006). For instance, Deng *et al.* showed increased secretion of seven NLEs in a *grlA* expression strain (Deng *et al.,* 2004). NleA and NleH2 have been shown to be directly under the control of Ler whereas NleH1 and NleB2 are not (García-Angulo *et al.,* 2012, Holmes *et al.,* 2012, Roe *et al.,* 2007). The latter study also identified an essential 13 bp NLE regulatory inverted repeat (NRIR) sequence upstream of a subset of NLEs. This NRIR sequence indicated a Ler independent mechanism of NLE regulation. NLEs identified in the *ΔyhaO* dataset (summarised in Table 3-1) comprised both Ler and NRIR dependent effectors, suggesting that the *yhaO* regulon consisted of an even more universal regulator

of virulence that is as yet unidentified. Surprisingly, one predicted NLE (EspR3) was significantly upregulated however this effector has not been characterised. Its coding sequence is not located on a cryptic prophage element associated with virulence but rather it is located directly adjacent to OI-77, which carries flagella-associated genes (Perna *et al.,* 2001). Its function is unknown. NLE annotations were obtained from the Virulence Factor Database (Chen *et al.,* 2012).

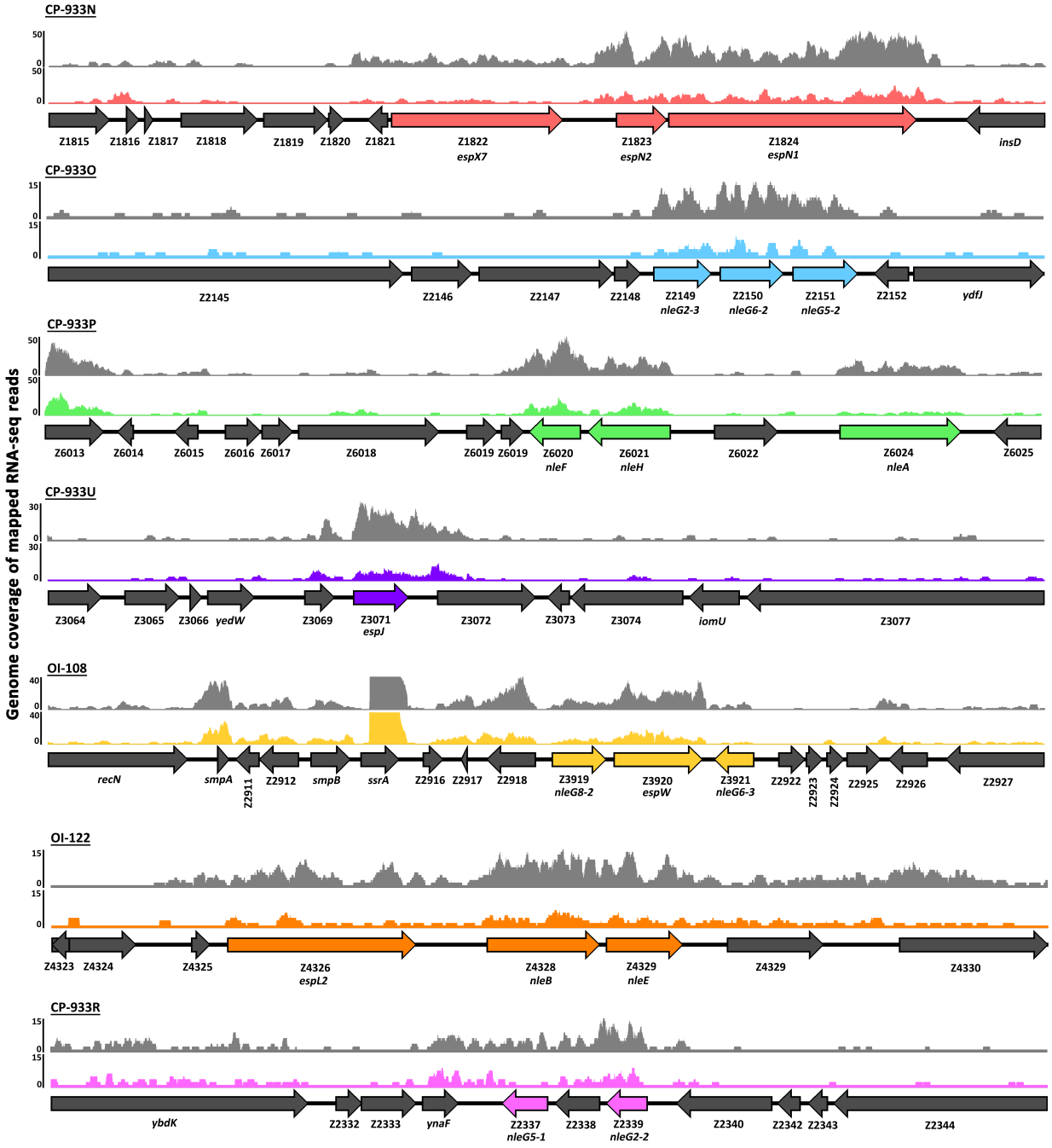

Figure 3-13 Multiple NLE expression patterns in the *ΔyhaO* mutant. Visualisation of raw reads from RNA-seq analysis of TUV93-0 (grey) and *ΔyhaO* (colour coded for each region) mapped to the various cryptic prophage and O-islands of the EDL933 genome. Scale bars are indicated on the left hand side. The coverage graphs are representative of a single biological replicate and were generated in EasyFig (Sullivan *et al.,* 2011).

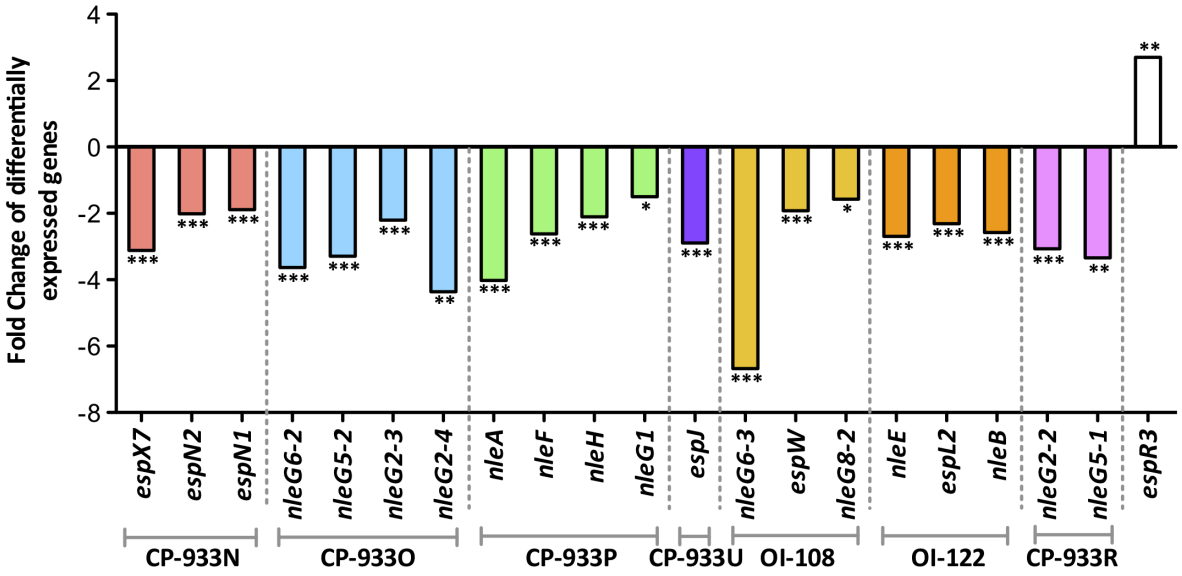

Figure 3-14 Differential expression of multiple NLEs in the *yhaO* mutant. RNA-seq data indicating the fold change of mean counts mapped per transcript for each NLE ORF identified in the *ΔyhaO* dataset. Bars are colour coded according to their genomic island location, which are labelled below and separated by grey dotted lines. DEG analysis was performed by the Edge test from three biological replicates. ***, ** and * denote $p \leq 0.001$, $p \leq 0.01$ and $p \leq 0.05$ respectively.

Table 3-1 Summary of NLEs differentially expressed in *ΔyhaO* as identified by RNA-seq

Z number	Name	Fold Change	P value	O-island	Phage designation	NRIR regulated
Z3071	*espJ*	-2.89	1.11E-17	OI-79	CP-933U	no
Z6024	*nleA*	-4.02	1.80E-13	OI-71	CP-933P	no
Z1822	*espX7*	-3.12	1.83E-11	OI-50	CP-933N	yes
Z6020	*nleF*	-2.62	4.72E-10	OI-71	CP-933P	no
Z1824	*espN2*	-2.01	9.28E-10	OI-50	CP-933N	no
Z6021	*nleH1*	-2.1	4.10E-09	OI-71	CP-933P	no
Z3921	*nleG6-3*	-6.68	6.88E-09	OI-108	_	yes
Z3920	*espW*	-1.92	5.06E-07	OI-108	_	no
Z2150	*nleG6-2*	-3.63	2.72E-05	OI-57	CP-933O	yes
Z4329	*nleE*	-2.69	3.36E-05	OI-122	_	no
Z4326	*espL2*	-2.31	6.14E-05	OI-122	_	no
Z1823	*espN1*	-1.89	8.90E-05	OI-50	CP-933N	no
Z2151	*nleG5-2*	-3.29	1.47E-04	OI-57	CP-933O	yes
Z4328	*nleB*	-2.58	4.59E-04	OI-122	_	no
Z2339	*nleG2-2*	-3.07	6.08E-04	_	CP-933R	yes
Z2149	*nleG2-3*	-2.2	9.34E-04	OI-57	CP-933O	yes
Z2337	*nleG5-1*	-3.34	4.04E-03	_	CP-933R	yes
Z2075	*nleG2-4*	-4.36	4.51E-03	OI-57	CP-933O	yes
Z3023	*espR3*	2.7	0.01	_	_	no
Z3919	*nleG8-2*	-1.57	0.04	OI-108	_	no
Z6010	*nleG1*	-1.5	0.04	OI-71	CP-933P	no

3.4.6 *Validation of RNA-seq by qRT-PCR*

The RNA-seq data revealed a number of interesting changes in global gene expression in the *ΔyhaO* mutant. Most striking were the effects on the LEE and multiple NLE encoding phages. Thus validation of RNA-seq was carried out on selected LEE encoded ORFs and NLEs. Primers were designed for *ler, espA, espD* and *tir* encoded on the LEE and the NLEs *nleA* and *nleG1.* Total RNA was extracted from TUV93-0 and *ΔyhaO* cultured in MEM-HEPES to an OD_{600} of 0.6. RNA was reverse transcribed and analysed by qRT-PCR in a one-step reaction. Analysis of three biological replicates confirmed the results obtained from the RNA-seq data.

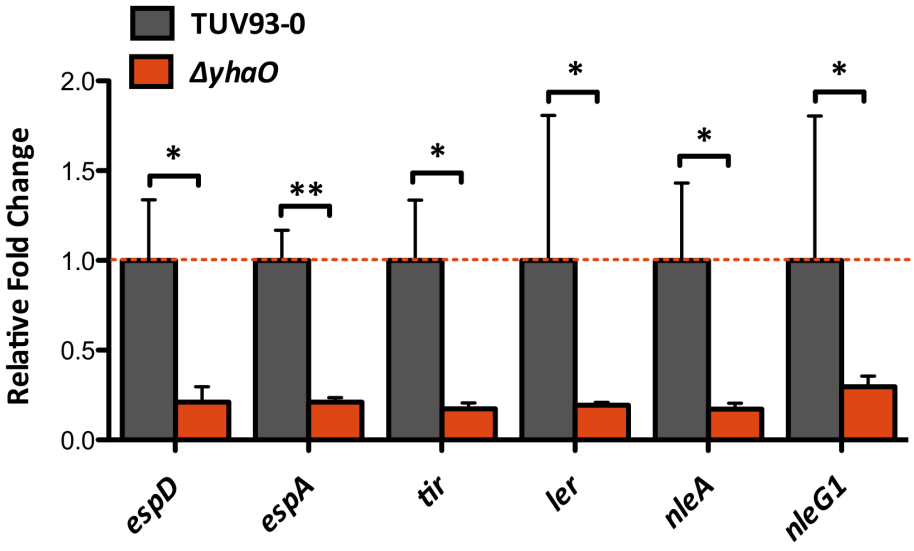

Figure 3-15 qRT-PCR validation of RNA-seq data. The expression of LEE encoded genes *espD, espA, ler* and *tir* in the *ΔyhaO* background (red) was determined relative to TUV93-0 (grey). Similarly, the relative expression levels of *nleA* and *nleG1* were determined. The red dotted line indicates the TUV93-0 expression level of the transcript. * and ** denote $p \leq 0.05$ and $p \leq 0.01$ respectively as calculated from three biological replicates.

3.5 Is YhaO controlled by YhaJ?

Regulation of the LEE and NLEs is a complex process. The LEE is induced when cultured in MEM-HEPES or DMEM. These conditions also induce expression of various NLEs. The expression levels of individual genes however are not uniform. Despite this, common regulators may exist that control the expression levels of the LEE and associated NLEs in a temporal manner. Due to the striking effects of the *ΔyhaO* deletion on the TUV93-0 transcriptome we hypothesised that an unknown regulator may be controlling LEE and also NLE expression. It was postulated that YhaJ, coded in close proximity to *yhaO* in a manner genetically similar to *dsdCXA,* might regulate both *yhaO* expression and virulence. YhaO is a D-serine transporter and YhaJ is a putative LTTR, which are proteins involved in transcriptional regulation. YhaK, coded divergently from *yhaJ,* has been implied to act as a cofactor for YhaJ. Interestingly, *yhaK* was found to be upregulated when grown in human urine, an environment rich in D-serine. Thus, one can envisage a system whereby D-serine is transported or sensed by YhaO and a signal is translated to YhaJ/YhaK, which is capable of binding DNA and regulating transcription. In a similar fashion to *yhaO*, *yhaJ* was deleted and investigated the effects this had on O157:H7 normal function and virulence.

3.5.1 *Analysis of growth and motility in the ΔyhaJ mutant*

The *yhaJ* mutant was cultured in MEM-HEPES and DMEM tissue culture media at 37°C and compared to growth of TUV93-0 under the same conditions. As with *ΔyhaO,* the growth behaviour of *ΔyhaJ* was virtually identical to that of the wild type. Similarly, motility was analysed also and no differences were observed.

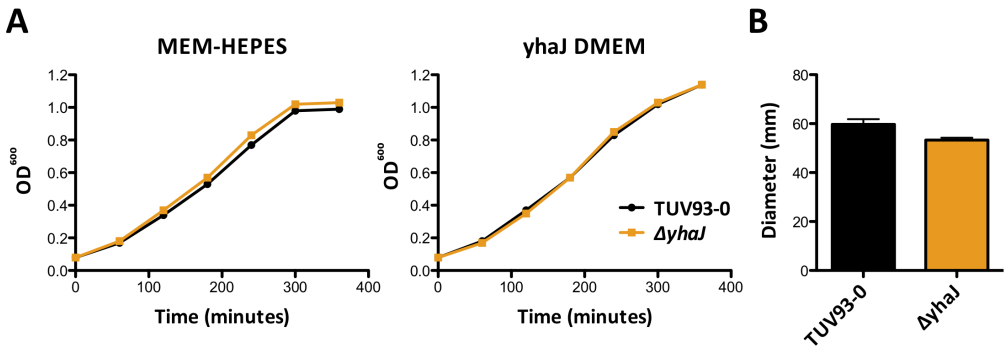

Figure 3-16 Deletion of *yhaJ* does not affect growth or motility. Growth (A) and motility (B) of *ΔyhaJ* were assessed in comparison to TUV93-0. Growth curves are representative graphs of triplicate experiments.

3.5.2 *yhaO is transcriptionally controlled by YhaJ*

The working hypothesis thus far was that YhaO has control over the LEE by feeding signals from the environment into the LEE regulatory network. Being a membrane protein however, these signals must be transmitted by a regulator that is capable of affecting gene expression directly. To assess if YhaJ was associated with YhaO, the effects of deleting *yhaJ* on the expression of *yhaO* were first determined. Interestingly, expression of *yhaO* was reduced > 11 fold in the *ΔyhaJ* strain as analysed by qRT-PCR. This result indicates that YhaJ is required for *yhaO* expression and suggests that the two genes share roles in the same processes. Thus, it is logical to hypothesise that signals transported or perceived by YhaO might be transferred to the genome by the LTTR YhaJ in a similar way to the *dsdCXA* system of UPEC. Furthermore, the *ΔyhaO* RNA-seq data indicated a > 5 fold increase in expression of *yhaM*, which was shown in Figure 3-2 to be co-transcribed with *yhaO.* This over-compensation at the *yhaOM* promoter in the *ΔyhaO* strain may account for the negative effects exerted on the LEE if this regulation occurs through YhaJ.

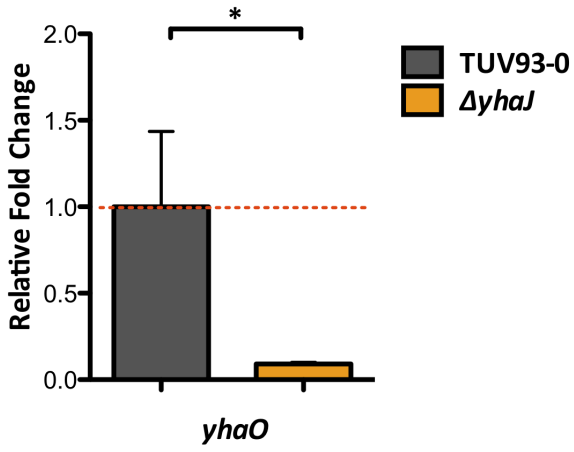

Figure 3-17 YhaJ controls *yhaO* expression. qRT-PCR analysis of *yhaO* expression levels in the *ΔyhaJ* background (orange) relative to TUV93-0 (grey). The red dotted line indicates baseline expression levels of *yhaO* in the wild type. * denotes $p \leq 0.05$ as calculated from three biological replicates.

3.5.3 *Deletion of yhaJ negatively affects LEE associated type 3 secretion*

As discussed in section 3.4.2, deletion of *yhaO* negatively affects the secreted protein profile in comparison to the wild type TUV93-0. As *yhaO* is under the control of YhaJ, the impact of a deletion in *yhaJ* on T3S was assessed by secreted protein profiling and

western blot analysis. TUV93-0 and *ΔyhaJ* were cultured in MEM-HEPES at 37°C to an OD_{600} of 0.7 and the secreted proteins were SDS-PAGE analysed. *ΔyhaJ* displayed a reduced secretion profile of known LEE encoded effectors. Furthermore, western blot analysis of EspD confirmed this repression in both the secreted and whole cell fractions. This experiment confirmed that EspD expression in *ΔyhaJ* was indeed reduced when compared with the wild type. As a loading control, anti-GroEL antibody was used and confirmed normal housekeeping expression in the whole cell lysate. The negative effects on T3S observed in the *ΔyhaJ* mutant could also be reversed by complementation with p*yhaJ* (Figure 3-18). These data confirm a role for YhaJ in regulation of the T3SS and associated secreted effectors.

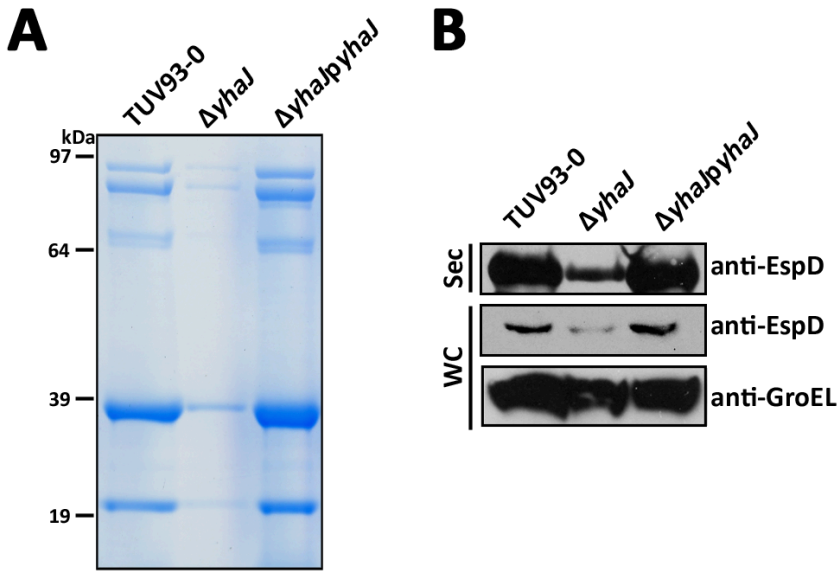

Figure 3-18 Deletion of *yhaJ* negatively effects T3S. (A) SDS-PAGE analysis of the T3S profile from TUV93-0, *ΔyhaJ* and *ΔyhaJpyhaJ*. Molecular weight markers are indicated on the left. (B) Western blot analysis of EspD and GroEL from secreted proteins (Sec) and whole cell lysate (WC) corresponding with lanes indicated in panel A.

3.5.4 *qRT-PCR validation of T3S repression in ΔyhaJ*

YhaJ was found to be necessary for secretion of effectors into the growth medium when cultured under T3S permissive conditions. To validate if these effects were occurring at the transcriptional level, similarly to *ΔyhaO,* qRT-PCR analysis was carried out on the LEE encoded genes *ler, espD* and *tir* as well as *nleA* and *nleG1.* As expected, deletion of *yhaJ* significantly downregulated the LEE encoded genes as well as both NLEs (Figure 3-19). These data indicate transcriptional repression of virulence in the *ΔyhaJ* mutant similarly

to *ΔyhaO* suggesting that the transcriptional effects observed in the *ΔyhaO* may be exerted through YhaJ.

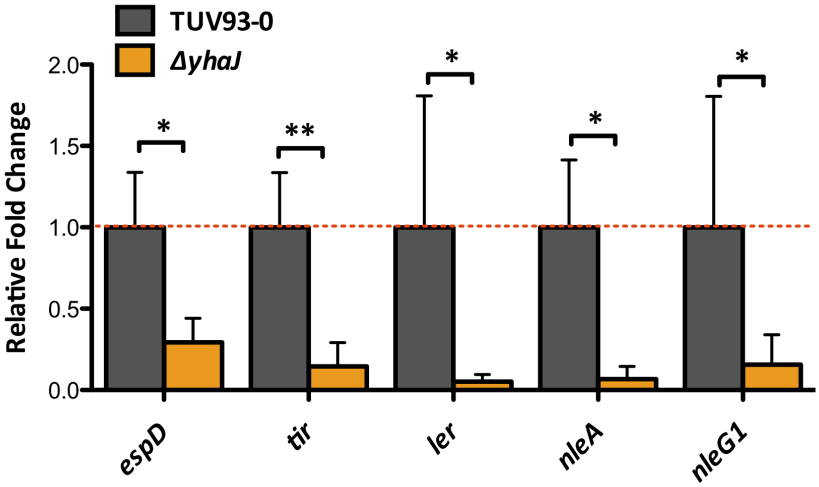

Figure 3-19 qRT-PCR analysis of LEE and NLE genes in *ΔyhaJ.* The expression of LEE encoded genes *espD, ler* and *tir* in the *ΔyhaJ* background (orange) was determined relative to TUV93-0 (grey). Similarly, the relative expression levels of *nleA* and *nleG1* were determined. The red dotted line indicates the TUV93-0 wild type expression level of the transcript. * and ** denote $p \leq 0.05$ and $p \leq 0.01$ respectively as calculated from three biological replicates.

3.5.5 *YhaJ is essential for normal binding to host cells*

T3S phenotyping and qRT-PCR analysis suggest that YhaJ would be required for attachment to host cells. To test this cell-adhesion assays were performed under the same conditions as *ΔyhaO.* The *ΔyhaJ* mutant was less capable of forming A/E lesions than TUV93-0 indicated by less prominently and frequently formed patches of condensed actin underneath bacterial attachment sites (Figure 3-20A). As previously, TUV93-0 successfully colonised >80% of cells visualised in a count of more than 30 random fields of view whereas *ΔyhaJ* colonised ~60% of cells ($p \leq 0.01$). As seen with *ΔyhaO,* the number of *ΔyhaJ* bacteria per host cell was ~15 on average, lower than the ~19 seen for TUV93-0 although these data were not significant. *ΔyhaJ* was significantly less able to form A/E lesions on host cells, with less than 50% of the attached bacteria forming patches of condensed actin ($p \leq 0.01$) (Figure 3-20B). These results indicate that *ΔyhaJ* is less capable of colonising host cells and that attached bacteria are significantly less likely to form A/E lesions, similarly to the *ΔyhaO* mutant.

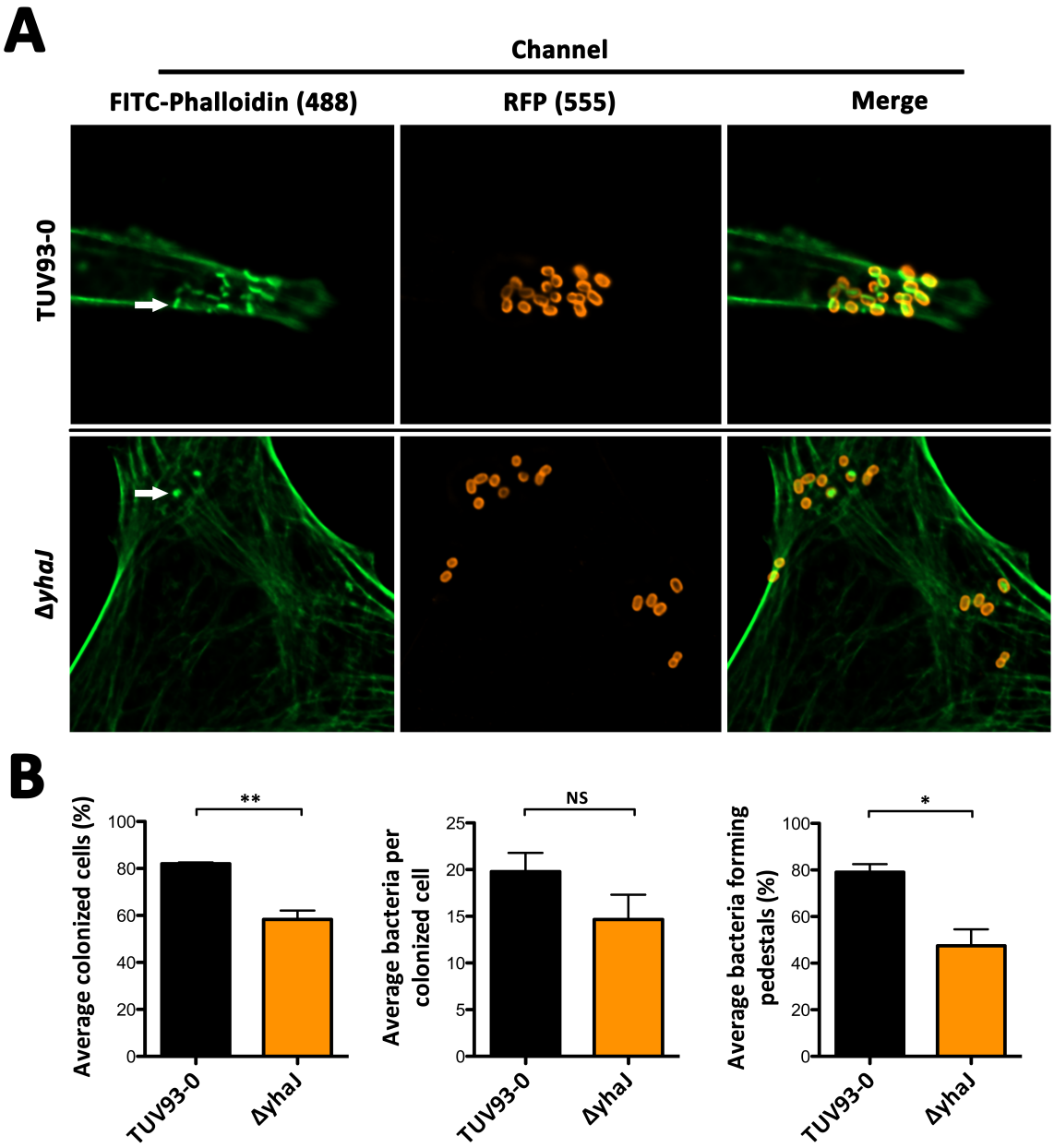

Figure 3-20 *in vitro* cell adhesion assay of TUV93-0 and *yhaJ*. (A) Representative immunofluorescence images of HeLa cells infected with TUV93-0 or *ΔyhaJ* captured by wide-field fluorescence microscopy. Cells were fixed and stained with Phalloidin, which fluoresces at 488 nm (green). Bacteria were transformed with the pRFP plasmid, which allowed constitutive expression of RFP and therefore visualisation in at 555 nm (red). White arrows indicate areas of condensed actin characteristic of A/E lesions. (B) Quantification of cell-adhesion data. TUV93-0 was compared to *ΔyhaJ* for the average number of colonised host cells, the average number of bacteria per colonised cell and the average number of bacteria forming A/E lesions on host cells. NS, ** and * denote no significance, $p \leq 0.01$ and $p \leq 0.05$ respectively calculated from three biological replicates (students T-test).

3.6 YhaJ – a novel transcriptional regulator of virulence

Phenotypic studies of *yhaJ* and *yhaO* deletion mutants show overlapping characteristics between the two strains – normal growth and motility, transcriptional down regulation of the LEE and numerous NLEs, and an impaired ability to colonise host cells via A/E lesions. *yhaO* is transcriptionally under the control of YhaJ leading to the hypothesis that YhaJ may regulate other genomic sites through sensing of signals via YhaO. LTTRs are a family of known transcriptional activators and repressors regulating processes such as virulence, motility and diverse branches of cellular metabolism (Maddocks & Oyston 2008). Its clear that YhaJ plays a role in regulating virulence through the complicated network of LEE and NLE control systems but whether this regulation is direct or indirect through additional factors was not clear. To address this important point, YhaJ was cloned into an overexpression vector, purified as a His-tagged recombinant protein and tested for its ability to bind the promoter regions of genes in question.

3.6.1 *Cloning, overexpression and purification of YhaJ*

The DNA sequence for *yhaJ* from EDL933 relative to its translational start and stop codons was amplified to include *Bam*HI and *Hind*III restriction sites in frame at its 5' and 3' ends respectively using primers specified in Table 2-16. The product was cloned into a StrataClone storage vector using a StrataClone PCR cloning kit according to the manufacturers guidelines and sub-cloned into pET28b. Clones were transformed into TOP10 cells and selected for growth on kanamycin. Positive colonies were confirmed for the presence of pET28b-YhaJ by Miniprep and restriction digest using *Bam*HI and *Hind*III followed by analysis on agarose gel electrophoresis, which yielded an 897 bp band relative to *yhaJ*. The pET28b-YhaJ construct was then re-transformed into BL21 DE3 cells for protein overexpression. A 100 ml small scale induction of YhaJ was first obtained and whole cell lysate was analysed for YhaJ overexpression by western blotting using anti-His antibody, which confirmed overexpression of YhaJ by the presence of two oligomeric states at ~33 kDa (the predicted molecular weight of YhaJ) and ~66 kDa (Figure 3-21A). Cells were subsequently cultured in five litre batches of LB at 37°C and 200 rpm, induced with 1 mM IPTG at an OD_{600} of 0.6 and incubated overnight at 16°C. Cell pellets were harvested by centrifugation, resuspended in His-column buffer and lysed by French press

or sonication. The resulting supernatant was used for purification of recombinant YhaJ by IMAC and SEC (Figure 3-21B). YhaJ purified in two forms (monomeric and dimeric) as confirmed by SDS-PAGE (Figure 3-21C). These bands were confirmed as being YhaJ by tandem mass spectrometry (MS-MS) performed by Dr. Richard Burchmore of University of Glasgow Polyomics. LTTRs generally function as protein dimers that bind DNA so the presence of stable multimeric complexes in solution is not a surprising result (Maddocks & Oyston 2008). Solubility is often an issue associated with the study of DNA binding proteins. YhaJ remained stable post SEC at room temperature in a buffer composition of 50 mM Tris, 150 mM NaCl and 150 mM imidazole but this solution was extremely temperature sensitive. Rhys Grinter, a PhD student from the lab of Dr Dan Walker in the University of Glasgow, assisted with the purification and is now currently optimising stabilisation conditions for YhaJ to attempt structural characterisation studies.

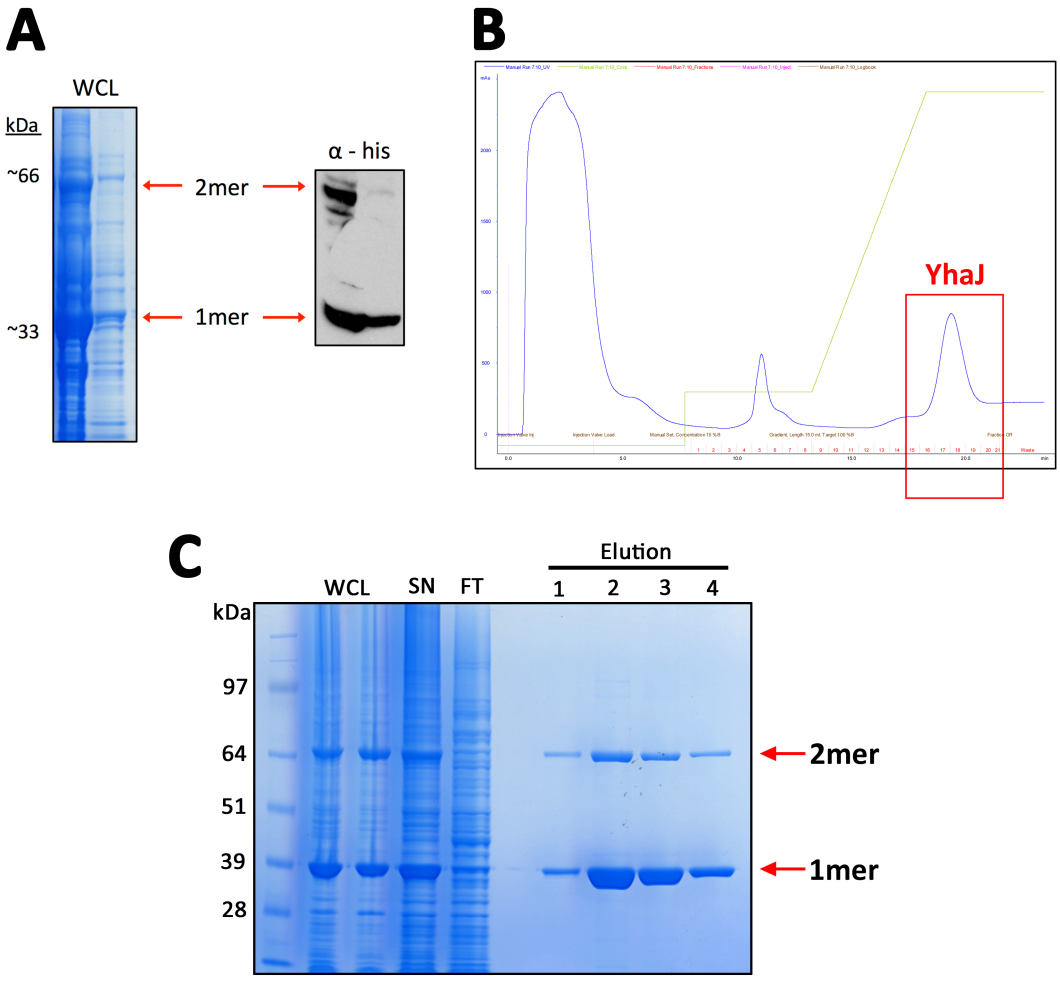

Figure 3-21 Overexpression and purification of YhaJ. (A) Small scale expression test of pET28-YhaJ. Cell lysate indicated large bands at ~33 and ~66 kDa and this was confirmed to be recombinant YhaJ by western blotting using anti-His antibody. (B) AKTAprime trace of YhaJ purification by IMAC. The red box indicates the elution peak that yielded YhaJ. (C) SDS-PAGE analysis of purified YhaJ confirms two oligomeric states. Cell lysate (WCL), supernatant (SN) and flow through (FT) samples are indicated also.

3.6.2 *Purified YhaJ binds multiple promoter region targets*

LTTRs are the most abundant family of transcriptional regulators in bacteria and they can function as transcriptional activators and repressors on a global scale. The 'classical' model of LTTRs is that they function as autorepressors of their own transcription in the absence of an inducing signal. When this signal is presented they activate their own transcription as well as other genes they control, usually organised in operons, thus forming a positive feedback loop. However, LTTRs are known to have diverse functions and can act as activators or repressors of their own expression and the expression of genes under their control (Maddocks & Oyston 2008). Typically, LTTRs bind two independent DNA binding sites upstream of the promoter they are controlling as dimers, the regulatory binding site (RBS) and the activation binding site (ABS), the latter of which is only occupied in the presence of an inducer. In the presence of the inducer signal, LTTRs undergo a conformational change and bend the DNA to form a functional tetramer, which activates transcription (Schell 1993). The RBS typically contains a T-N_{11}-A sequence motif called an 'LTTR box' although this sequence can vary in length and composition and is often subject to much degeneration, even in promoters of the same regulon (Maddocks & Oyston 2008).

The ability of YhaJ to bind specific sequences of DNA was assessed by EMSA using PCR fragments of its own promoter region. Three regions were amplified: +12 to -291, -270 to -569 and -504 to -797 in respect to the ATG start codon. DNA fragments were DIG labelled, mixed with increasing concentrations of purified YhaJ in a binding reaction and run on a DNA retardation gel as per the manufacturers guidelines (Roche). Protein-DNA complexes were electro-transferred to a nylon membrane and UV crosslinked before detection of DIG labelled probes using a specific anti-DIG antibody. Anti-DIG is alkaline phosphatase tagged and therefore allows chemi-luminescent detection of DNA by addition of the substrate CSPD and exposure to x-ray film. EMSA analysis of these regions confirmed that YhaJ bound the +12 to -291 region upstream of its coding sequence by the presence of a DNA-protein complex shift suggesting that YhaJ bound this region of DNA specifically (Figure 3-22A). Analysis of this 303 bp sequence revealed 15 putative LTTR boxes according to the T-N_{11}-A motif. To attempt to identify a more specific binding site, PCR fragments were generated for three subregions of this 303 bp fragment: +12 to -84, -85 to -208 and -182 to -287. EMSA analysis revealed selective binding to the -182 to -287

fragment by the presence of a DNA-protein complex shift (Figure 3-22B). This region contained three putative LTTR boxes, two of which overlapped therefore revealing three putative YhaJ boxes (Figure 3-22C). Attempts to observe specific binding to the individual sites of this 105 bp fragment were made using synthesised oligonucleotide probes but were inconclusive.

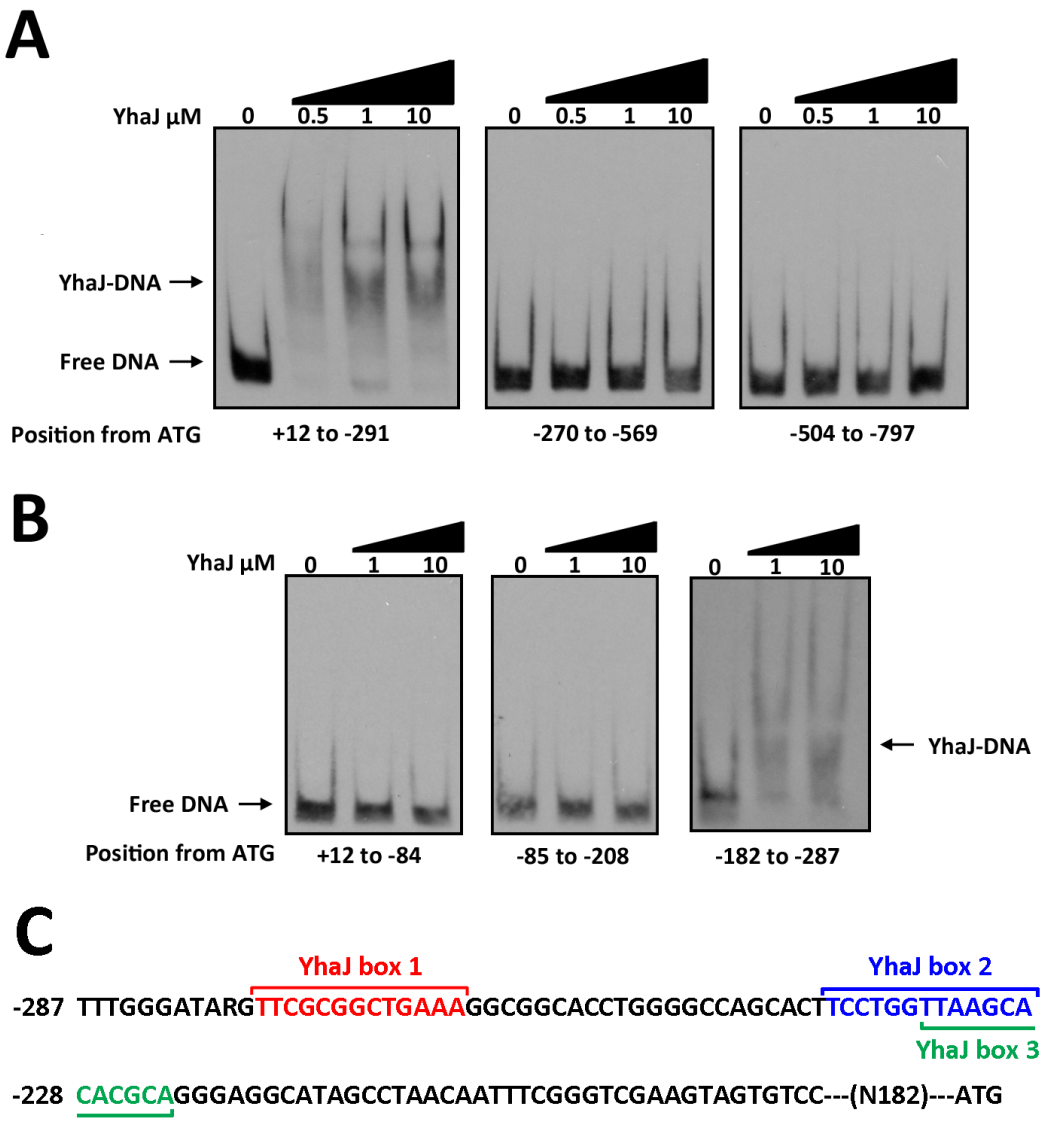

Figure 3-22 YhaJ binds upstream of its own promoter. (A) EMSA assays of three DNA fragments covering the +12 to -797 region with respect to the ATG. Increasing concentrations of YhaJ are indicated by black triangles above. DNA retardation was observed in the +12 to -291 fragment with a YhaJ-DNA complex being formed. Free unbound DNA is labelled also. (B) EMSA assays of three DNA fragments covering the +12 to -287. DNA retardation was observed in the -182 to -287 fragment with a YhaJ-DNA complex being observed. EMSA experiments were performed in triplicate with similar results obtained each time. (C) Sequence of the -182 to -287 fragment with the three putative YhaJ binding boxes indicated in red, blue and green respectively. N182 represents the remaining sequence length between this fragment and the ATG.

The capability of YhaJ to bind its own promoter suggests that YhaJ autoregulates its own expression independent of an inducing signal. This is likely to maintain constitutive expression of the system under various conditions. The ability of YhaJ to bind the promoter regions of *ler* and *yhaO* was therefore tested. The *ΔyhaJ* mutant displays reduced expression of the LEE and this was hypothesised to occur through transcriptional control by YhaJ. Indeed, YhaJ was able to bind a 451 bp fragment containing the LEE1 regulatory region (Figure 3-23A). Alignment of this region with each of the three putative YhaJ boxes revealed the presence of degenerate forms of each (box 1, 61% conserved; box 2, 69% conserved; box 3, 61% conserved) (Figure 3-23B). Box 1 lies at position -91 and position +27 relative to the LEE1 P2 and P1 transcriptional start sites (Russell *et al.,* 2007). Box 2 overlaps with the P2 -10 element, however the T-N_{11}-A motif has not been maintained. Finally, box 3 overlaps with the *ler* ATG translational start site (Figure 3-23C). LTTRs usually bind the RBS at -35 to +20 and the ABS at -40 to -20, however distant binding sites both up and downstream of promoters have been described (Maddocks & Oyston 2008). LTTRs often bind dyadic (inverted repeat) regions within the T-N_{11}-A motif. Each YhaJ box contains imperfect dyadic characteristics although box 2 and box 3 within the LEE1 promoter regions have conserved only one of these residues. These data suggest box 1 is likely to be the primary YhaJ binding site but the possibility of multiple binding sites cannot be ruled out as seen in other studies (Chugani *et al.,* 1998). Furthermore, the positions of box 2 and box 3 relative to the P2 promoter suggest they would act as transcriptional repressors blocking P2 promoter activity whereas the upstream position of box 1 suggests a more likely position for transcriptional activation via LEE1 P2 according to previous studies on the P2 promoter activity (Islam *et al.,* 2011a, Sharp & Sperandio 2007). This is the preferred hypothesis due to the fact that *ΔyhaJ* displays decreased T3S but the precise role of YhaJ as a direct activator or repressor of the LEE1 promoters has not been investigated here.

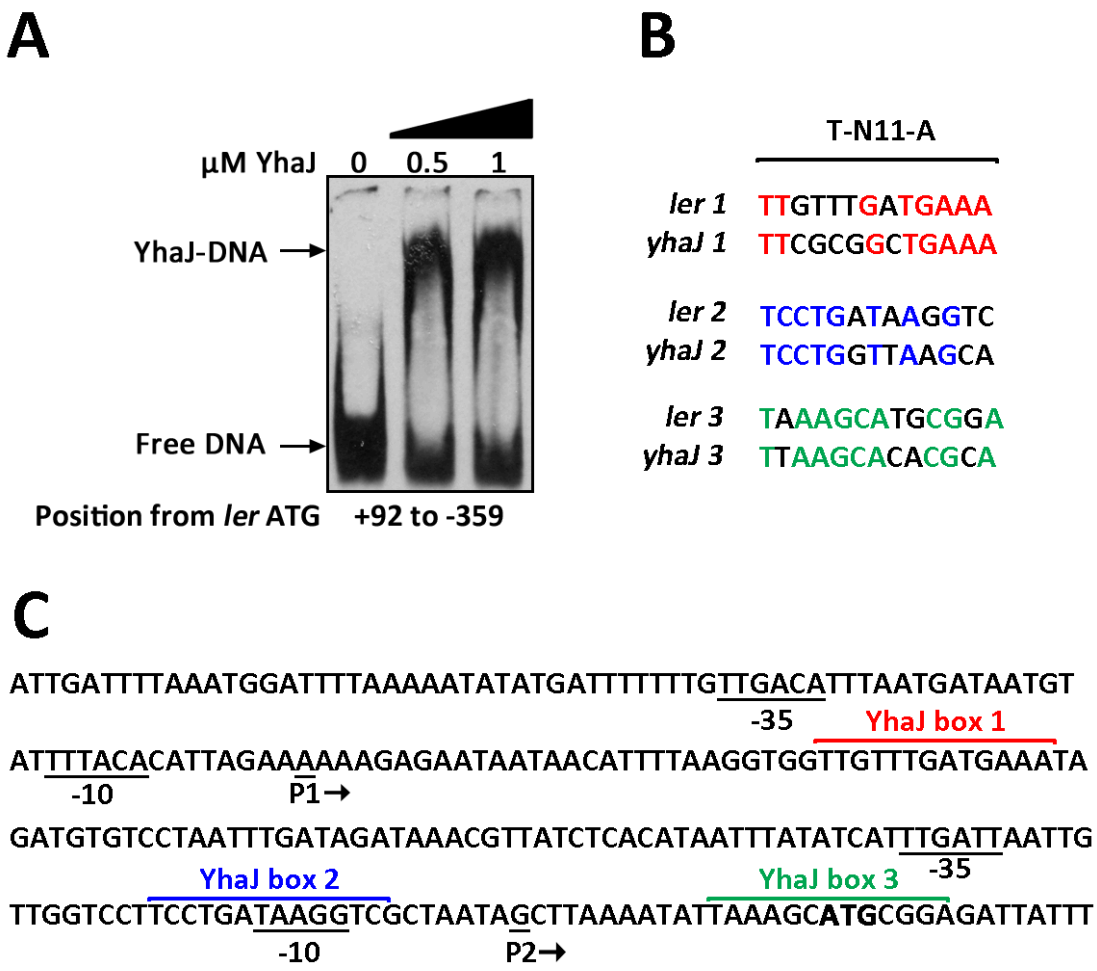

Figure 3-23 YhaJ binds the LEE1 promoter region. (A) EMSA of YhaJ binding a 451 bp fragment including the LEE1 promoter region. (B) Sequence conservation of the three putative YhaJ boxes in the LEE1 fragment. Conserved residues are colour coded by which YhaJ box they correspond to. (C) LEE1 promoter region sequence. The ATG start codon is in bold. The P1 and P2 transcriptional start sites as well as corresponding -10 and -35 promoter elements are underlined and labelled below the sequence. YhaJ boxes 1 to 3 are labelled above the corresponding sequence and colour coded red, blue and green respectively. The LEE1 promoter details were obtained from Sharp and Sperandio 2007.

The *ΔyhaO* RNA-seq data shows a significant decrease not only on LEE expression but also on expression of multiple NLE genes (Figure 3-10). This phenotype was similar in *ΔyhaJ* as analysed by qRT-PCR (Figure 3-19). As YhaJ is capable of binding the LEE1 promoter, repression of the LEE is implicated to occur through *ler* regulation. However, NLE regulation can occur both dependently and independently of Ler (García-Angulo *et al.,* 2012, Holmes *et al.,* 2012, Roe *et al.,* 2007). Therefore, binding of YhaJ to a selection of NLE promoter fragments was tested by EMSA. Four NLEs were chosen, *nleA, nleB, nleG1* and *nleG6-2*, due to their location on different phage elements (Table 3-1). Binding of

YhaJ was not observed in an ~400 bp fragment of each NLE promoter region suggesting that YhaJ does not regulate NLE expression directly (Figure 3-24). It is likely that inhibition of LEE expression as a whole has downstream effects on NLE expression as translocation of effectors is dependent on the T3SS and temporal translocation.

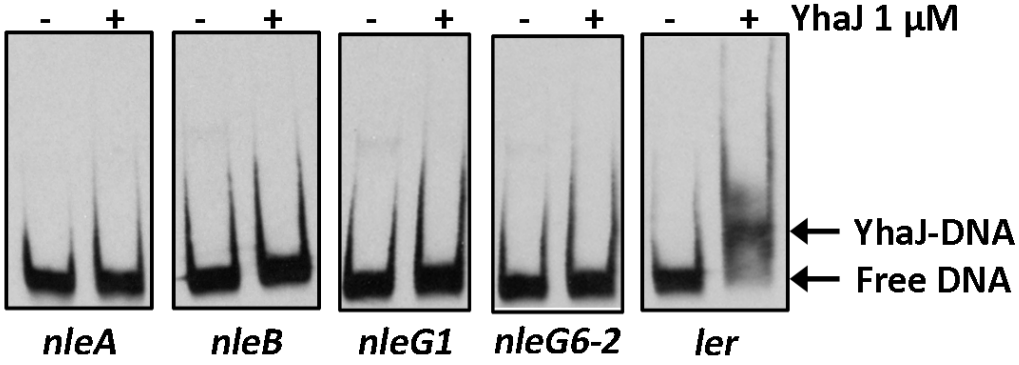

Figure 3-24 YhaJ does not regulate NLE expression directly. EMSA analysis of YhaJ interaction with ~400 bp fragments of promoter regions of *nleA, nleB, nleG1* and *nleG6-2*. As a control, interaction of YhaJ with the *ler* promoter is indicated by a positive gel shift. Free DNA and DNA-YhaJ complexes are indicated by black arrows. Lanes containing none or 1 µM of YhaJ are indicated above by a – or + respectively.

Deletion of *yhaJ* resulted in an ~11 fold reduction in *yhaO* expression as analysed by qRT-PCR (Figure 3-17) suggesting that *yhaO* is under the control of YhaJ. Investigation of the *yhaO* upstream sequence revealed the presence of degenerate forms of each putative YhaJ binding site (box 1, 76% conserved; box 2, 46% conserved; box 3, 59% conserved) (Figure 3-25A). Interestingly, the same overlapping pattern of YhaJ box 2 and box 3 that was observed for *yhaJ* was found in the *yhaO* promoter region (Figure 3-25B). Therefore, a 306 bp DNA fragment containing all three putative binding sites was tested for YhaJ binding by EMSA. No shift in DNA mobility was observed for this fragment. As YhaO is a D-serine transporter protein and is predicted to sense signals that are interpreted by YhaJ at the transcriptional level, binding was tested in the presence of D-serine as an inducing signal but no obvious band shift was observed (Figure 3-25C). Together, these data imply that YhaJ is capable of binding and regulating both LEE1 and its own expression in a constitutive manner independently of an inducing signal, whereas binding to and therefore differential regulation of the *yhaO* promoter region may be indirect or require additional signals or cofactors. A candidate for this would be YhaK that has been suggested to act as a YhaJ co-factor. The YhaJ box sequences are well conserved between

the *yhaJ* and *yhaO* promoter regions and even the distance of the overlapping box 2 and box 3 from the ATG differs by only a single nucleotide suggesting YhaJ binding potential.

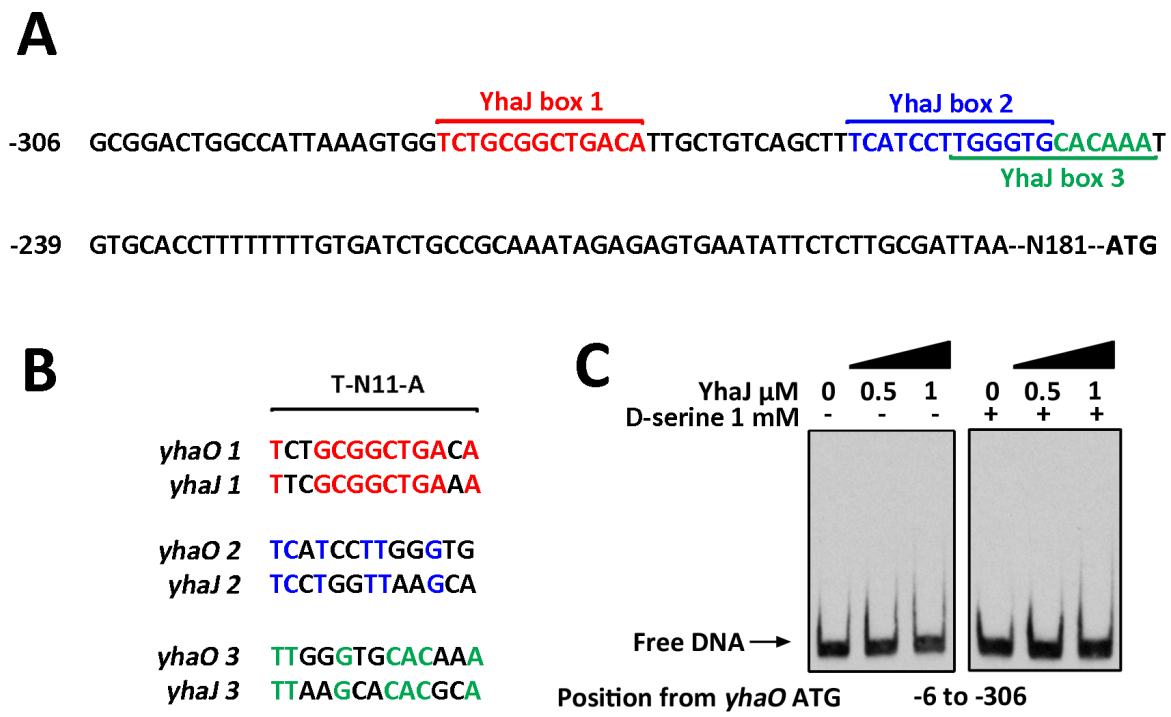

Figure 3-25 YhaJ does not directly bind the *yhaO* promoter region. (A) *yhaO* promoter region sequence. The ATG start codon is in bold. YhaJ boxes 1 to 3 are labelled above or below the corresponding sequence and colour coded red, blue and green respectively. (B) Sequence conservation of the three putative YhaJ boxes in the *yhaO* fragment. Conserved residues are colour coded by which YhaJ box they correspond to. (C) EMSA of YhaJ and a 306 bp fragment upstream of the *yhaO* ATG with and without 1 mM D-serine as an inducing signal.

3.7 Do YhaO and YhaJ affect virulence *in vivo*?

Until this point studies on the role of the *yhaOMKJ* system were performed *in vitro.* To assess if this system plays a role in regulation of virulence *in vivo* a *C. rodentium* mouse model of infection was used. *C. rodentium* has been widely used to study EHEC and EPEC infections *in vivo* due to its similarities in virulence mechanisms (colonisation via a LEE encoded T3SS) and its ability to colonise mice naturally. As EHEC and EPEC are human pathogens that do not naturally infect mice, *C. rodentium* offers an alternative small animal model for studying A/E pathogens *in vivo* (Collins *et al.,* 2014, Deng *et al.,* 2004,

Wiles *et al.,* 2006). Therefore, deletion mutants of *yhaJ, yhaO* and *ler* were made in *C. rodentium* ICC168 (a gift from Professor Brett Finlay, University of British Columbia, Vancouver; NCBI accession NC_013716.1) using Lambda Red mutagenesis and tested for their ability to colonise mice compared to the wild type strain (Datsenko & Wanner 2000).

3.7.1 *Generation and LUX marking of C. rodentium deletion mutants*

YhaJ and YhaO are currently uncharacterised in *C. rodentium.* To address if they play a role in *C. rodentium* virulence deletion mutants were generated. The *yhaJ* (ROD_47691) sequence shares 82% identity with that of EDL933 over 100% of the sequence (translated sequence shares 96% sequence identity). The *yhaO* (ROD_47651) sequence from *C. rodentium* shares 84% identity with EDL933 over 95% of the sequence. The sequence carries an in-frame truncation of 21 AAs at the N-terminus but maintains high identity (95%) over the rest of the translated sequence. Whether this has any bearing on the functionality of YhaO is unknown. The Lambda Red mutagenesis system is a one-step gene knockout technique that uses a PCR amplified kanamycin (Kan^R) marker (amplified from pKD4) flanked with DNA overhangs that are complimentary to the chromosomal flanks of the gene of interest. By induced expression of the Lambda red recombinase genes, from a temperature sensitive plasmid pKD46, one-step replacement of genes with resistance markers is achieved. The resulting mutant is then recovered at 42°C over several generations to cure the strain from pKD46 and eliminate any further recombinase activity (Datsenko & Wanner 2000). Using this technique, deletions in *yhaJ* and *yhaO* were successfully made in ICC168. Ler has been shown previously to be essential for T3S in *C. rodentium* (Deng *et al.,* 2004) and therefore a deletion in *ler* was made as a control mutant. Gene deletions were confirmed by PCR (Figure 3-26A). Each mutant derivative of ICC168 was tested for growth in LB and MEM-HEPES with no observable defects in comparison to the wild type (Figure 3-26B).

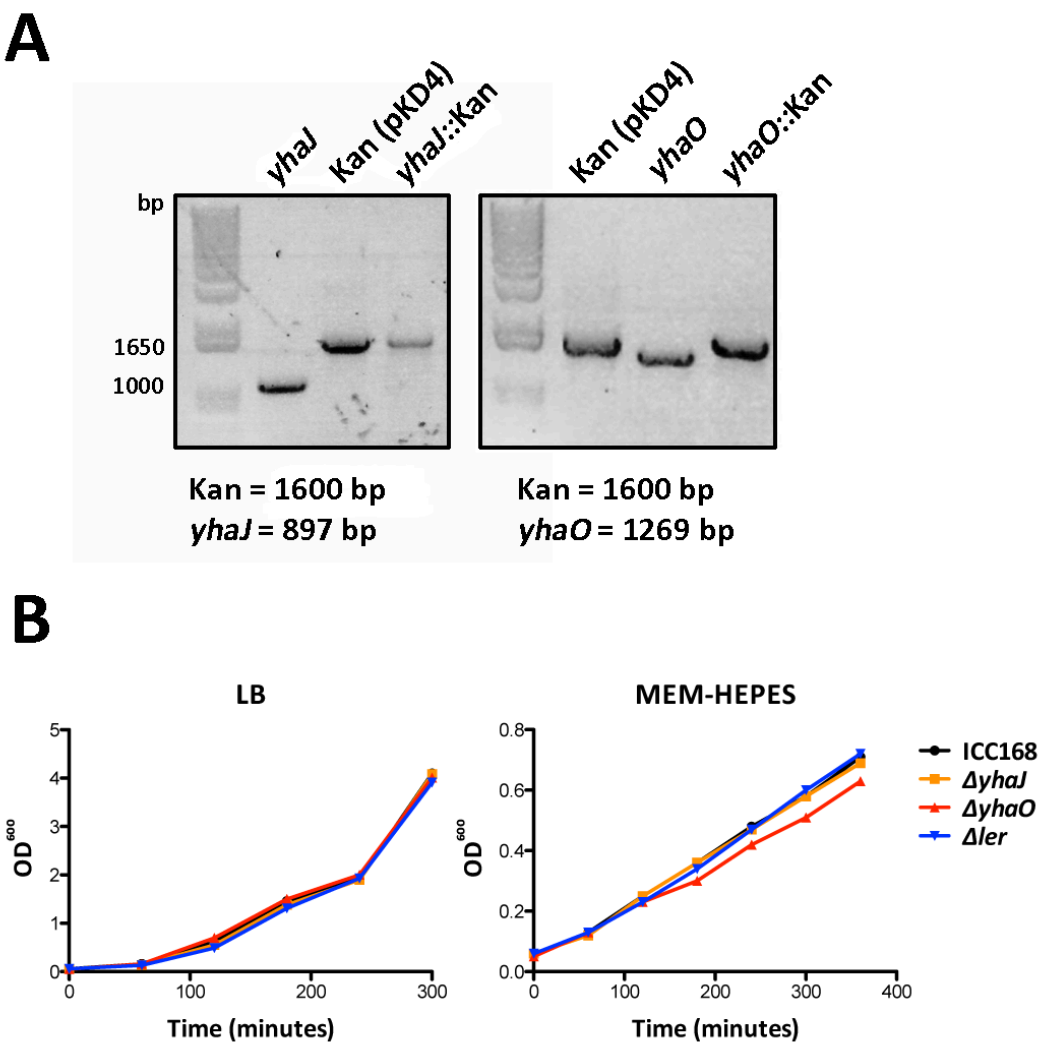

Figure 3-26 Generation of ICC168 mutants by Lambda Red. (A) Agarose gel confirmation of *yhaJ* and *yhaO* deletion mutants. PCR products were analysed for the Kan^R cassette amplified from pKD4, *yhaJ* and *yhaO* amplified from ICC168 gDNA and the Kan^R cassette amplified from gDNA after Lambda Red mutagenesis using *yhaJ* and *yhaO* region specific primers. Band sizes are indicated. (B) Growth curves of each mutant compared to the wild type in LB and MEM-HEPES. Graphs show single replicates that are representative of three biological repeats.

In order to visually track colonisation of mice at different time points during the infection, each strain was marked with the *lux* operon at the 16S locus allowing constitutive bioluminescence of the bacteria (Riedel *et al.,* 2007). This enabled visualisation of live mice for colonisation, via the *in vivo* imaging system (IVIS). Luminescence was tested by IVIS imaging of colonies or liquid culture using a FLUOstar Optima Fluorescence Plate Reader (Figure 3-27).

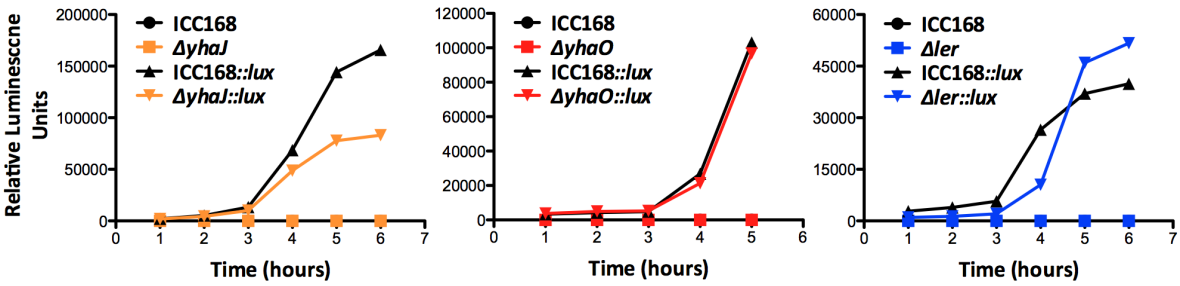

Figure 3-27 Confirmation of *lux* tagged ICC168 mutants. Liquid cultures of *ΔyhaJ::lux, ΔyhaO::lux* and *Δler::lux* were analysed for *lux* expression in LB media by luminescence monitoring using a FLUOstar plate reader. The relative luminescence units were plotted against time. Unmarked mutants were used as a control and displayed no background luminescence.

3.7.2 *BALB/c infection model using ICC168, ΔyhaJ, ΔyhaO and Δler*

The *lux* marked strains were cultured in DMEM under normal growth conditions prior to infection. Bacteria were resuspended in PBS and the five BALB/c mice per strain were orally challenged with a dose of 1 x 10^9 CFU. Mice were inoculated and monitored for colonisation by Dr. Gillian Douce. The IVIS system allows imaging of colonisation by measuring luminescence flux from the animals at different time points. Measurements were taken on days two to five post infection and then finally on day eight when the animals were culled (Figure 3-28). Colonisation was established by day three in ICC168, *ΔyhaJ* and *ΔyhaO,* although no significant difference in luminescence was observed. At day eight, *ΔyhaJ* and *ΔyhaO* maintained average levels of colonisation but the wild type ICC168 displayed a sharp average increase in luminescence flux (Figure 3-29A). To confirm these results, CFU counts were performed on faecal samples from each mouse and showed the same pattern of infection with respect to bacterial shedding in the faeces (Figure 3-29C). These results were striking but not significant and were likely due to an excessive colonisation phenotype observed in a subset (mice two and three) of the wild type mice, thus raising the overall average of the luminescence flux and CFU counts observed. It is also worth noting that mouse one of the ICC168 group and mouse four of the *ΔyhaJ* group did not establish a measureable colonisation phenotype. In agreement with the role of A/E lesions in colonisation of the host, the *Δler* mutant strain was incapable of colonising mice in all animals tested. Together, these data suggest that *ΔyhaJ* and *ΔyhaO* are not essential for colonisation by *C. rodentium in vivo*.

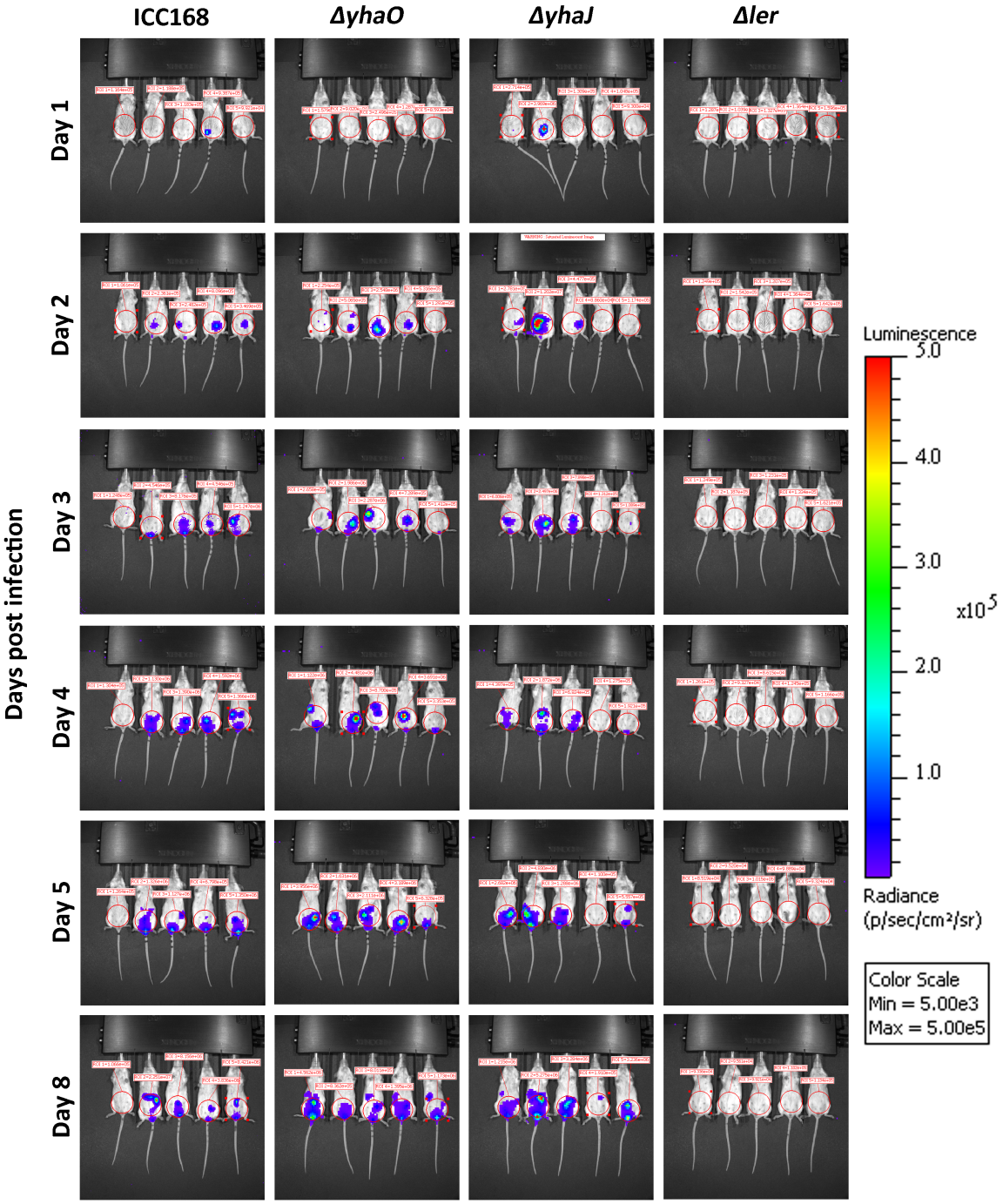

Figure 3-28 Colonisation of BALB/c mice with ICC168, *ΔyhaJ*, *ΔyhaO* and *Δler* mutants. Images of mice taken from the IVIS at days two to five and day eight post infection. The scale bar represents luminescence flux in photons per second per cm^2. Flux was normalised for all mice to allow direct comparison. The day eight ICC168 group was saturated at this scale and thus the luminescence output was adjusted for this one image.

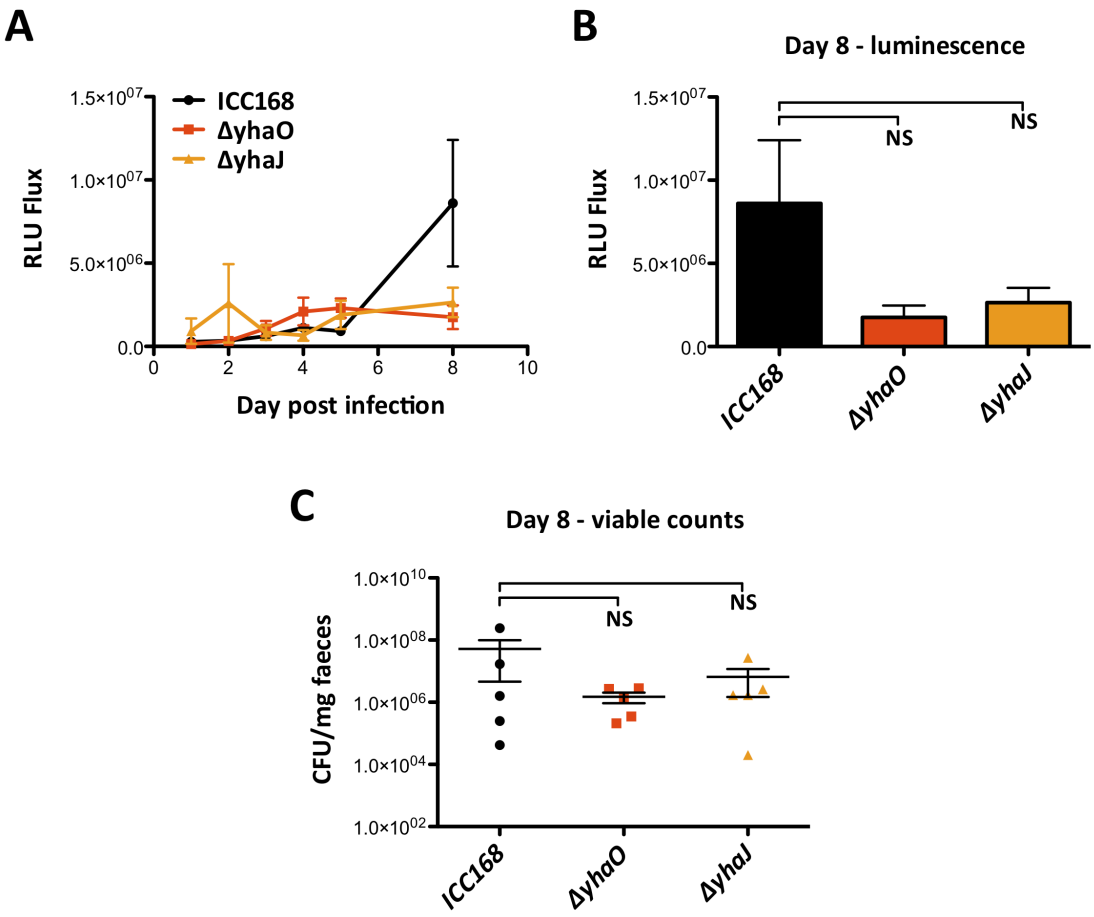

Figure 3-29 Analysis of ICC168, *ΔyhaJ* and *ΔyhaO* BALB/c colonisation data. (A) Scatter plot of the relative luminescence unit (RLU) flux at each time point. (B) Bar chart of day eight RLU flux data from figure panel A. (C) Viable counts corresponding to day eight (panel B) calculated as CFU per mg of faeces. Data is colour coded in black, red and orange for ICC168, *ΔyhaO* and *ΔyhaJ* respectively. NS denotes no significance between two groups (students T-test).

IVIS imaging and CFU counts suggested that the *ΔyhaO* and *ΔyhaJ* mutants were capable of colonising the host to the same extent as the wild type ICC168. An enhanced colonisation phenotype was observed in one of the wild type mice although this was not significant, possibly due to the low number of replicates. This suggests that although YhaJ and YhaO are required for full expression of the T3SS from TUV93-0 *in vitro,* the same may not be said for *C. rodentium.* However, this experiment is indicative of colonisation alone and not of the discrete mechanisms used by the bacteria to colonise the host. Indeed, the lack of measureable colonisation in the *Δler* mutant suggests that expression of the T3SS is absolutely required to establish an infection but these data cannot identify if an overriding mechanism may compensate for the loss of YhaO and YhaJ. To help clarify this colon samples were taken from mice from each group post-cull and prepared for

imaging by electron microscopy to identify if the *ΔyhaO* and *ΔyhaJ* mutants colonised the tissue in a similar fashion to the wild type. Scanning electron microscopy (SEM) identified clear A/E lesion formation in the wild type characterised by deep embedding of bacteria into the epithelial surface. *ΔyhaO* was found in colonies also but appeared to form less pronounced A/E lesions. *ΔyhaJ* was mostly found in large clusters in epithelial folds but appeared to be forming self-adhesive colonies as opposed to distinct A/E lesions. *ΔyhaO* and *ΔyhaJ* also displayed an unusual surface morphology not seen with the wild type. One could speculate that deletion of *yhaO* and *yhaJ* resulted in expression of alternative adhesins to overcompensate for the decrease in T3S ability. Transmission electron microscopy (TEM) revealed that A/E lesions are indeed formed in both the wild type and *ΔyhaJ* strains. Together, these data confirm the ability of *ΔyhaO* and *ΔyhaJ* to colonise the epithelium via A/E lesions but leave unanswered questions as regarding the molecular mechanisms involved in this process.

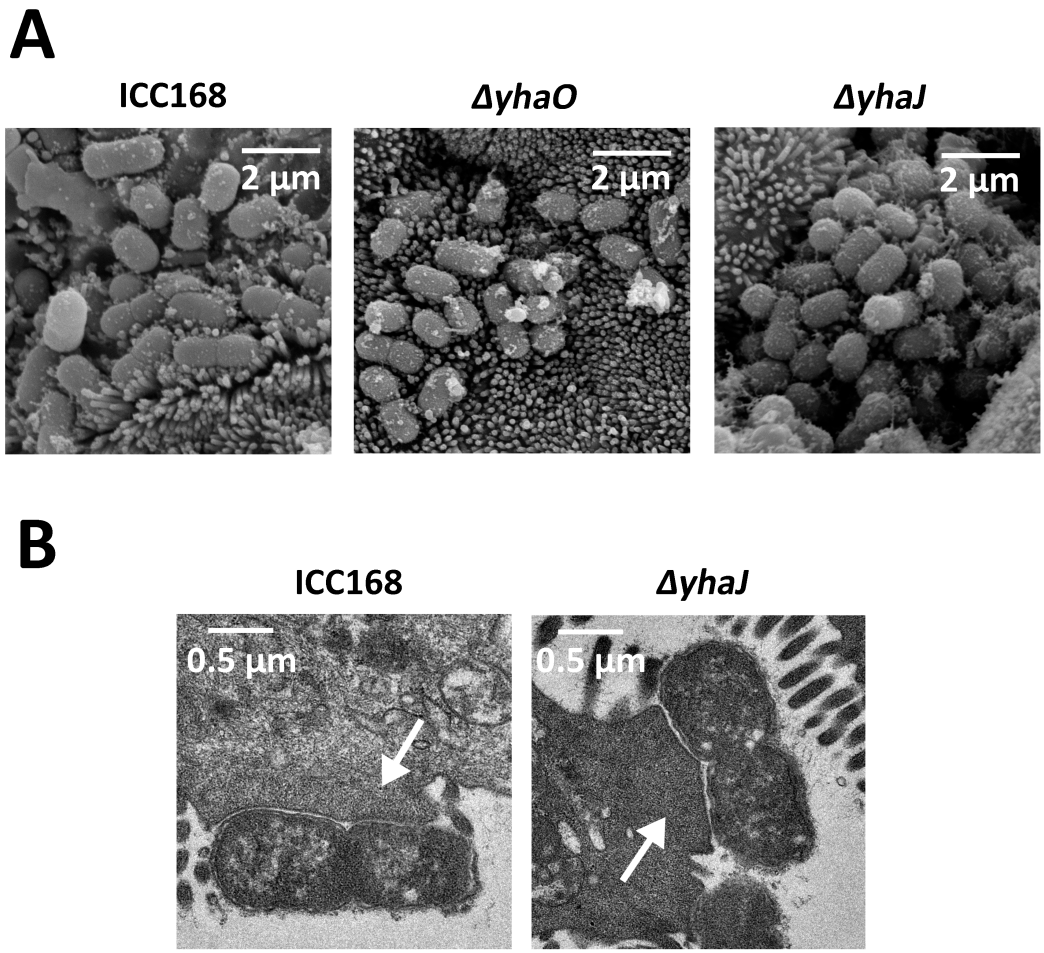

Figure 3-30 Electron microscopy of ICC168, *ΔyhaJ* and *ΔyhaO* colonisation phenotypes. (A) SEM images of ICC168, *ΔyhaO* and *ΔyhaJ* colonies from colon samples of mice used during infection models. (B) TEM images of ICC168 and *ΔyhaJ* forming A/E lesions on the host epithelium. A/E lesion pedestals are indicated by a white arrow. Scales for each image are indicated by a white bar.

3.8 Discussion

3.8.1 *A novel D-serine transportation system is expressed with the LEE*

In early microarray studies aimed at identifying *E. coli* O157:H7 ORFs that may be co-regulated under LEE inducing conditions, *yhaO* was identified. The gene was found to be upregulated along with the LEE encoded T3SS when cultured in LEE inducing conditions (MEM-HEPES tissue culture medium). YhaO was subsequently characterised as a putative inner membrane serine transporter capable of transporting both L- and D-isomers of this amino acid. The first aim of this project was to investigate the context of YhaO both genetically and physiologically. The capacity of *E. coli* to transport and metabolise D-serine has been explored previously. D-serine is known to be detrimental to the growth of *E. coli* strains lacking a D-serine deaminase, DsdA, through inhibition of β-alanine associated biosynthetic pathways (Cosloy & McFall 1973). Thus, DsdA plays a key role in D-serine tolerance. *dsdA* is co-transcribed with *dsdX,* a D-serine outer membrane transporter and the expression of this operon is controlled by the LTTR DsdC from a divergently transcribed adjacent gene (Nørregaard-madsen & Fall 1995). The *dsdCXA* system has since been characterised in detail in the UPEC strain CFT073 (Anfora & Welch 2006, Anfora *et al.,* 2007). The need for a UPEC strain to detoxify itself from exogenous sources of D-serine is obvious when one considers that D-serine is a metabolite in abundance in the urinary tract, reaching concentrations of ~1 mM (Anfora *et al.,* 2007). In fact, carriage of the *dsdCXA* locus is thought to be a key characteristic of ExPEC strains whereas InPEC strains have evolved to utilise sucrose (via the *cscRAKB* locus) as an alternative. Known InPEC including the prototypical O157:H7 strain EDL933 were found to have a conserved truncation in *dsdX* resulting in loss of most of the *dsdX* coding sequence and the entire *dsdC* gene, rendering the remaining *dsdA* useless. The *cscRAKB* genes were found in place of *dsdCX* as well as a number of adjacent genes in the InPEC isolates studied (Jahreis *et al.,* 2002, Moritz & Welch 2006). This discovery made sense, as the gastrointestinal tract is a sugar-rich environment and not known to be abundant in D-serine. Therefore, it was assumed that responding to D-serine was considered redundant in InPEC strains and was eliminated through selective pressure.

This story made the discovery of YhaO very intriguing. Why would an *E. coli* isolate known to be susceptible to D-serine toxicity encode and upregulate an alternative D-serine transporter? The *dsdCXA* system offered a useful comparator for studying YhaO. Indeed, *YhaO* is co-transcribed with the gene *yhaM,* encoding a D-serine deaminase, similarly to *dsdX* and *dsdA* from UPEC. YhaM, however, is not capable of detoxifying the bacterial cell from intracellular D-serine either in its native TUV93-0 background or in CFT073 when supplied in trans to a *ΔdsdA* mutant However, YhaO was capable of transporting D-serine in both organisms. This means that should an O157:H7 encounter environmental D-serine, they would not have any way of metabolising it, irrespective of whether they can transport it intracellularly or not, suggesting an alternative role for YhaO in O157:H7, such as D-serine 'sensing' rather than uptake. Additionally, sequence analysis confirmed the presence of *yhaOMKJ* in CFT073 but its functionality is unclear. qRT-PCR analysis of *yhaO* revealed that it was expressed lower in CFT073 cultured in MEM-HEPES suggesting a specific role in TUV93-0 under LEE inducing conditions.

The genes *yhaM* and *yhaO* are also in close proximity to a LTTR, *yhaJ,* however they are not divergently transcribed. Instead, *yhaJ* is adjacent to a divergent ORF, *yhaK,* which encodes a structurally characterised pirin like protein capable of redox control (Gurmu *et al.,* 2009). Interestingly, the authors described how the growth of two *E. coli* strains (83972 and VR50) in human urine resulted in a 12 fold increase in *yhaK* expression, as reported by another study (Hancock & Klemm 2007). They went on to hypothesise that due to its genetic positioning and similarity to known pirin like proteins, YhaK may serve as a transcriptional cofactor for YhaJ, aiding DNA binding under certain conditions. A separate study characterising global protein-protein interactions in *E. coli* K-12 further bolstered this theory by identification of YhaJ and YhaK as binding partners (Arifuzzaman *et al.,* 2006). This led to the hypothesis that YhaJ and/or YhaK likely regulates *yhaOM* similarly to the *dsdCXA* system. Expression of *yhaJ* is not particularly high in MEM-HEPES however qRT-PCR analysis of *yhaO* transcript levels in a *ΔyhaJ* background revealed a > 11 fold ($p \leq 0.05$) decrease in *yhaO* expression suggesting that YhaJ may regulate *yhaO* in a constitutive manner. Thus, the potential substrate of YhaO, D-serine, could act as an inducing signal perceived by YhaJ and/or YhaK to allow transcriptional regulation in response to this stress. Purified YhaJ was capable of binding a short fragment of the *yhaJ* promoter region that contains three putative LTTR binding boxes. However, YhaJ did not bind the *yhaO* promoter region. This result was surprising as the *yhaO* promoter fragment

tested contained all three of the putative YhaJ boxes, albeit with a certain degree of degeneracy. These results combined with the qRT-PCR analysis suggest a mechanism whereby YhaJ likely autoregulates its own expression with or without an inducing signal to maintain constitutive expression levels in the cell. YhaJ clearly plays a role in *yhaO* regulation but this response probably requires another cofactor that was not present during the EMSA experiments. Studies on the *dsdCXA* system share similarities to this whereby DsdC controls its own expression and that of *dsdXA* but this control is reliant on the presence of D-serine and is enhanced largely by the presence of an inducing cofactor, the cyclic AMP receptor protein (Bloom *et al.,* 1975, Heincz & McFall 1978, Nørregaard-madsen & Fall 1995). Binding of YhaJ was neither enhanced or diminished in the presence of D-serine under the conditions tested suggesting that if D-serine is indeed an inducing signal a cofactor is absolutely required. This offers a tantalising model whereby YhaJ regulates multiple genes in a complex manner, possibly to respond to multiple environmental signals, not just D-serine alone. Additionally, *yhaK* was also found to be upregulated by the nitric oxide donor nitroso-glutathione (GSNO) (Gurmu *et al.,* 2009, Hancock & Klemm 2007). Previous studies did not identify an upregulation of *yhaO* in human urine but this gene has been identified in two different studies as being induced by GSNO (Flatley *et al.,* 2005, Pullan *et al.,* 2007). Together these data provide a convincing genetic organisation and transcriptional model for *yhaOMKJ* in responding to environmental stresses such as D-serine and NO but the discrete mechanisms involved remain elusive.

3.8.2 *YhaO feeds into the LEE and NLE regulons in O157:H7*

YhaO was presumed to play a role in virulence by virtue of the fact that its expression was increased with the LEE under inducing conditions. Indeed, deletion of *yhaO* resulted in reduced expression and secretion of LEE associated effectors EspD and Tir into the surrounding growth media. Furthermore, the *ΔyhaO* mutant was significantly less able to colonise host cells and form A/E lesions *in vitro*. These results begged the question, how can an inner membrane transporter of D-serine have such profound effects on the expression of a pathogenicity island? Membrane proteins involved in virulence regulation have been well characterised on many occasions but these usually involve two-component regulatory systems involved in transducing environmental signals via a signal transduction cascade. The role that a membrane transporter would play in virulence

therefore implies transport of a substrate into the cell where it can act directly as a regulatory signal. RNA-seq analysis was therefore used to assess the *yhaO* global regulon and to identify co-regulated genes that may uncover a regulatory mechanism of action. Deletion of *yhaO* resulted in differential expression of 114 genes, 29 of which were enhanced and 85 of which were repressed. Most striking was the downregulation of the entire LEE PAI. All 41 ORFs were significantly repressed ($p \leq 0.001$) suggesting that YhaO feeds into regulation of the entire LEE. The LEE is under the control of Ler, which activates operons LEE2 through LEE5 (Elliott *et al.,* 2000, Haack *et al.,* 2003, Mellies *et al.,* 1999, Sánchez-SanMartín *et al.,* 2001). The *grlRA* feedback regulatory system was also negatively affected by *ΔyhaO*. This system is activated by Ler and forms a feedback loop to enhance *ler* expression from the LEE1 operon (Deng *et al.,* 2004). Thus, YhaO feeds into the core regulatory circuit of the LEE. This may explain why *yhaO* is expressed in parallel with the LEE under inducing conditions. Perhaps the *yhaOMKJ* system helps regulate steady state expression levels of the LEE in order to quickly respond to environmental stresses sensed through YhaO. Indeed, the regulatory circuit controlling the LEE is extremely complex and multiple global regulators have evolved to feed into this system for temporal control in response to the environment (Mellies *et al.,* 2007). Interestingly, a recent study characterising the Ler regulon in EPEC and EHEC identified *yhaO* as being repressed by Ler during late log phase in EPEC (Bingle *et al.,* 2014). The authors described how the expression of the LEE is growth phase dependent. Our experiments were performed at mid log phase but it is interesting to note that *yhaO* has been independently identified as a member of the Ler regulon in another study. Perhaps as the bacteria enter a later stage of their lifecycle during the course of an infection the expression of the LEE must be regulated differently to maintain colonisation of the host. It is also important to note that EPEC harbour unique regulatory mechanisms not employed by EHEC to elicit control over the LEE, such as the plasmid encoded regulator (Per) system, that may account for variations in *yhaO* expression (Bingle *et al.,* 2014, Mellies *et al.,* 1999).

Deletion of *yhaO* not only resulted in LEE repression but also the downregulation of 20 NLEs. Genes encoding NLEs are located throughout the O157:H7 genome on horizontally acquired O-islands and insertion sequence elements (Perna *et al.,* 2001, Tobe *et al.,* 2006). Regulation of NLEs is not understood to the same extent as that of the LEE. For instance, Ler and GrlA have been shown to be directly necessary for expression of *nleA*

and *nleH2* whereas *nleH1* and *nleB2* are regulated differently via a conserved NRIR sequence (García-Angulo *et al.,* 2012, Holmes *et al.,* 2012). Its not surprising that NLEs have multiple regulatory mechanisms as they have been acquired independently on phage elements but there must be some level of control that relates to the expression of the LEE as the T3SS is their export machine that facilitates translocation to the host cell cytoplasm. It is thought NLEs are temporally regulated to minimise competition for T3SS associated export and allow establishment of A/E lesions before further subversion of the host cell. Indeed, *nleH1* and *nleH2* were both shown to be repressed upon initial contact with host cells (Holmes *et al.,* 2012). The *ΔyhaO* RNA-seq data also indicates that expression levels of most NLEs is lower than that of LEE encoded genes, an observation that has been made previously (Roe *et al.,* 2007). These findings suggest that YhaO feeds directly into the LEE and this has a knock on effect on the subsequent expression of NLEs irrespective of their regulatory mechanisms.

Aside from the negative effects on virulence, patterns of gene expression in relation to cellular processes were fairly undefined in this dataset. There was significant up and down regulation of genes involved in varied metabolic processes, membrane protein composition and amino acid metabolism, none of which related to serine. Its possible that YhaO may have been involved in metabolic processes in a more ancestral sense and thus linked to the regulation of more than one cellular process, although this is speculative. Interestingly, deletion of *yhaO* resulted in a > 5 fold increase in *yhaM* expression. As *yhaO* and *yhaM* are cotranscribed, this appears to be attributed to overcompensation at the *yhaOM* promoter. Levels of *yhaJK* were unaffected by this deletion but that does not rule out this hypothesis. As described above, YhaJ can autoregulate without an inducing signal or a cofactor. Therefore, loss of YhaO may result in an environmental sensory defect rendering its tight regulation redundant. It is possible that growth in MEM-HEPES, an environment containing no D-serine or NO stress, will have no affect on excessive expression of *yhaO* above necessary levels but deletion of *yhaO* removes this checkpoint resulting in *yhaM* overexpression. It is also possible that other regulatory elements may act in concert with YhaJ to regulate *yhaOM* and removing this careful balance may have downstream effects on other sections of the genome potentially regulated by such elements or YhaJ. It is important to note that most genes affected in the *ΔyhaO* mutant that are not LEE or NLE associated displayed fold changes of < 2. These changes were significant but may be a more transient effect of the deletion

as opposed to being directly involved in the *yhaO* regulon, which may explain the lack of clear expression patterns. A more targeted approach is to look at expression levels of transcriptional regulators affected by deletion of *yhaO.* Three transcriptional regulators were downregulated (*leuO, yhcK* and *rcsA*) and only one (*csgD*) was enhanced in *ΔyhaO.* CsgD is the regulator of curli fimbriae, structural adhesins essential for biofilm formation, and is part of complex regulation network involving five cellular transcription factors (Ogasawara *et al.,* 2010). RcsA regulates synthesis of the bacterial capsule, YhcK is a regulator for sialic acid uptake systems and LeuO is an LTTR involved in global transcriptional regulation as well as prolonged expression of the LEE (Ebel & Trempy 1999, Plumbridge & Vimr 1999, Takao *et al.,* 2014). These regulators are all involved in independent processes but its interesting to note that LeuO and RcsA have both been shown to form part of the Ler regulon and are enhanced by expression of Ler (Bingle *et al.,* 2014, Takao *et al.,* 2014). Therefore, downregulation of the LEE in *ΔyhaO* appears to have significant effects on the larger LEE regulon, not just *ler* itself.

3.8.3 *YhaJ as a mediator the ΔyhaO phenotype*

Transcriptional profiling of *ΔyhaO* showed a clear pattern of LEE and NLE associated repression but how this signal is transmitted to the genome was unclear. As discussed above, YhaJ is absolutely required for *yhaO* expression, which is in turn required for full LEE expression in TUV93-0. It was therefore hypothesised that YhaJ, being a LTTR, was capable of transmitting signals transported or perceived by YhaO to the genome. Surprisingly, YhaJ did not bind the *yhaO* promoter region under the conditions tested independently or in the presence of D-serine as an inducing signal but YhaJ was capable of binding its own promoter alone. Expression levels of *yhaJ* were also not altered in the *ΔyhaO* mutant suggesting that YhaJ autoregulates its own expression probably to maintain a constant basal expression level within the cell. Based on this assumption, a *ΔyhaJ* mutant was tested for its ability to secrete LEE encoded effectors and bind host cells. Similarly to *ΔyhaO,* the *ΔyhaJ* mutant showed reduced expression and secretion of the LEE encoded effector EspD by SDS-PAGE and western blot analysis and also an impaired ability to colonise host cells via A/E lesions. qRT-PCR analysis of *ler, espD, tir, nleA* and *nleG* expression levels confirmed the repression of multiple LEE operons, including the master regulator, and NLE regions. These data highlighted the indistinguishable phenotype observed between *ΔyhaO* and *ΔyhaJ.* Combined with the

data described above, YhaJ emerged as the likely candidate to mediate the *ΔyhaO* phenotype and a novel regulator of the LEE. Indeed, purified YhaJ was capable of binding the LEE1 promoter region independently of an inducing signal. This fragment contained all three putative LTTR boxes found in the *yhaJ* promoter region but positioning relative to the translational start site was unique. Ler has been shown previously to be absolutely required for expression of the LEE and interruption of another regulator that is feeding into its expression would likely disrupt this process (Deng *et al.,* 2004). Indeed, the LEE repression observed in the *ΔyhaO* background was overcome by supplying Ler *in trans* on an arabinose inducible promoter. This result confirmed that the repression was at the core transcriptional level. How control of the LEE was affected by deletion of *yhaO* is unclear but, as discussed above, deletion of *yhaO* resulted in overcompensation at the *yhaOM* promoter. It is therefore conceivable that this over activity of a YhaJ dependent promoter may have knock on effects on other genes that are regulated by such a low-expressed transcriptional regulator. YhaJ was also tested for its ability to bind four unique NLE promoter regions. These were selected from different phage elements and included Ler and NRIR dependent and independent genes. No binding was observed to any NLE promoter agreeing with the earlier observation that the LEE is likely regulated by YhaO and subsequent deletion results in downstream repression of temporally associated genes. Together, these data allowed us to construct a novel model of LEE regulation, whereby YhaJ is expressed constitutively under LEE inducing conditions helping to regulate steady state virulence expression and this regulation is capable of responding to YhaO and signals that it has perceived or transported from the environment through potential interaction with YhaK (Figure 3-31).

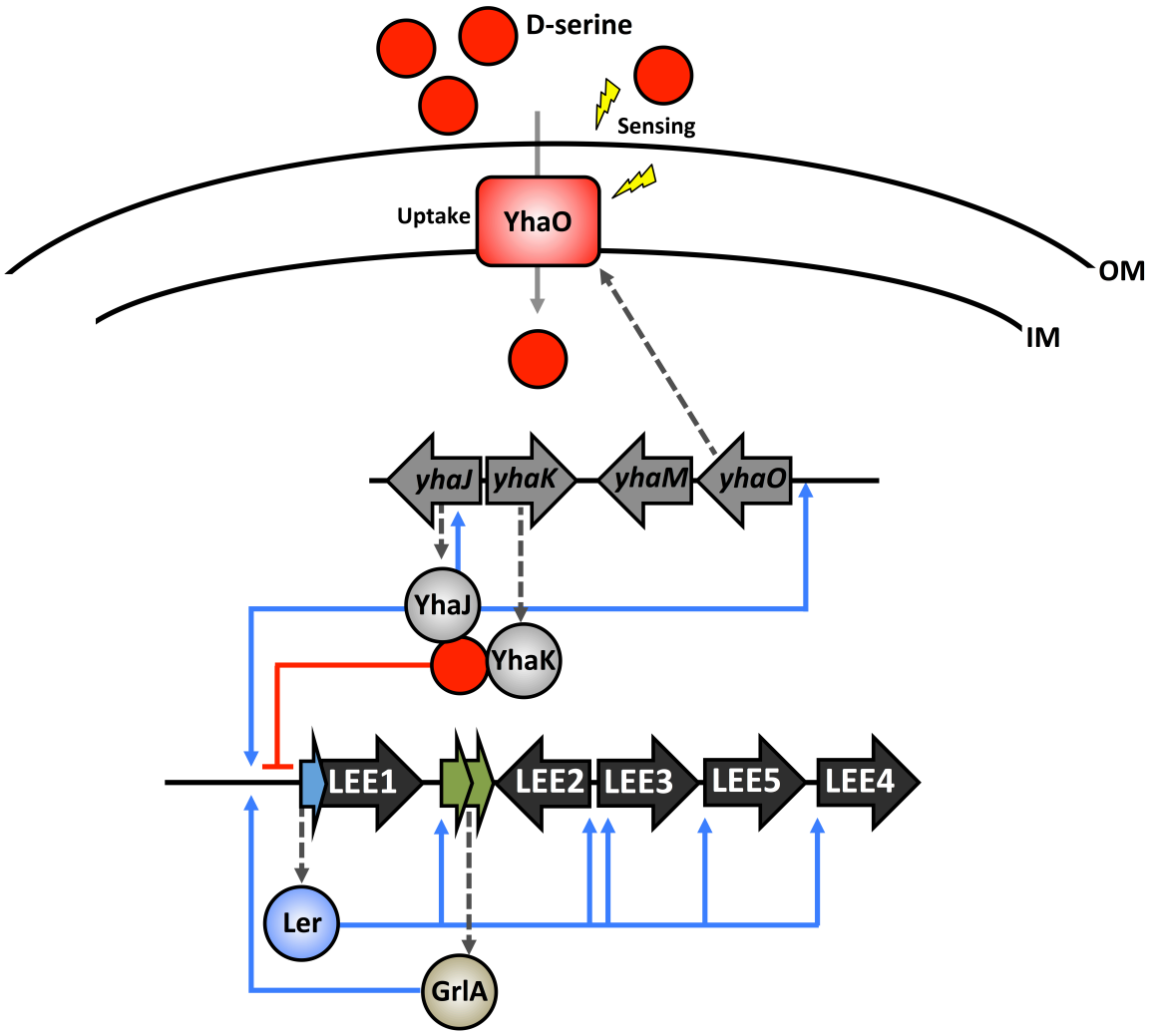

Figure 3-31 Regulation of the LEE by YhaJ. Schematic model of how YhaJ may regulate the LEE through YhaO. YhaJ autoregulates its own expression constitutively and helps too maintain steady state levels of *ler* expression. YhaJ regulates *yhaO* expression in tandem with YhaK. This YhaJ-YhaK interaction is hypothesised to be capable of responding to environmental signals such as D-serine although whether these signals are direct or sensed by YhaO is unknown. This system therefore allows rapid response to environmental stresses so as to alter expression levels of the LEE accordingly. Blue lines indicate positive regulation and red lines indicate negative regulation.

3.8.4 *YhaO and YhaJ control of the LEE appears to be EHEC specific*

The data described above provide an intriguing model of regulation by the *yhaOMKJ* system in *E. coli* O157:H7. Culture of LEE positive pathogens in DMEM and MEM-HEPES is known to mimic the intestinal environment and promote expression of the LEE allowing detailed studies on the functionality of this system *in vitro* (Deng *et al.,* 2004, Kenny *et al.,* 1997a, Roe *et al.,* 2007). Although this has proved incredibly informative, the importance of virulence factors can only be truly assessed by using an *in vivo* model of infection. As

described in section 1.4.6 the intestinal physiology has emerged as a key source of LEE modulating factors with components such as SCFAs, available sugars, hormones and bacterially produced quorum sensing factors all being capable of affecting virulence expression (Nakanishi *et al.,* 2009, Pacheco *et al.,* 2012, Sperandio *et al.,* 1999, Takao *et al.,* 2014). This network of regulation is extremely complex and must be carefully tuned to allow adaptive responses to numerous signals.

EHEC is a human pathogen and does not naturally colonise the intestine of mice, the common small animal model for infection studies. Rabbit and infant calf models of EHEC infection have been established although these are expensive systems that do not offer the advantages of a small animal model (Dean-Nystrom *et al.,* 1998, Marchès *et al.,* 2000, Ritchie & Waldor 2005). *C. rodentium* is a natural mouse pathogen that encodes the LEE, which is imperative for *C. rodentium* virulence in mice. Thus, *C. rodentium* has become the model organism for studying A/E pathogenicity *in vivo* (Collins *et al.,* 2014, Deng *et al.,* 2004). *yhaJ* and *yhaO* mutants were therefore tested in the prototypical *C. rodentium* strain ICC168 and tested for their ability to colonise mice in comparison to the wild type and a *Δler* mutant. In contrast to *in vitro* studies using TUV93-0, *ΔyhaJ* and *ΔyhaO* mutants were not significantly attenuated for virulence in ICC168 whereas colonisation in the *Δler* mutant was completely abolished. Both *ΔyhaJ* and *ΔyhaO* showed a luminescence flux and faecal shedding count comparable to that of the wild type. At day eight of infection however, a sharp spike in luminescence was observed for a subset of the wild type infected mice only. Although only two of the four mice in this group exhibited this strong phenotype, this result suggests that the wild type ICC168 is capable of persisting more aggressively in a longer-term infection than *ΔyhaJ* or *ΔyhaO.* Indeed, *in vitro* studies of *ΔyhaJ* or *ΔyhaO* in TUV93-0 did not suggest that these genes are absolutely required for host cell colonisation but rather that they may play a specific role in a portion of LEE regulation resulting in strong attenuation of the pathogen as opposed to abolition of infectious capability. This experiment was not without limitations. Only five mice per strain were tested and in the wild type and *ΔyhaJ* groups, colonisation was established in four of these mice. The trend of colonisation and faecal shedding did not suggest *ΔyhaJ* or *ΔyhaO* were less virulent than the wild type and these data were not significant therefore the experiment was not carried on past day eight. *C. rodentium* can be detected in faecal samples up to two weeks post infection (Collins *et al.,* 2014). To validate this result in detail the experiment would need to be repeated using larger

groups of mice per strain and perhaps over a longer time course to establish whether these results are truly significant.

The molecular mechanism of YhaO and YhaJ mediated control of the LEE is still currently under investigation in O157:H7 but the data presented in this thesis strongly suggests modulation of Ler in response to environmental signals. Although EHEC, EPEC and *C. rodentium* are comparable in their pathogenesis via A/E lesions, one must not forget that they are indeed unique pathogens with unique host range. The core regulation of the LEE via Ler/GrlRA control in all three pathogens is established and conserved but unique mechanisms are present in each case that assists in this control. Deng *et al.* first described the discrepancies between the *C. rodentium* LEE1 promoter region and that of EHEC by highlighting that only 194 bp upstream of the *ler* start codon can be aligned with the EHEC region (79% homology) whereas the same regions of EHEC and EPEC are almost identical (98% homology) (Deng *et al.,* 2001). As discussed in section 1.4.1 the LEE1 upstream region in EHEC contains two promoters designated P1 and P2. Activity is thought to be regulated primarily through P1 but P2 is still required for maximal expression of the LEE1 operon (Islam *et al.,* 2011a, Porter *et al.,* 2005, Sharp & Sperandio 2007, Sperandio *et al.,* 2002). Interestingly the P2 promoter is not present in EPEC or *C. rodentium* and may indicate an intricate EHEC specific regulatory system of *ler.* Furthermore, each of the three putative YhaJ binding boxes were located downstream of the P1 promoter, with the predicted primary YhaJ box located between this and the EHEC specific P2 promoter. This suggests that regulation of the LEE by YhaJ may be acting on the EHEC specific LEE1 P2 promoter thus explaining the lack of a phenotype in *C. rodentium* (alignment of the LEE1 promoter region from EHEC, EPEC and *C. rodentium* is illustrated in Figure 8-1). Additionally, EPEC LEE1 is regulated by the plasmid encoded regulator (Per) system whereas EHEC LEE1 is under the control of a chromosomally encoded system homologous to Per, known as Pch (Iyoda, S. and Watanabe 2004, Mellies *et al.,* 1999). It is important to note that no Per-like homologue has been identified in *C. rodentium.* Other well established regulators of the LEE such as the quorum sensing system have also not been thoroughly explored for this pathogen (Yang *et al.,* 2010) and conflicting observations on the role of RpoS have been described for EHEC (Dong & Schellhorn 2009b, Iyoda & Watanabe 2005, Sperandio *et al.,* 1999). Recently, a *C. rodentium* specific global transcriptional regulator, RegA, was identified as being a key activator of *grlA* expression in response to environmental bicarbonate (Hart *et al.,* 2008,

Yang *et al.,* 2009) but no homologue of RegA exists in EHEC. Finally, It is also important to note the observed differences in colonisation phenotypes of the colon between ICC168, *ΔyhaJ* and *ΔyhaO.* One could hypothesise that YhaJ may indeed be required for A/E lesion formation but *C. rodentium* may harbour unique adhesive properties that can overcompensate for defects in A/E lesion formation. This however is highly speculative and would require in depth investigation. Studies on ICC168 *ΔyhaJ* and *ΔyhaO in vitro* would be useful in elucidating the molecular mechanisms of these results and may explain the observed pathogen specificity.

Taken together, these findings indicate the LEE regulation is more complex than previously anticipated and displays a large degree of pathogen specificity. This may be to account for differences in the human and mouse intestinal physiology for instance or it may be as a result of the adaptation of the LEE PAI into different genetic backgrounds. *C. rodentium* is a relatively recent pathogen and has been hypothesised to have emerged through the development of laboratory mouse models. Its genome is unstable and subject to large scale recombination (Petty *et al.,* 2011). Therefore its existence may be considered somewhat forced with respect to the natural mouse host and therefore imply unique adaptations of specific genetic systems to allow the expression of the LEE PAI.

3.8.5 *Conclusion*

The ability of temporally controlled virulence factors to be regulated requires systems that can sense and respond to the surrounding environment. This study has identified and characterised the genetic, transcriptional and physiological context of a novel environmental sensory system, *yhaOMKJ,* in *E. coli* O157:H7. This system is regulated by environmental stimuli such as urine and NO stress. Additionally, the study specifically identified D-serine as a substrate for the system. It is proposed that this system responds to environmental signals to regulate the T3SS, a major colonisation factor in A/E pathogens, in a novel manner. The precise mechanism of how this regulation occurs is still under investigation but it is proposed that YhaJ might be a direct regulator of LEE expression, eliciting control in a novel manner. This study provides a strong framework for further research and an intriguing story of how *E. coli* O157:H7 can respond to stresses encountered in the environment.

4 Investigating the response of *E. coli* O157:H7 to the host metabolite D-serine

4.1 Introduction

Chapter 3 described the discovery and characterisation of a novel stress sensory system, *yhaOMKJ,* in *E. coli* O157:H7. It was discovered that this system was capable of responding to the host metabolite D-serine by intracellular transport. D-serine has been previously characterised as being toxic to *E. coli* strains lacking a functional DsdA. The exact mechanism of toxicity is not well defined but it has been proposed to inhibit L-serine biosynthesis by interfering with the initial enzymatic steps of its synthesis and pantothenate biosynthesis by competitively inhibiting β-alanine coupling to pantothenate pre cursors. This results in bacteriostasis when grown in minimal media supplemented with D-serine (Cosloy & McFall 1973, Maas & Davis 1950). Conversely, the metabolite can be used as a carbon source in bacteria that harbour a functional DsdA by breaking down D-serine into non-toxic pyruvate and ammonia, a positive fitness trait commonly associated with uropathogenic bacteria (Roesch *et al.,* 2003, Sakinç *et al.,* 2009). These studies led to the hypothesis that ExPEC strains have maintained the ability to metabolise D-serine allowing them to disseminate to colonisation sites beyond the gastrointestinal tract. In contrast, InPEC strains have mutated the *dsdCXA* locus and replaced it with the sucrose utilisation locus *cscRAKB* presumably as the ability to metabolise D-serine as a carbon source is not a required trait for survival in the gastrointestinal tract, an environment containing more readily available sugars for growth (Jahreis *et al.,* 2002, Moritz & Welch 2006). It was therefore surprising to discover the function of the *yhaOMKJ* system in *E. coli* O157:H7. YhaO allows uptake of D-serine but the deaminase YhaM cannot detoxify the cell. This ultimately led to the hypothesis that transport or "sensing" of D-serine was an important trait of InPEC strains irrespective of the ability to metabolise D-serine.

D-serine has been proposed to act not only as a nutrient source for uropathogens but also as a regulatory signal for expression of virulence factors. Anfora and colleagues described how CFT073 *ΔdsdA* displayed a hypercolonisation phenotype that they attributed to D-serine accumulation and subsequent overexpression of virulence genes. Moreover, a *ΔdsdA/dsdX/cycA* triple mutant was attenuated for kidney colonisation indicating a requirement for D-serine uptake in prolonged UTI (Anfora *et al.,* 2007). More recently, D-serine has been identified as a signal for both niche fitness and expression of

at least one virulence factor in the uropathogen *S. saprophyticus* (Korte-Berwanger *et al.,* 2013). A role for D-serine has not previously been investigated in O157:H7 but based on the above studies, the role of the D-serine transporter YhaO in expression of the T3SS and the knowledge that InPEC strains expressing *yhaO* can at least transport D-serine, it was hypothesised that this metabolite may act as a physiologically relevant signalling molecule involved in virulence gene regulation in O157:H7. Given that UPEC and *S. saprophyticus* are distinct, it is possible that D-serine may have broader functions in a wide variety of bacteria.

High concentrations of D-serine have previously been quantified in the brain and urine (Anfora *et al.,* 2007, Wolosker *et al.,* 1999, 2008) however an analysis of D-serine concentrations in the intestine has not been performed. D-serine is known to be endogenously produced but may also be readily available in the gut depending on dietary intake (Asakura & Konno 1997, Man & Bada 1987). Recent studies have also highlighted the emerging roles and bacterial origin of NCDAAs in bacteria (Cava *et al.,* 2011a,b; Lam *et al.,* 2009). Thus, it is possible that InPEC strains may encounter inhibitory metabolites such as D-serine outside of the urinary tract. This chapter focuses on the phenotypic and transcriptional effects that D-serine has on *E. coli* O157:H7 specifically. Selected data from this chapter has been published in the ISME journal and is presented in the section 9 of this thesis (Connolly *et al.,* 2015).

4.2 The effects of D-serine on growth and motility in O157:H7

The main goal of this work was to assess whether D-serine affected gene expression in O157:H7 using the strain TUV93-0. However, in order to compare treatment with D-serine to the wild type alone it was first essential to assess what, if any, effects that supplementation of D-serine into the chosen media had on the cells. The D-serine transporter described in chapter 3, YhaO, was upregulated co-ordinately with the T3SS and therefore the hypothesis was that D-serine sensing from the environment might be important for expression of the LEE under conditions that mimic the gastrointestinal tract. However, before examining the effects of D-serine on virulence gene expression a

more general approach to examine the effects of D-serine on growth and motility was carried out.

4.2.1 *D-serine is not bacteriostatic to TUV93-0 in MEM-HEPES/DMEM*

D-serine has previously been demonstrated to be bacteriostatic to *E. coli* unable to express a functional DsdA when grown in minimal media at a concentration of ~475 µM (Cosloy & McFall 1973). The environmental concentrations of D-serine are capable of exceeding this range. For instance, the concentrations of D-serine in urine have been measured from ~28 µM to > 1 mM (Anfora *et al.,* 2007). This upper limit was therefore used in growth experiments as it represents a physiologically relevant concentration. Three growth media were tested – MEM-HEPES, DMEM (both T3S inducing media) and M9 minimal media. Bacteria were grown overnight in LB and sub-cultured into each of the above media for growth at 37°C. Growth of TUV93-0 in MEM-HEPES and DMEM was not affected by supplementation of the media with 1 mM D-serine however a > 3 fold reduction in growth was observed in M9 minimal media containing 1 mM D-serine (Figure 4-1). This repression could be overcome by supplementation of L-serine into the media, confirming an observation reported previously (Cosloy & McFall 1973). These data confirm previous findings that at high concentrations D-serine are capable of repressing growth in minimal media. Interestingly, growth in MEM-HEPES or DMEM was not affected suggesting that other factors in this media may overcome the growth defects. MEM-HEPES contains ~238 µM L-serine and ~4 µM pantothenic acid which are 500 and 1000 (respectively) fold greater than the amount reported to overcome D-serine inhibition in minimal media (Cosloy & McFall 1973). These data highlight the ability of TUV93-0 to tolerate high concentrations of D-serine in growth conditions that mimic the gastrointestinal tract. This finding is important as it allows analysis of gene expression in response to D-serine under growth conditions comparable to the wild type alone.

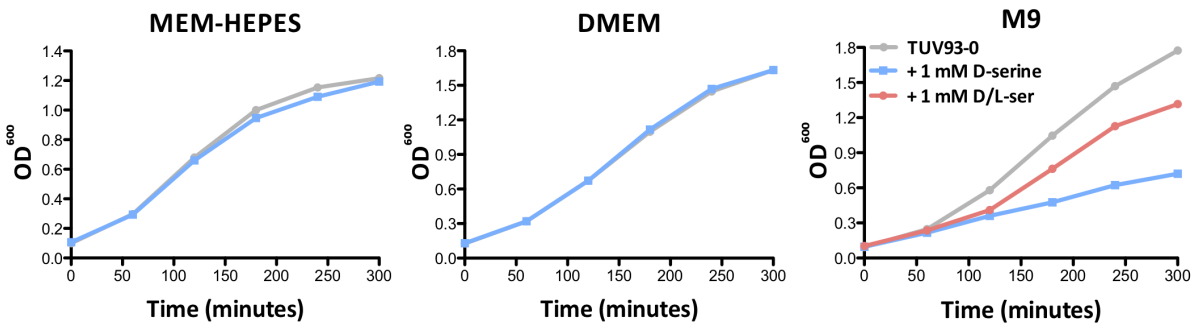

Figure 4-1 Effects of D-serine on the growth of TUV93-0. OD_{600} was measured over time for TUV93-0 (grey) cultured in MEM-HEPES, DMEM and M9 minimal media. No difference in growth was observed for MEM-HEPES or DMEM but D-serine (blue) severely repressed growth in M9. This repression was partially overcome by supplementation of the media with L-serine (red) also. Growth curves are representative of three biological replicates.

The viability of TUV93-0 cultured in MEM-HEPES supplemented with 1 mM D-serine was assessed also. Samples of growth culture were removed at OD_{600} of 0.8, serially diluted and spot plated onto LB agar plates in triplicate for calculation of CFUs. No significant affects were observed on the viability of bacterial cultures from growth experiments in MEM-HEPES confirming that TUV93-0 cultured with and without D-serine under T3S inducing conditions was indistinguishable.

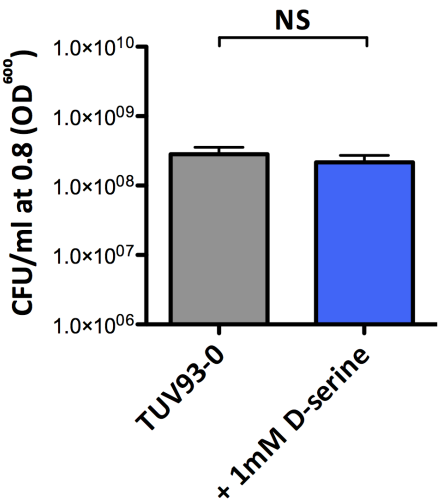

Figure 4-2 Viability of TUV93-0 after exposure to D-serine. Viable counts were performed on TUV93-0 from cultures in MEM-HEPES (with and without 1 mM D-serine) at an OD_{600} of 0.8. No significant difference in viability was observed between the two growth conditions (students T-test). Experiments were performed in biological triplicate and samples at each dilution for each replicate were plated in triplicate.

As a positive control for the growth experiments on TUV93-0, two strains were selected for comparative growth in MEM-HEPES and M9 minimal media supplemented with 1 mM D-serine; CFT073 as it is the prototypic UPEC strain that has been used for studies on the *dsdCXA* system and *E. coli* K-12 (MG1655) as it is the common laboratory workhorse and a commensal strain. Both CFT073 and K-12 have been characterised previously as harbouring the full *dsdCXA* system and can therefore use D-serine as a sole carbon source in stark contrast to TUV93-0 (Moritz & Welch 2006). CFT073 and K-12 had very similar growth profiles. Both strains grew in M9 supplemented with D-serine just as well as the wild type alone, in contrast to TUV93-0, which was inhibited for growth in M9 plus D-serine. Interestingly, growth in MEM-HEPES with D-serine was improved over the wild type for strains expressing the *dsdCXA* system. This indicates that these strains can use D-serine not only as a carbon source but also as a positive supplement used alongside other nutrients to improve fitness in a more defined environment.

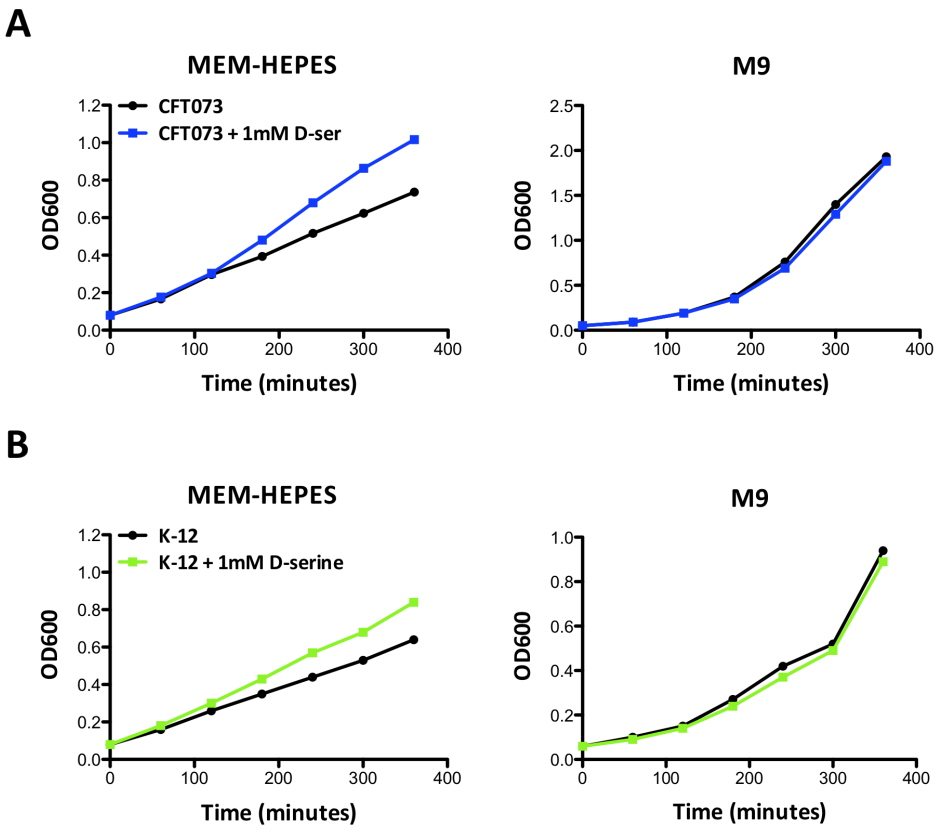

Figure 4-3 Growth profiles of CFT073 and K-12 in media supplemented with D-serine. Growth curves of CFT073 (A) and K-12 (B) in MEM-HEPES and M9 minimal media with and without D-serine measured as OD_{600} over time. Growth curves are representative of triplicate biological replicates.

4.2.2 *D-serine increases bacterial motility*

The motility of *E. coli* allows the bacteria to navigate their environment in search of nutrients or a particular niche for colonisation. Additionally the use of bacterial chemotaxis allows spatial and temporal control over the response to certain nutrients in real time. L-serine is known to be a major chemo-attractant in *E. coli* along with a number of other L-AAs (Adler 1975, Mesibov & Adler 1972). Recently a number of D-AAs have been suggested to act as weak chemo-attractants in CFT073 although the data obtained were not statistically significant (Raterman & Welch 2013). Based on these data the idea of D-serine contributing to bacterial motility was explored further. TUV93-0 is non-motile in MEM-HEPES and DMEM as this media mimics physiological LEE inducing conditions (Beckham *et al.,* 2014, Tobe *et al.,* 2011). The motility of TUV93-0 with and without D-serine was assessed by liquid culture in MEM-HEPES with and without D-serine followed by sub-inoculation onto motility agar plates containing no or 1 mM D-serine. The wild type grew to a diameter of ~35 mm after eight hours growth on motility agar. Exposure to D-serine increased the average bacterial diameter to ~45 mm ($p = 0.0043$) suggesting that the presence of D-serine in the growth media can induces changes in key cellular processes such as motility.

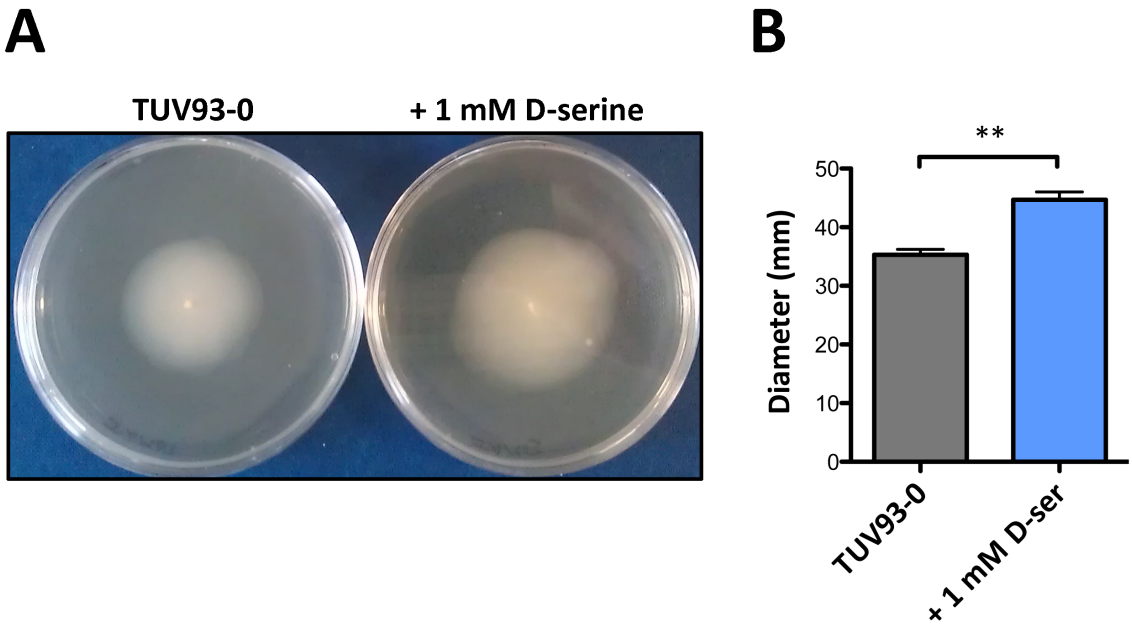

Figure 4-4 Motility of TUV93-0 after exposure to D-serine. (A) Representative images of TYV93-0 inoculated on motility agar plates alone and after exposure to D-serine. (B) Measurement of TUV93-0 bacterial growth diameter (mm) after eight hours inoculation on motility agar plates with and without exposure to D-serine. ** denotes $p \leq 0.01$ (students T-test). Experiments were performed in biological triplicate.

D-serine appeared to increase bacterial motility on motility agar plates. As TUV93-0 is normally non-motile under LEE inducing conditions (Beckham *et al.,* 2014, Tobe *et al.,* 2011), expression of FliC (flagellin) was analysed from TUV93-0 in MEM-HEPES liquid culture. Whole cell lysates were obtained at OD_{600} of 0.6, proteins separated by SDS-PAGE and analysed by western blot using anti-FliC antibodies. FliC (~60 kDa) was barely detectable in the wild type alone but was markedly increased after exposure to 1 mM D-serine (Figure 4-5A). The expression of *fliC* was transcriptionally monitored using a GFP-promoter fusion construct to measure relative fluorescence units (RFU) over time (Figure 4-5B). Expression of the *fliC* gene was increased during growth but this trend was not significant. Furthermore, anti-FliC antibodies were used to detect the formation of flagellar structures on the surface of bacterial cells by fluorescence microscopy (Figure 4-5C) however the proportion of flagellated bacteria per replicate varied with a number of cells still lacking flagella. This result confirmed not only the over-expression of FliC but also the formation of flagella on the cell surface. Finally, SEM analysis confirmed the formation of flagellar structures on D-serine exposed bacterial samples whereas the wild type alone was largely non-motile (Figure 4-5D).

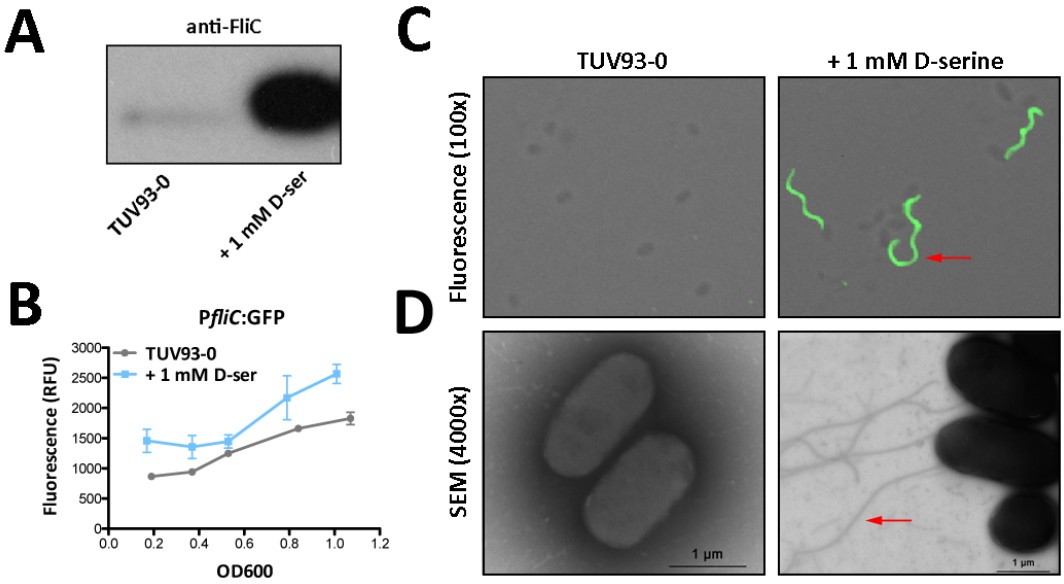

Figure 4-5 D-serine induces FliC over-expression in TUV93-0. (A) Western blot analysis of FliC expression from whole cell lysate of TUV93-0 with and without 1 mM D-serine cultured in MEM-HEPES. (B) Expression of *fliC* measured over time using a p*fliC*:GFP fusion reporter plasmid and plotted against OD_{600}. (C) Fluoresence microscopy (100x magnification) of corresponding TUV93-0 samples stained with anti-FliC primary and AlexaFluor 488 secondary antibodies. The red arrow indicates the appearance of an intact flagella (green). (D) SEM images of wild type TUV93-0 in MEM-HEPES. Supplementation of 1 mM D-serine results in the formation of flagella on the bacterial cell surface as indicated by a red arrow.

In these experiments D-serine clearly showed an impact on bacterial motility but the mechanism behind this was unclear. Due to the strong chemotactic properties of L-serine as an attractant to *E. coli* it was postulated that D-serine may also have chemotactic properties and may even act as a repellent due to its toxic nature. The chemotactic properties of D-serine were investigated using a number of chemotactic assays. Firstly, the chemical "in-plug" assay was used (Tso & Adler 1974). This experiment involved pouring 0.3% Tryptone broth agar containing an OD_{600} of 0.1 bacteria washed in chemotaxis medium around a series of solid 4% agar plugs containing high concentrations of the desired chemotactic substrate. The substrate diffuses into the surrounding soft agar in a gradient and allows visual assessment of chemo-repellence to the plug. Plugs containing a range of D-serine concentrations resulted in zones of clearing however this was inconclusive as a *ΔfliC* mutant strain showed the same phenotype as the wild type. An agar plug containing an excess of L-serine displayed no zone of clearance suggesting that the high concentration of D-serine was simply inhibitory to the bacteria irrespective of motility or not (Figure 4-6A). Indeed this method has been reported to be prone to false positive results and should not be used alone to assess chemotaxis (Li *et al.,* 2010). To further investigate this two other methods were employed. The chemical "in-capillary" method involved a thin capillary filled with a chemotactic substrate that is exposed to a culture of bacteria (Tso & Adler 1974). After an incubation period the capillary contents were serially diluted and plated on LB agar. A capillary containing L-serine retrieved ~2.8 x 10^7 CFU/ml whereas a capillary containing D-serine retrieved ~5 x 10^6 CFU/ml (Figure 4-6B). These experiments were performed in triplicate with similar results. A final experiment was performed using a method described by Deloney-Marino *et al. E. coli* forms concentric rings when migrating outwards on TB motility agar plates, the outer of which is representative of preferential L-serine consumption. When an excess of L-serine was spotted next to the migrating outer ring this resulted in an accumulation of bacteria around the source of excess nutrient (Deloney-marino *et al.,* 2003). A spot of excess D-serine however did not effect the migration of the bacteria (Figure 4-6C). Together these data are somewhat inconclusive but suggest D-serine does not have chemo-attractant properties. Indeed, a decrease in capillary uptake was observed for D-serine when compared with L-serine however this disagrees with a recent observation that D-serine may act as a weak chemo-attractant (Raterman & Welch 2013). This discrepancy may be due to a lack of replicates or may be indicative of specific differences between bacterial

strains occupying different niches. Further studies are required on this topic to resolve these uncertainties and were not explored further in this project.

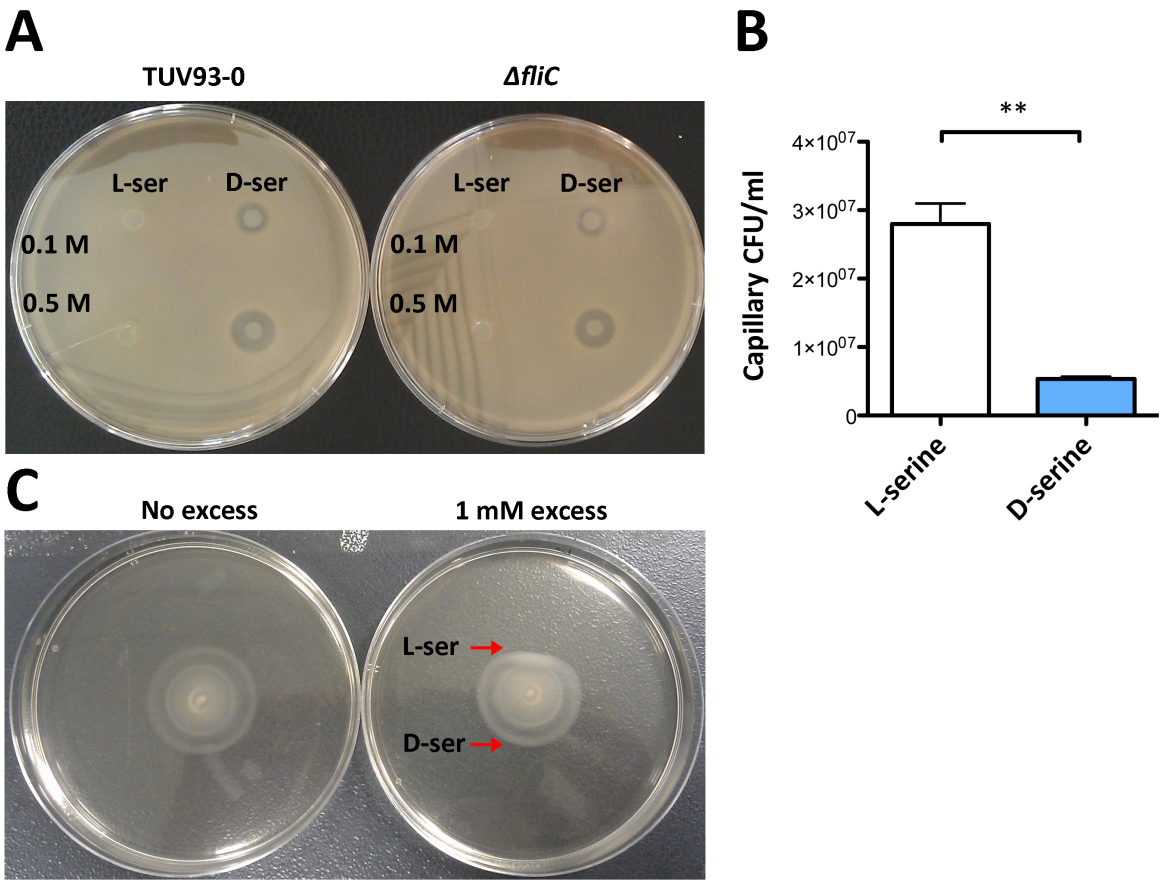

Figure 4-6 Chemotaxis of TUV93-0 towards D-serine. (A) Chemical "in-plug" analysis of L and D-serine chemotaxis. Solid 4% agar plugs containing 0.1 and 0.5 M serine excesses were submerged into a surrounding mixture of TUV93-0 or *ΔfliC* and motility soft agar. (B) Quantification of chemical "in-capillary" assays. The bars represent the number of CFU recovered from a 1 mm capillary containing an excess of L or D-serine. Experiment was performed in triplicate. ** denotes $p \leq 0.01$. (C) Serine migration assay of TUV93-0. A 1 mM excess of L or D-serine was spotted next to the migrating bacteria as indicated by the red arrows. Experiments were performed in triplicate with similar results.

4.3 The effects of D-serine on virulence of O157:H7

The data described above established that exposure to D-serine is not detrimental to *E. coli* O157:H7 when grown in an environment mimicking the gastrointestinal tract. However, it was apparent that TUV93-0 could respond to D-serine, as observed by the shift in motility. This observation was surprising as *E. coli* grown in MEM-HEPES are known to be non-motile (Beckham *et al.,* 2014, Tobe *et al.,* 2011) indicating that the presence of D-serine is causing a shift in the regulation of key cellular functions. The expression of flagella and the T3SS is thought to be a temporal process and the two factors are cross-regulated so as to allow efficient expression of each system under the correct conditions (Iyoda *et al.,* 2006). The flagella has adhesive properties but is usually expressed in the early stages of infection to allow niche determination and initial adherence, with later colonisation involving expression of the T3SS (Mahajan *et al.,* 2009). It is conceivable however that the system could be reversed in order to evade prolonged attachment to particular sites upon exposure to toxic compounds. D-AAs have recently gained interest in the literature when it was discovered that they play novel roles in temporally regulating key cellular processes such as cell wall remodelling (Cava *et al.,* 2011a, Lam *et al.,* 2009). It was therefore proposed that D-serine might have the ability to modulate expression of key virulence factors.

4.3.1 *D-serine inhibits the secretion of LEE encoded effectors*

The next aim was to investigate whether D-serine affected the ability of TUV93-0 to secrete LEE encoded effector proteins when grown in MEM-HEPES. D-serine has been implicated in niche specific virulence regulation for uropathogens but is unexplored in enteric pathogens (Anfora *et al.,* 2007, Korte-Berwanger *et al.,* 2013). Bacteria were cultured to an OD_{600} of 0.7 in MEM-HEPES with and without 1 mM D-serine to establish expression of the LEE. At this point the bacteria were harvested, supernatant separated from the cell pellet and secreted proteins precipitated using 10% TCA. Profiling of equal quantities of secreted proteins by SDS-PAGE showed a severe decrease in secreted proteins (Figure 4-7A). Western blotting confirmed the reduction in the LEE encoded effectors EspD and Tir (Figure 4-7B).

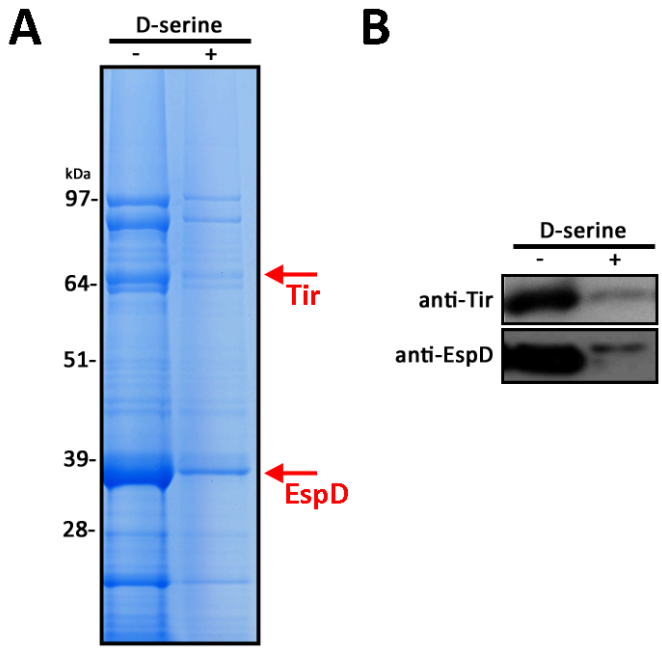

Figure 4-7 D-serine reduces T3S in TUV93-0. (A) Secreted protein profile of TUV93-0 with (+) and without (-) 1 mM D-serine supplemented into the growth medium MEM-HEPES. The molecular weight marker is indicated on the left. The positions of the major LEE encoded effectors EspD and Tir are indicated by a red arrow. (B) Western blot analysis of secreted protein fractions using anti-Tir and anti-EspD antibodies confirms the reduction in protein secretion of LEE encoded effectors.

4.3.2 *D-serine inhibits the LEE transcriptionally*

Having identified D-serine as an inhibitor of T3S the next obvious question was is this effect transcriptional or merely an inhibition of protein secretion? To investigate this a selection of transcriptional fusions were used. The promoter regions of LEE1, LEE2, LEE3 and *tir* were fused to GFP in the pAJR70 vector, which has been previously described (Roe *et al.,* 2003). This vector allows the expression of promoters fused to GFP to be monitored by measurement of green fluorescence over time. Exposure to 1 mM D-serine resulted in significant repression ($p \leq 0.001$) of all LEE promoters measured. As a control the expression of a p*rpsM*:GFP fusion was also measured and no decrease was observed over the wild type. This result suggests that repression of the LEE by D-serine targets the central regulation of the LEE. The master regulator Ler, the product of the first ORF located in LEE1, controls all subsequent LEE operons and expression of the GrlRA feedback loop (Mellies *et al.,* 2007).

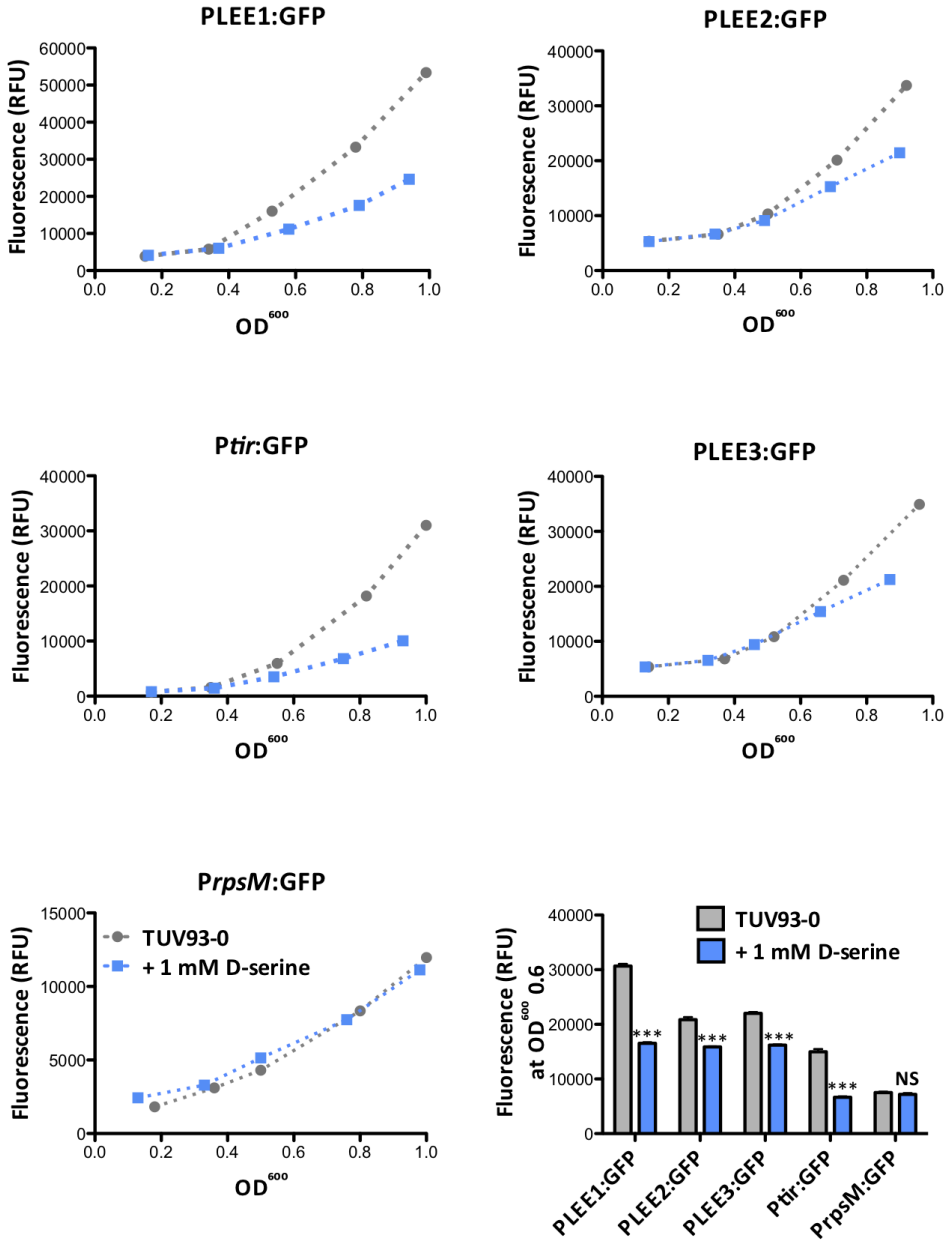

Figure 4-8 D-serine repression of the LEE acts at the transcriptional level. Fusions of the LEE1, LEE2, LEE3, *tir* and *rpsM* promoter regions to GFP were used to monitor the expression of each operon over time in MEM-HEPES with and without 1 mM D-serine. Data was plotted against OD_{600}. Additionally the data from OD_{600} 0.6 was used to calculate statistical significance at this point. *** and NS denote $p \leq 0.001$ and no significance respectively (students T-test). Graphs are representative of three replicates.

The concentration dependence of LEE repression by D-serine was assessed by measuring pLEE1:GFP expression over time in response to a concentration gradient of D-serine supplemented into the growth medium (1 mM, 0.5 mM, 0.25 mM, 0.1 mM and 0.05 mM). The effect was found to be most prominent at 1 mM and decreased accordingly along the

gradient. Plotting the fluorescence at an OD_{600} of 0.6 revealed that D-serine significantly reduced the expression of LEE1 in TUV93-0 at a concentration as low as 0.1 mM, which is > 4 fold lower than that required to inhibit growth in minimal medium (Cosloy & McFall 1973). This suggests that irrespective of inhibiting growth, D-serine can still cause significant repression of the T3SS major colonisation factor in O157:H7.

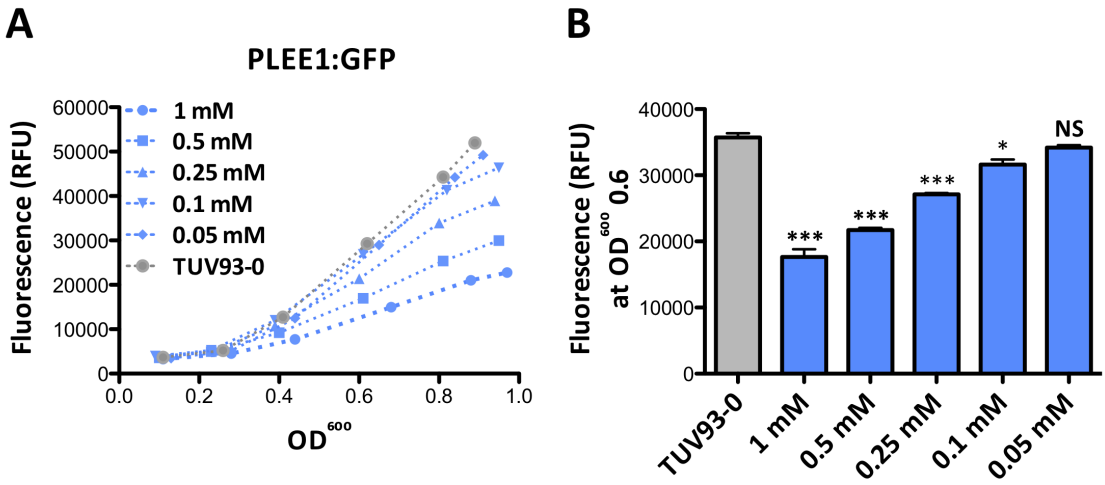

Figure 4-9 The negative effects of D-serine on the LEE1 promoter are concentration dependent. (A) Expression of the LEE1 promoter was measured over time and plotted against OD_{600} under a range of D-serine concentrations (1 mM, 0.5 mM, 0.25 mM, 0.1 mM and 0.05 mM). (B) Fluorescence at an OD_{600} of 0.6 was plotted and used to calculate statistical significance of the effects. ***, * and NS denote p ≤0.001, p ≤0.05 and no significance respectively (students T-test). Graphs are representative of three replicates.

Due to the increasing knowledge of the diverse roles that D-AAs play in bacterial processes it was important to assess if other D-AAs had significant effects on expression of the T3SS. A selection of D-AAs were tested for their ability to inhibit LEE1 expression similarly to D-serine. D-glutamate and D-alanine were chosen as they are known to be involved in cell wall biogenesis (Turner *et al.,* 2014, Typas *et al.,* 2012). D-glutamine, D-tyrosine and D-methionine were selected as they are structurally unique and have been recently identified as naturally occurring racemisation products of *V. cholerae* and *B. subtilis* (Lam *et al.,* 2009). Growth in MEM-HEPES supplemented with these 1 mM D-AAs was similar to the wild type in all cases except for D-glutamate, which was slightly slower (Figure 4-10A). D-Glutamate, D-Glutamine, D-Tyrosine and D-Methionine had no significant effect on LEE1 expression as seen for D-serine (Figure 4-10B). Exposure to D-

alanine however did result in a 29% reduction in LEE1 expression with no negative effects on *rpsM* control expression (Figure 4-10C). Although significant ($p \leq 0.01$) this repression was not as strong as for D-serine, which resulted in 47% reduction in LEE1 expression ($p \leq 0.001$) after exposure to 1 mM. Structurally D-serine and D-alanine are similar which could possibly account for the mimicked effect when high concentrations are present in the growth medium (Figure 4-10D). The effects of D-alanine were not studied further in this project, as the physiological significance of this finding was unclear.

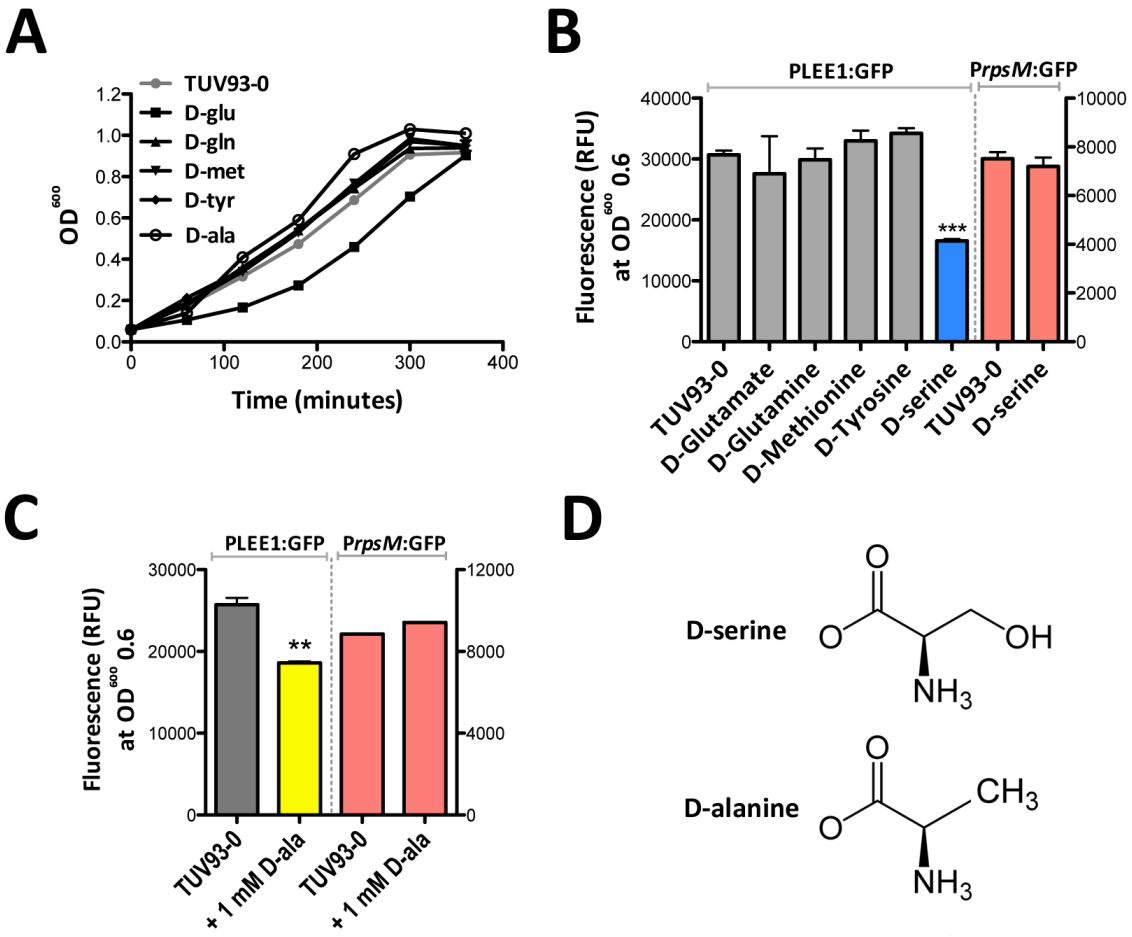

Figure 4-10 The effects of diverse D-AAs on LEE1 expression. (A) Growth of TUV93-0 in MEM-HEPES supplemented with 1 mM D-Glutamate, D-Glutamine, D-Tyrosine, D-alanine and D-Methionine measured as OD_{600} over time. The curves are representative of three biological replicates. (B) Transcriptional reporter assay monitoring LEE1 expression at an OD_{600} of 0.6 for various D-AAs supplemented at a concentration of 1 mM to the growth medium. The blue bar indicates data for D-serine. The expression of *rpsM* is highlighted for D-serine also as a control of housekeeping gene expression (red bars). *** denotes $p \leq 0.001$ (students T-test). (C) Transcriptional reporter assay monitoring LEE1 expression at an OD_{600} of 0.6 for D-alanine (yellow bar). *rpsM* expression is indicated also (red bars). ** denotes $p \leq 0.01$ (students T-test). (D) Chemical structures for D-serine and D-alanine.

4.3.3 *D-serine inhibits host cell binding of TUV93-0*

The data described above shows convincing evidence that exposure to D-serine at physiologically relevant conditions has severely detrimental effects on expression of the T3SS encoded on the LEE PAI. Expression of the LEE is concurrent with the formation of the A/E lesion phenotype on host cells colonised by LEE encoding pathogens (Elliott *et al.,* 1999). Therefore, the ability of TUV93-0 to colonise host cells via A/E lesions after exposure to D-serine was assessed. TUV93-0 was cultured in MEM-HEPES with and without 1 mM D-serine and was used to infect HeLa cells *in vitro.* Addition of D-serine to the growth medium during infection resulted in 77% fewer infected host cells relative to the wild type alone ($p \leq 0.001$). Furthermore, the numbers of bacteria adhering to successfully colonised host cells was significantly reduced by the presence of D-serine ($p \leq 0.01$). Finally, it was clear that even though a small number of bacteria managed to remain attached to host cells, the formation of A/E lesions as defined by areas of condensed actin underneath individual bacteria was greatly reduced (Figure 4-11). These data provide evidence that in a physiologically relevant infection model D-serine is capable of inhibiting bacterial adhesion to host cells and suggests a role for responding to D-serine encountered in the natural environment.

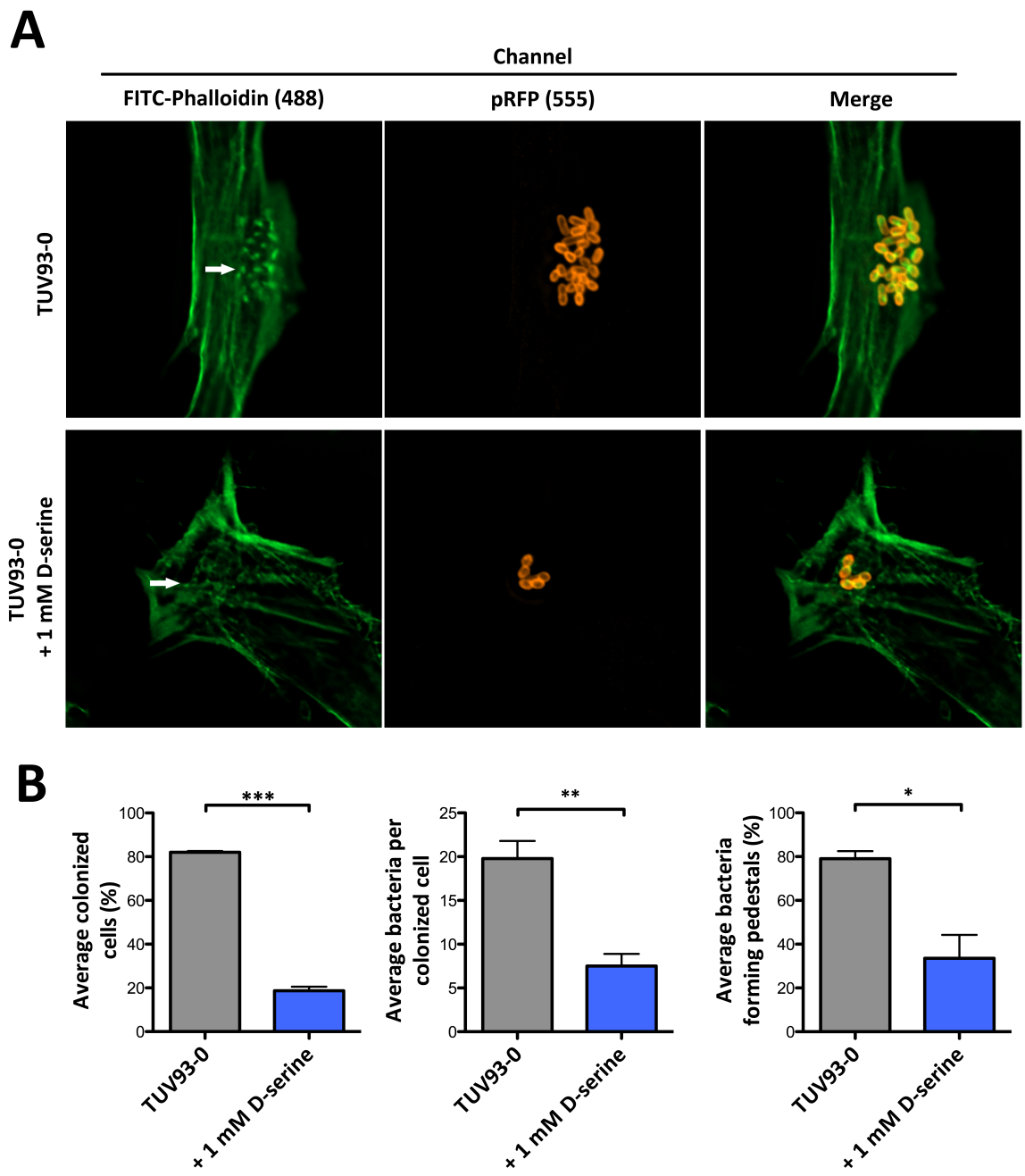

Figure 4-11 D-serine inhibits A/E lesion formation by TUV93-0 on host cells. (A) Representative fluorescence microscopy images of HeLa cells infected with TUV93-0 (with and without 1 mM D-serine). Host cells were stained with FITC-Phalloidin (green) and bacteria were carrying a constitutively expressed pRFP plasmid (red). Actin condensation corresponding to A/E lesion formation is indicated by a white arrow. (B) Histograms displaying the average percentage of colonised host cells, the average number of bacteria per host cell and the percentage of bacteria forming actin pedestals (A/E lesions) during cell-adhesion assays, with and without exposure to 1 mM D-serine. *, ** and *** denote $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$ respectively as calculated from three biological replicates (students T-test).

4.3.4 *D-serine induces diverse changes to the TUV93-0 transcriptome*

Exposure of TUV93-0 to D-serine resulted in striking phenotypic changes when cultured in MEM-HEPES. The LEE was repressed across operons LEE1 through LEE5 resulting in downregulation of the T3SS, reduced secretion of LEE encoded effectors and a severely impaired ability to colonise host cells. Additionally to becoming less virulent, the bacteria became more motile and these responses occurred without any negative effects on the ability to grow and divide. In order to identify global regulatory patterns of gene expression when exposed to D-serine, RNA-seq analysis was performed. Total RNA from three replicates of TUV93-0 alone in MEM-HEPES and two replicates of TUV93-0 supplemented with D-serine was purified at an OD_{600} of 0.6. The samples were enriched for mRNA, reverse transcribed into cDNA and sequenced at the University of Glasgow Polyomics facility. Raw data was aligned to the EDL933 reference genome, normalised and analysed for differential gene expression using the EdgeR package of Bioconductor (Robinson *et al.,* 2010).

Comparative transcriptomic analysis revealed 411 differentially expressed genes in the TUV93-0 transcriptome after exposure to D-serine. 188 genes were significantly upregulated with functions including cellular metabolism (42), cell envelope transport and biogenesis (32), various stress responses (25), nucleotide transport and metabolism (19), transcriptional regulators (16), amino acid transport and metabolism (14), phage association (7), post-transcriptional regulation (1) and 32 hypothetical genes of unknown function. Conversely, 222 genes were significantly downregulated in response to D-serine. This subset included genes involved in cellular metabolism (45), cell envelope transport and biogenesis (38), amino acid transport and metabolism (22), transcriptional regulation (14), various stress response (13), nucleotide transport and metabolism (8), colanic acid biosynthesis (8), adhesion (3), phage association (2), post-transcriptional regulation (1) and 41 hypothetical genes. As expected the LEE PAI was downregulated with 26 of the 41 ORFs displaying significant repression. Importantly, this repression was not restricted to a certain function encoded within the LEE and genes from each operon (LEE1 to LEE5) as well as non-operonic ORFs encoded within the LEE were repressed. Furthermore, one NLE gene (Z3921; *nleG6-3*) was also downregulated in the dataset. The conditions promoting expression of NLEs is less defined than that of the LEE but differential expression of NLEs has been identified previously in MEM-HEPES (Holmes *et*

al., 2012, Roe *et al.,* 2007). Differentially expressed genes not discussed in detail are highlighted in Table 8-2.

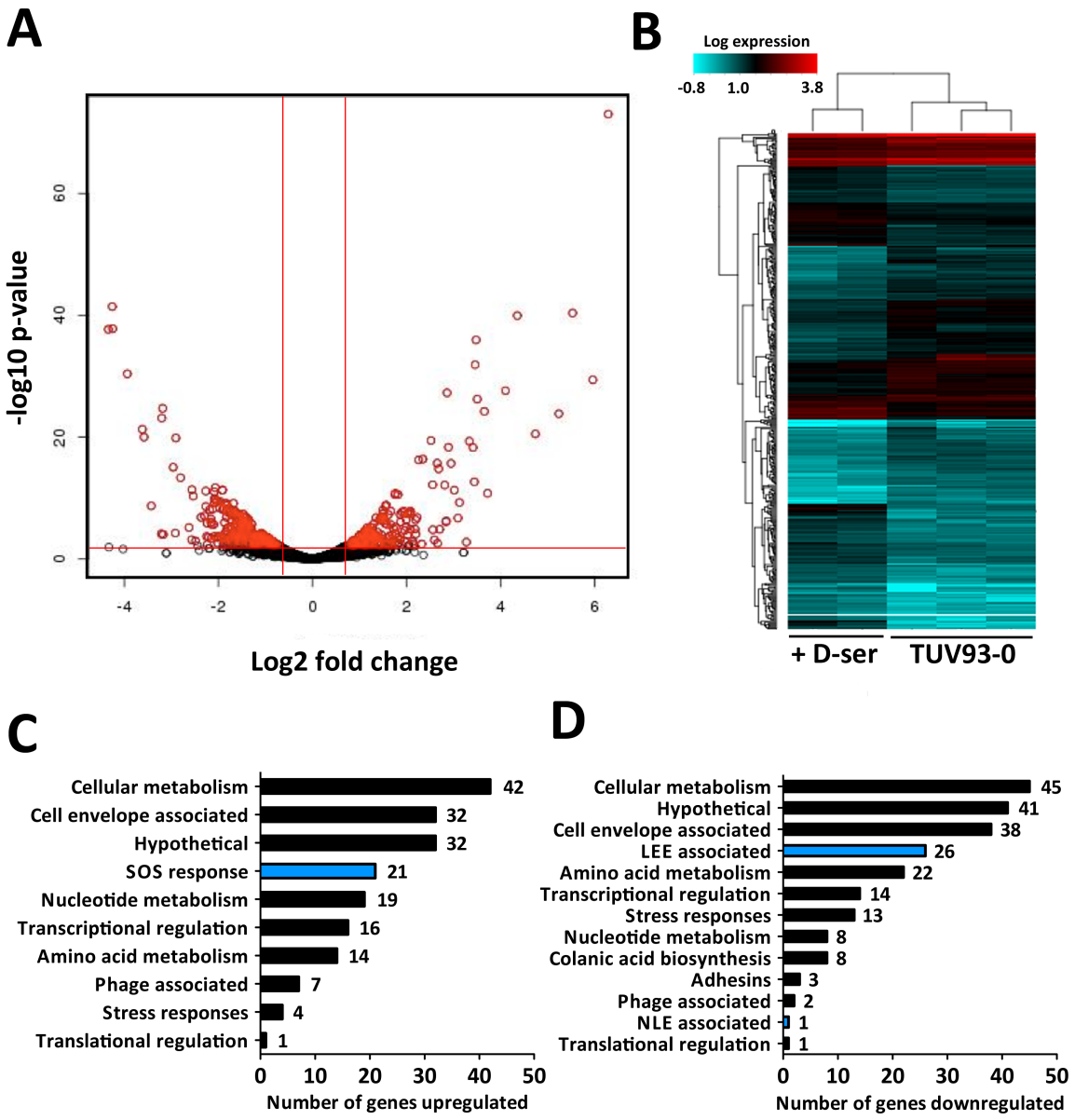

Figure 4-12 Global analysis of TUV93-0 exposure to D-serine by RNA-seq. (A) Volcano plot illustration of the gene expression profile of TUV93-0 treated with D-serine relative to the wild type alone. Significantly differentially expressed genes ($p \leq 0.05$) are indicated as red circles whereas genes not differentially expressed are indicated as black circles. The cut-offs for significant changes are outlined in red. (B) Heatmap of the 411 differentially expressed genes identified by RNA-seq illustrated by hierarchical clustering highlighting the expression patterns between TUV93-0 alone and treated with D-serine in the growth media (+ D-ser). Broad functional grouping of significantly upregulated (C) and downregulated (D) genes. The number of genes in each category is indicated adjacent to the corresponding bar. Groups of particular interest are highlighted in blue.

The negative effects on the LEE were striking and are summarised in Table 4-1. As hypothesised by the GFP-promoter transcriptional fusion assays, the effects are controlling the master regulation of the LEE. *ler* was not identified in the RNA-seq data but was confirmed to be significantly downregulated by qRT-PCR, validating the results seen in Figure 4-8. Indeed, examination of the raw reads would suggest that the entire LEE is downregulated after D-serine exposure although only 26 ORFs were found to be significant (Figure 4-13). This discrepancy may be due to technical aspects of the experimental design. Three wild type replicates were used but only two D-serine treated replicates were available for comparison. LEE expression for TUV93-0 grown in MEM-HEPES evidently can vary and a greater number of replicates would likely resolve this issue. The LEE4 and LEE5 operons are expressed to a higher extent with the lowest expression being observed in LEE1 to LEE3. RNA-seq data also requires a method of normalisation to account for differences in the total number of reads obtained by sequencing of the sample (for instance the three wild type replicates obtained 5.6, 4.9 and 4.7 million reads respectively whereas the D-serine treated samples obtained 4.5 and 4.3 million reads). Nonetheless, the repression of the LEE was significant and confirmed a shift in the TUV93-0 virulence gene expression profile in response to D-serine.

Table 4-1 Summary of the effects of D-serine on LEE expression as identified by RNA-seq

Gene name	Log^2 fold change	P-value	Function
escL	-2.242	1.28E-06	ATPase regulator
escR	-2.238	4.52E-03	Unknown
escP	-1.961	1.65E-04	Needle length regulator
orf4	-1.738	4.24E-06	Unknown
cesA2	-1.710	3.16E-08	Chaperone
cesAB	-1.681	6.40E-05	Chaperone
escA	-1.584	2.20E-05	Structural protein; chaperone
espF	-1.552	8.53E-08	Secreted effector protein
sepQ	-1.513	1.02E-06	C-ring component
escF	-1.475	9.49E-07	Needle protein
espB	-1.469	4.33E-08	Translocon
cesD5	-1.435	1.00E-07	Chaperone
espA	-1.396	1.83E-07	Needle filament
espG	-1.372	1.98E-05	Secreted effector protein
espD	-1.350	4.24E-07	Translocon
escE	-1.276	3.88E-03	Chaperone
cesF	-1.212	1.32E-04	Chaperone
escN	-1.106	8.52E-05	ATPase
escD	-1.073	2.09E-04	T3SS apparatus
sepL	-1.051	9.55E-05	Gatekeeper protein
escV	-1.044	1.89E-04	Cytoplasmic substrate docking
espH	-1.010	4.15E-04	Secreted effector protein
cesD	-0.982	1.76E-03	Chaperone
tir	-0.960	2.64E-04	Translocated intimin receptor
escC	-0.900	1.39E-03	Secretin
eae	-0.840	1.32E-03	Intimin

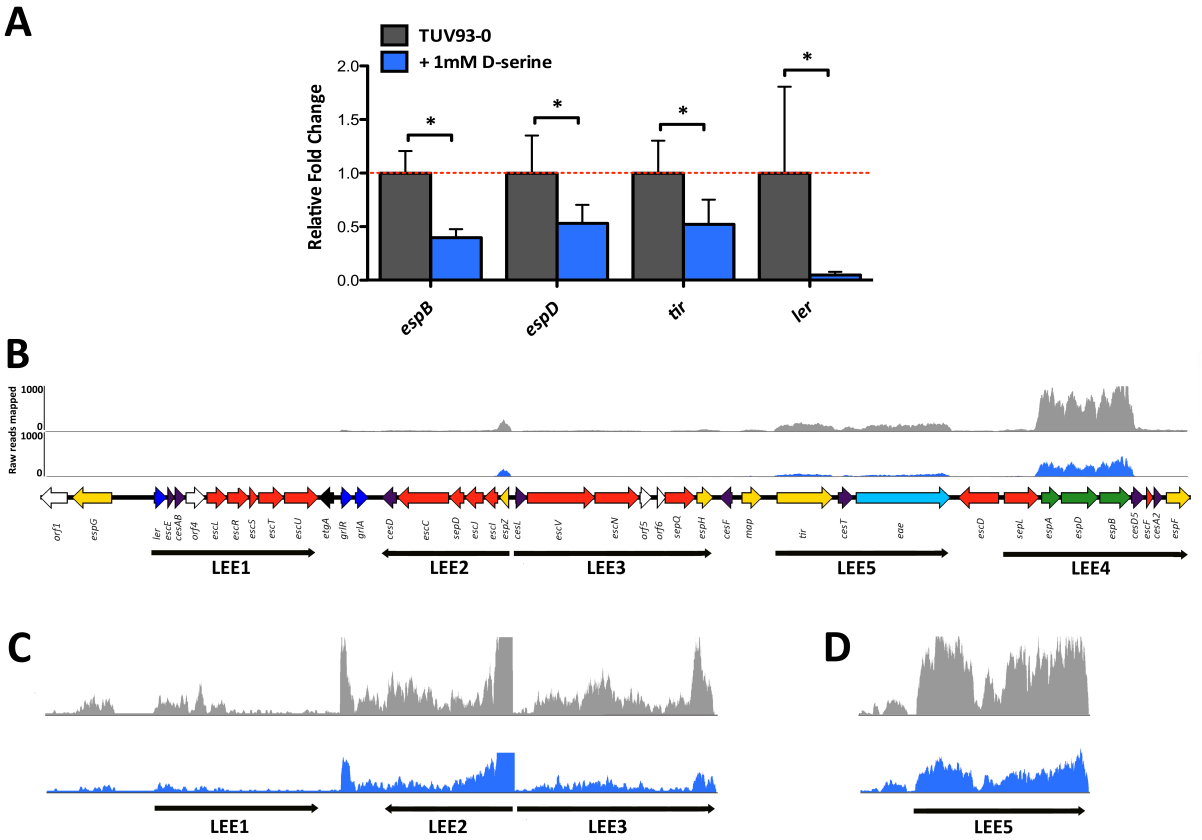

Figure 4-13 D-serine downregulates the LEE in TUV93-0. (A) qRT-PCR validation of RNA-seq data. The expression of *ler, tir, espD* and *espB* was determined for TUV93-0 cultured in MEM-HEPES supplemented with 1 mM D-serine relative to the untreated wild type. The red dotted line indicates baseline expression for the untreated control. * denotes $p \leq 0.05$. (B) Coverage graph illustrating the mapping of raw sequencing reads to the LEE PAI of EDL933. The grey peaks indicate the coverage of TUV93-0 and the blue peaks indicate coverage of the D-serine treated sample. Each graph has been scaled equally for comparison. Expansion of the LEE1 to LEE3 (C) and LEE5 (D) regions to demonstrate the differential expression of lower expressed ORFs within the LEE. Coverage graphs are representative of raw data from single replicates and are used only for visualisation purposes.

RNA-seq analysis revealed a diverse range of effects on cellular metabolism and nutrient transport induced by treatment with D-serine. The LEE PAI was downregulated as well as putative adhesins and members of the colanic acid biosynthetic pathway, which has been implicated in O157:H7 stress tolerance (Mao *et al.,* 2006). A total of 30 transcriptional regulators (16 upregulated, 14 downregulated) were significantly differentially regulated in the dataset, many of which were of putative function. Disrupting the normal regulation of a large number of transcription factors may explain the diverse gene expression profile observed. Individual analysis of candidate regulators from this dataset would be useful in exploring the regulatory networks of currently uncharacterised regulators. 38 genes

associated with various bacterial stress responses were differentially expressed in the dataset. These were involved in processes such as acid stress, osmotic stress and the phage shock response, the latter of which was interesting as it implies expression of the phage shock response under conditions that promote expression of the T3SS. The phage shock response is involved in responding to extracytoplasmic stresses that affect the inner membrane and thus the energy status of the cell (Darwin 2005). It has been proposed that mislocalisation of secretins involved in T3SS localisation can induce this response. It is conceivable therefore that growth in MEM-HEPES, which induces LEE expression, may therefore increase the risk of secretin mislocalisation due the abundance of the LEE encoded secretin EscC (Gauthier *et al.,* 2003). A downregulation of *escC,* as observed in this dataset, would reduce the amount of secretin available in the cell and thus reduce the phage shock response activated, although this explanation is speculative. The most striking pattern observed in the dataset in regards to stress responses was the activation of the SOS response. The SOS response is a network of genes that are usually activated in response to DNA damage although emerging mechanisms of SOS activation have been proposed in response to β-lactam antibiotics (Baharoglu & Mazel 2014, Maiques *et al.,* 2006, Michel 2005, Miller *et al.,* 2004). 21 members of the SOS regulon were significantly upregulated and summarised in Table 4-2, with some being amongst the highest fold changes observed (Fernández De Henestrosa *et al.,* 2000, Salgado *et al.,* 2013). The SOS response has also been described to play a role in regulation of the LEE in EPEC making this finding particularly intriguing (Mellies *et al.,* 2007b).

Table 4-2 Summary of the SOS response activation by D-serine as identified by RNA-seq

Gene name	Log^2 fold change	P-value	Function
dinD	4.107	2.35E-28	RecA filament disassembly
recN	3.460	1.36E-32	DNA repair ATPase
recX	3.128	5.24E-10	RecA filament extension inhibitor
recA	2.863	5.43E-28	LexA antirepressor
umuC	2.655	1.92E-16	DNA polymerase V subunit
umuD	2.559	6.65E-13	DNA polymerase V subunit
sulA	2.522	3.62E-20	SOS cell division inhibitor
lexA	2.260	5.70E-17	Global regulator (repressor) for SOS regulon
dinI	2.087	1.57E-08	RecA filament stabiliser
dinF	2.022	3.80E-08	Oxidative DNA damage protection
sbmC	1.950	1.50E-08	DNA gyrase inhibitor; SOS induced
dinP	1.863	4.81E-08	DNA polymerase IV
yafN	1.576	3.20E-03	Putative antitoxin
yebG	1.494	3.89E-08	DNA damage induced protein
uvrD	1.456	2.13E-07	DNA helicase II
uvrY	1.236	2.14E-05	Response regulator for BarA
uvrA	1.236	6.14E-06	Excision nuclease subunit A
dinG	1.215	8.54E-05	ATP-dependent DNA helicase
ssb	1.209	6.50E-05	Single-stranded DNA-binding protein
polB	1.137	2.73E-04	DNA polymerase II
uvrB	1.118	2.66E-05	Excision nuclease subunit B

Comparison of the D-serine dataset with the *ΔyhaO* dataset from chapter 3 revealed genes which were regulated in both instances making them candidates for genes directly regulated by D-serine or its associated sensing systems and not just downstream effects. Both datasets shared three upregulated genes among the 188 for D-serine and 29 for *ΔyhaO,* whereas 37 downregulated genes overlapped between the 222 and 85 genes identified in the D-serine and *ΔyhaO* datasets respectively Figure 4-14. Functions of overlapping genes that are not LEE or NLE encoded are summarised in Table 4-3 and included regulation of colanic acid biosynthesis, methionine metabolism and putative BFP-like adhesins. Interestingly, the methionine metabolic genes identified (*metF, metN* and *metK*) were downregulated in response to D-serine but upregulated in the *ΔyhaO* background when compared to the TUV93-0 transcriptome under the same conditions suggesting that D-serine may interfere with methionine biosynthesis.

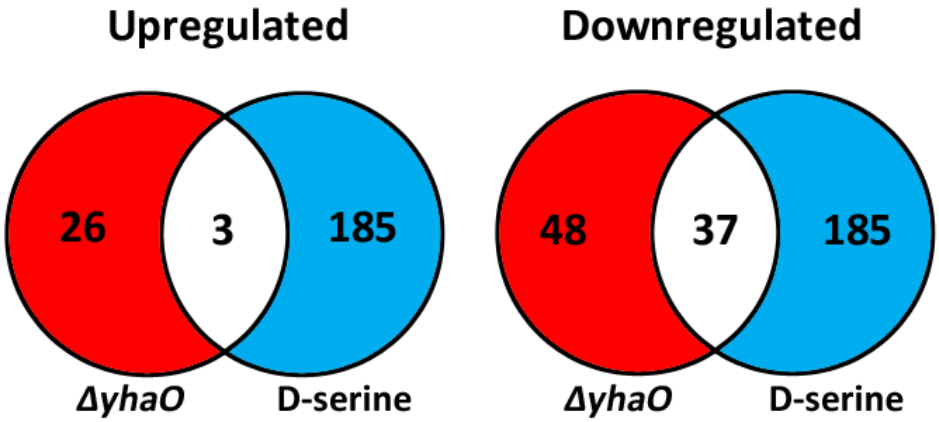

Figure 4-14 Overlap between the D-serine and *ΔyhaO* transcriptomes. Venn diagrams highlighting the number of upregulated and downregulated genes that overlap between the transcriptomes of TUV93-0 cultured in MEM-HEPES supplemented with D-serine and *ΔyhaO* grown in the same conditions. The number of genes is indicated with red corresponding to *ΔyhaO,* blue corresponding to D-serine and white corresponding to overlap between the two sets.

Table 4-3 RNA-seq identification of genes from the *ΔyhaO* regulon affected by D-serine

Gene	Log^2 fold change	P-value	Function
wcaG	-2.369	1.4E-07	NAD dependent epimerase/dehydratase family
rcsA	-2.184	8.7E-09	Positive response regulator for ctr capsule biosynthesis
nadB	2.073	4.3E-08	L-aspartate oxidase
Z0957	-1.847	4.2E-05	Hypothetical; shares sequence similarity to TfpB BFP in EPEC
gmd	-1.808	8.7E-06	NAD dependent epimerase/dehydratase family; colanic acid biosynthesis
Z0955	-1.728	5.2E-08	Hypothetical; IpxR homologue
metF	-1.614	3.0E-07	Methylenetetrahydrofolate reductase
proP	-1.406	1.2E-06	Proline/glycine betaine transporter (MFS transporter)
Z3931	-1.373	1.5E-03	Hypothetical; shares sequence similarity to TfpB BFP in EPEC
metN	-1.094	6.2E-04	Methionine transporter ATP-binding subunit
metK	-1.072	1.1E-04	Methionine adenosyltransferase
Z4386	0.955	7.3E-04	TonB dependent siderophore receptor
livJ	0.905	2.9E-03	ABC-type branched-chain amino acid (leucine/isoleucine/valine) transporter
rmf	-0.766	3.5E-03	Ribosome modulation factor

Together, the RNA-seq data described here reveal diverse effects of D-serine on TUV93-0, with the activation of the SOS response being particularly interesting. D-serine was not detrimental to growth under the conditions of this experiment indicating that the D-AA maintains the ability to regulate many cellular functions irrespective of the classical model of bacteriostasis. The SOS response was chosen for further investigation due to its strong pattern among the data observed and its proposed role in LEE regulation.

4.3.5 *Specific activation of the SOS response by D-serine in O157:H7*

Exposure to D-serine resulted in two clear transcriptional effects – downregulation of the LEE and activation of the SOS response. The classical scenario of SOS activation is in response to DNA damage. This activates the expression of a number of LexA repressed genes coding for DNA repair proteins that are controlled by RecA antirepression of the LexA repressor (Baharoglu & Mazel 2014, Michel 2005). The SOS response has also been implicated in regulation of the LEE and NLE ORFs in a LexA/RecA dependent manner although this study was carried out in EPEC (Mellies *et al.,* 2007b). It was therefore of interested to determine if the negative effects observed on the LEE from our RNA-seq data were mediated through activation of the SOS response. Specific activation of the SOS response by D-serine exposure was concurrent with LEE downregulation as analysed by RNA-seq. This was confirmed by both qRT-PCR and western blotting for the SOS antirepressor RecA on whole cell lysate of TUV93-0 cultured in MEM-HEPES with 1 mM D-serine (Figure 4-15A-C). Anti-GroEL antibody was used to verify equal amounts of protein loading per sample and mitomycin C was used as a positive control to induce the SOS response. Interestingly, the SOS cell division inhibitor SulA was transcriptionally upregulated without any negative effects on growth in MEM-HEPES as illustrated in Figure 4-1. As determined by transcriptional reporter assays, the negative effect on the LEE was largely D-serine specific. Western blot analysis for RecA confirmed that the same result could be said for the SOS response. Of the D-AAs tested for LEE expression, only D-serine induced excessive expression of RecA above wild type levels including D-alanine, which surprisingly resulted in a small repression of LEE expression (Figure 4-15D).

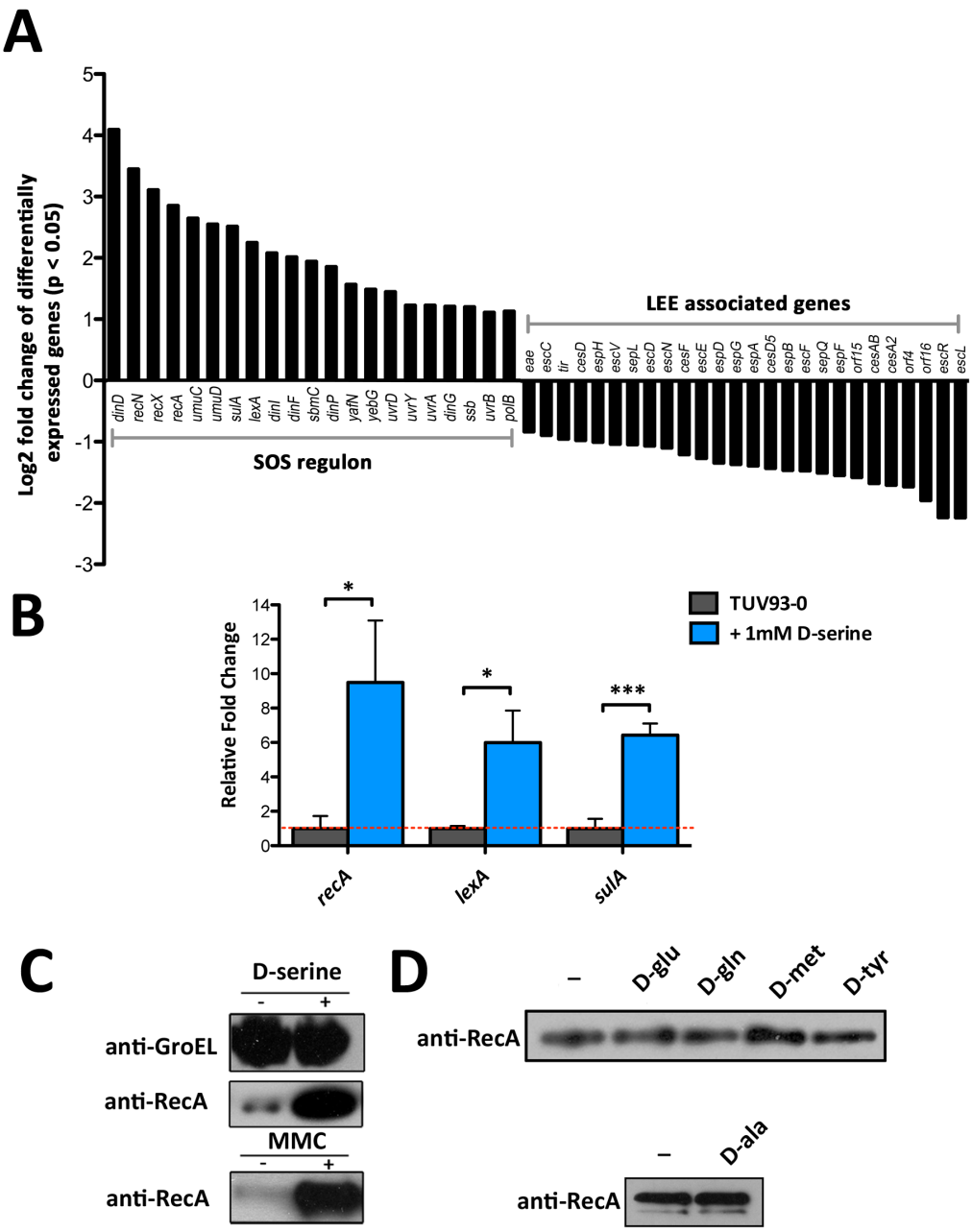

Figure 4-15 SOS induction by exposure to D-serine. (A) Concurrent induction of the SOS response and repression of the LEE as identified by RNA-seq analysis. 21 members of the SOS regulon and 26 LEE encoded ORFs from operons LEE1 through LEE5 were significantly ($p \leq 0.05$) differentially expressed. Bars indicate Log^2 expression values for TUV93-0 cultured in MEM-HEPES supplemented with 1 mM D-serine in comparison with untreated TUV93-0. (B) Relative expression of *recA, lexA* and *sulA* was analysed by qRT-PCR. The red dotted line indicates baseline expression in the TUV93-0 untreated control. * and *** denotes $p \leq 0.05$ and $p \leq 0.001$ respectively. (C) Western blot analysis to confirm the induction of the SOS response by detection of RecA levels from whole cell lysate. A + or – indicates the presence of 1 mM D-serine or not. Detection of GroEL was used as a loading control. Supplementation of MEM-HEPES with 5 µM mitomycin C was used as a positive control for SOS induction. (D) Western blot analysis for RecA expression in MEM-HEPES supplemented with 1 mM D-glutamate, D-glutamine, D-methionine, D-tyrosine and D-alanine in comparison to background RecA expression levels of TUV93-0 alone.

Seeing as the negative effects of D-serine on LEE1 expression were found to be concentration dependent the same approach was taken to assess if varying concentrations of D-serine could induce the SOS response to differing extents. MEM-HEPES was supplemented with D-serine at concentrations of 1 mM, 0.5 mM, 0.2 mM, 0.1 mM and 0.05 mM and whole cell lysate from TUV93-0 cultured in each concentration was analysed by western blot using anti-RecA antibody. RecA expression was observed at increased levels from the wild type over the entire gradient of D-serine with the strongest band at the 1 mM concentration (Figure 4-16A). Furthermore, growth in M9 minimal media supplemented with the concentration gradient of D-serine showed a recovery in growth capability as the concentration of D-serine was reduced with growth at 0.05 mM being indistinguishable from the wild type (Figure 4-16B). This suggests a point in the concentration gradient where TUV93-0 can tolerate D-serine exposure. Interestingly, this is the same point at which LEE expression was no longer significantly affected in MEM-HEPES as seen in Figure 4-9.

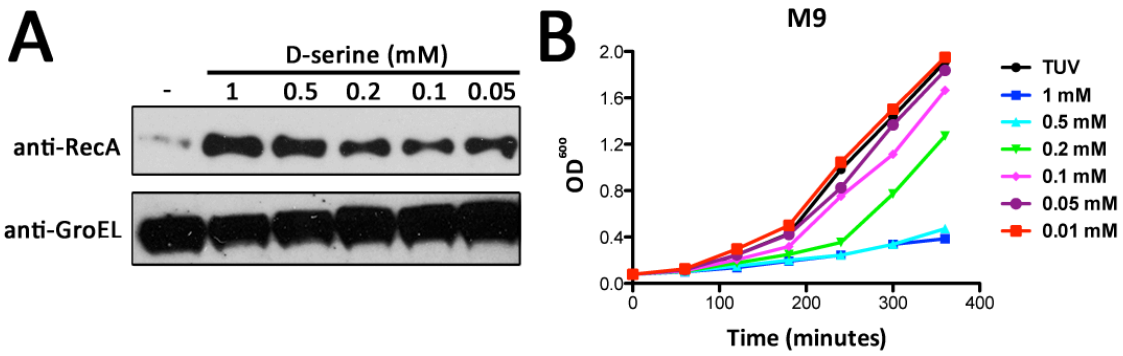

Figure 4-16 Activation of the SOS response by D-serine is concentration dependent. (A) Western blot analysis for RecA expression over a 1 mM, 0.5 mM, 0.2 mM, 0.1 mM and 0.05 mM concentration gradient of D-serine. GroEL is indicated to represent a loading control from the same samples. (B) Concentration dependent inhibition of TUV93-0 growth in M9 supplemented with D-serine at the concentrations indicated in panel A. Growth curves are indicative of three biological replicates.

4.3.6 *SOS induction is dependent on a non-functional DsdA*

Activation of the SOS response was clearly occurring in the presence of D-serine in concentrations that also repress the LEE PAI. It was hypothesised that this response may be due to intracellular accumulation of D-serine in O157:H7. In chapter 3 it was

demonstrated that the D-serine sensing transporter YhaO is capable of transporting D-serine even though O157:H7 cannot metabolise it (Figure 3-4) and that O157:H7 complemented with DsdA from CFT073 can in fact grow on D-serine as a carbon source (Figure 3-6). Indeed, it was shown previously that accumulation of D-serine in a CFT073 *ΔdsdA* mutant resulted in hypervirulence (Anfora *et al.,* 2007). By extracting bacterial metabolites of a culture grown in MEM-HEPES with and without 1 mM D-serine, analysing the intracellular metabolite pools in a chiral-specific manner (performed in collaboration with Dr. Karl Burgess) and comparing them to controls of known amounts, the relative concentrations of D-serine were determined (Figure 4-17A). As expected, TUV93-0 accumulated intracellular D-serine whereas the UPEC CFT073 strain did not, consistent with its ability to metabolise the D-AA. The CFT073 *ΔdsdA* mutant however accumulated D-serine comparably to TUV93-0. Complementation of *ΔdsdA* with a plasmid borne copy of the gene (p*dsdA*) resulted in an ~7 fold reduction in D-serine accumulation. Unsurprisingly, complementation of TUV93-0 with p*dsdA* (TUVp*dsdA*) resulted in an ~7 fold reduction also (Figure 4-17B). In agreement with this reduction of D-serine accumulation, growth of TUVp*dsdA* on MOPS minimal agar with D-serine as a sole carbon source was observed but abolished in CFT073 *ΔdsdA* (Figure 3-6). Once again, complementation with p*dsdA* reversed this effect and these results were confirmed by growth in MEM-HEPES supplemented with a concentration gradient of D-serine (Figure 4-17C-D). CFT073 Δ*dsdA* was susceptible to D-serine toxicity in MEM-HEPES but this was reversed when complemented with p*dsdA.* These results confirmed that intracellular accumulation of D-serine was entirely dependent on a functional DsdA. In addition to growth, the levels of RecA in *ΔdsdA* and *ΔdsdA*p*dsdA* were analysed by western blot. Strikingly, RecA expression was observed in *ΔdsdA* in a concentration dependent manner, similarly to TUV93-0 (Figure 4-16) but *ΔdsdA*p*dsdA* had levels of RecA similar to the untreated control consistently across all concentrations of D-serine (Figure 4-17E). Together, these data indicate that activation of the SOS response is concurrent with intracellular accumulation of D-serine and that this is in turn dependent on a non-functional DsdA.

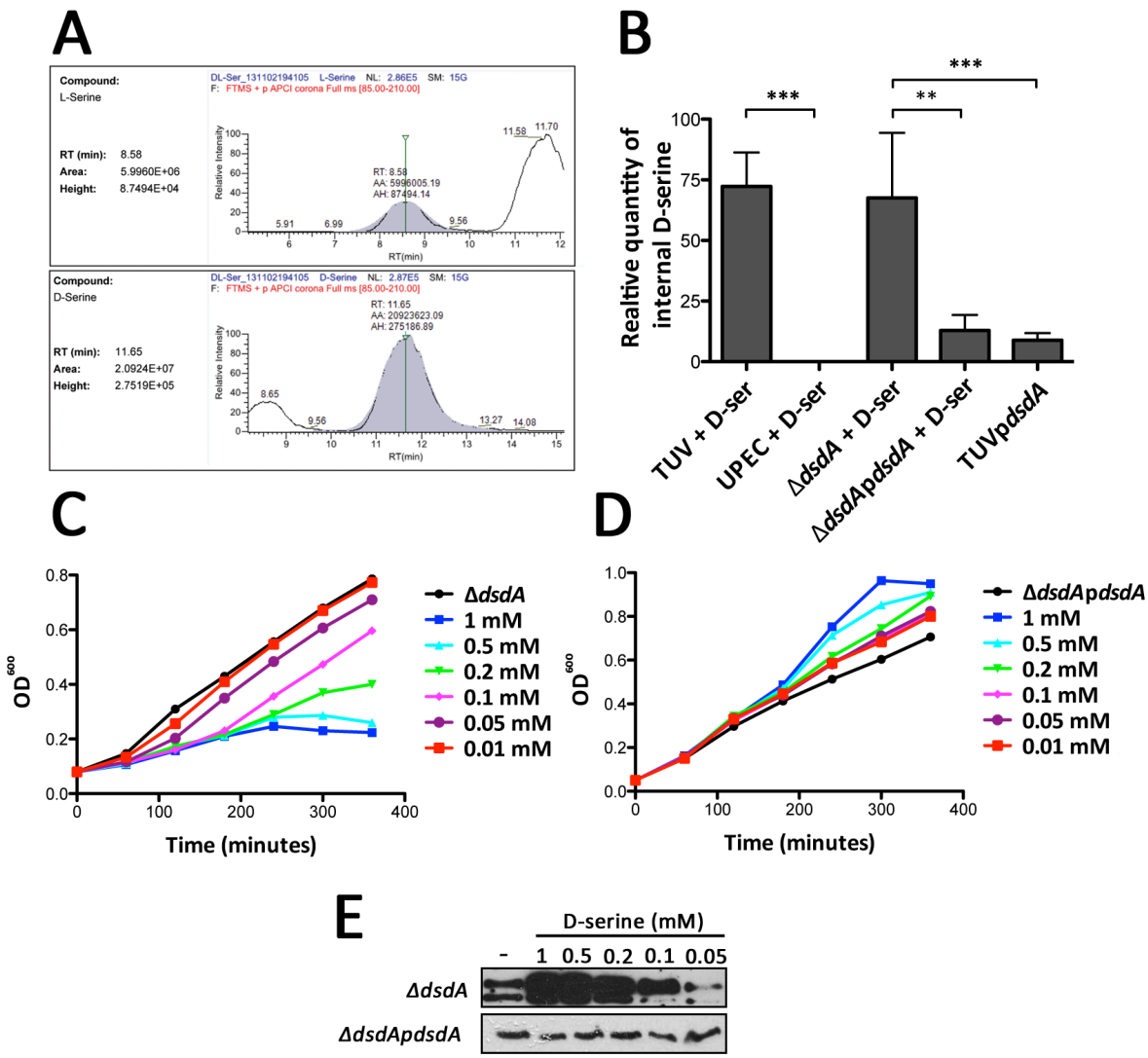

Figure 4-17 The SOS response is concurrent with D-serine intracellular accumulation. (A) Chromatographic peaks of standard L-serine and D-serine showing differential retention times as measured by LC-MS. This standard allows for relative quantification of D-serine from whole cell metabolite pools. (B) Relative quantity (%) of D-serine obtained from whole cell metabolite pools of TUV93-0, UPEC CFT073, UPEC Δ*dsdA,* UPEC Δ*dsdA*p*dsdA* and TUVp*dsdA* cultured in MEM-HEPES supplemented with 1 mM D-serine. ** and *** denote $p \leq 0.01$ and $p \leq 0.001$ respectively calculated from three biological replicates (students T-test). (C) Growth of CFT073 Δ*dsdA* and Δ*dsdA*p*dsdA* (D) in MEM-HEPES supplemented with D-serine at concentrations 1 mM, 0.5 mM, 0.2 mM, 0.1 mM and 0.05 mM. Graphs are representative of three biological replicates. (E) Western blot analysis of RecA expression in Δ*dsdA*p*dsdA* and TUVp*dsdA* exposed to this D-serine concentration gradient. Experiments were performed in triplicate.

4.3.7 *LEE repression by D-serine is independent of the SOS response*

The SOS response has been proposed to regulate the LEE in EPEC with identification of a LexA binding site at the LEE2/3 promoter (Mellies *et al.,* 2007b). It was therefore proposed that the SOS response induced by exposure to D-serine might be the mediator of the concurrent LEE downregulation observed under the same conditions and that the acquisition of a functional DsdA might overcome this effect. To test this the secreted protein profile was analysed for TUVp*dsdA* cultured in MEM-HEPES with and without 1 mM D-serine to an OD_{600} of 0.7 as well as western blot analysis of the whole cell lysate for RecA expression and compared to that of the wild type under the same conditions (Figure 4-18A-B). As expected, the SOS response as indicated by RecA expression was not detectable in the TUVp*dsdA* background when exposed to D-serine. However, there was no discernable difference in secreted protein profiles between the two scenarios and this was confirmed by western blotting for EspD, which showed a reduced secretion in TUVp*dsdA* similar to TUV93-0 after exposure to D-serine. These data were confirmed by qRT-PCR analysis of *espD, ler* and *recA* expression in either background (Figure 4-18C). Both *espD* and *ler* were significantly downregulated ($p \leq 0.05$) in response to D-serine but *recA* levels were comparable to that of the wild type showing no significant difference. This result was quite surprising and suggests that D-serine represses the LEE and activates the SOS response by two independent mechanisms.

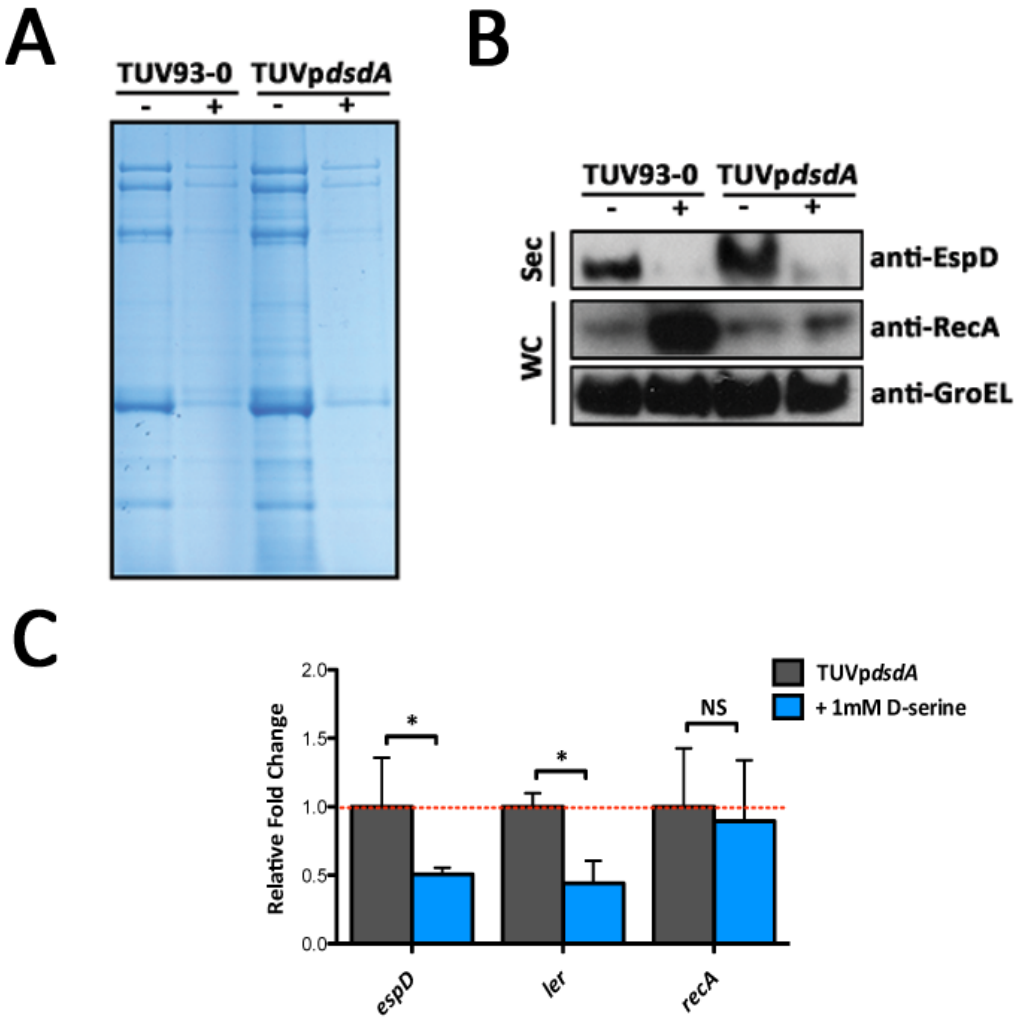

Figure 4-18 SOS-independent repression of the LEE by D-serine. (A) Secreted protein profile of TUV93-0 and TUV*pdsdA* grown in MEM-HEPES with (+) and without (-) 1 mM D-serine. (B) Western blot analysis of EspD from the secreted protein profile and RecA from the corresponding whole cell lysate. GroEL was also detected in the whole cell lysate to act as a loading control. (C) Validation of panels A and B by qRT-PCR. Relative expression of *espD, ler* and *recA* was analysed in the TUV*pdsdA* background grown in MEM-HEPES with and without 1 mM D-serine. The red dotted line indicates baseline expression of each gene in the untreated control. * and NS denote $p \leq 0.05$ and no significance respectively. All experiments were performed in biological triplicate.

This result indicated that the ability to metabolise D-serine was redundant in terms of LEE regulation but absolutely required for tolerance of D-serine induced stress. Inhibition of the LEE required an exogenous concentration of ≥ 100µM D-serine whereas SOS activation was evident at a concentration of 50 µM. D-serine is abundant in the urinary tract (> 1 mM) but its abundance in the gastrointestinal tract is unknown and likely to be affected by diet and the surrounding microbiota. O157:H7 successfully colonise the gastrointestinal tract suggesting D-serine may not be a major constituent but these

bacteria can clearly respond to the D-AA irrespective of the ability to metabolise it. The relative concentrations of serine in colon samples of BALB/c mice (in collaboration with Dr Gillian Douce and Dr Karl Burgess) were therefore measured. L-serine was present at a concentration of ~125 µM but D-serine was only detectable in the range of ~1 µM, some 100 times lower than that required to inhibit LEE expression. This indicates that the niche of O157:H7 does not contain inhibitory concentrations of D-serine, allowing successful colonisation. These bacteria maintain the ability to respond to D-serine however, and use this signal to regulate expression of colonisation factors in the gastrointestinal tract.

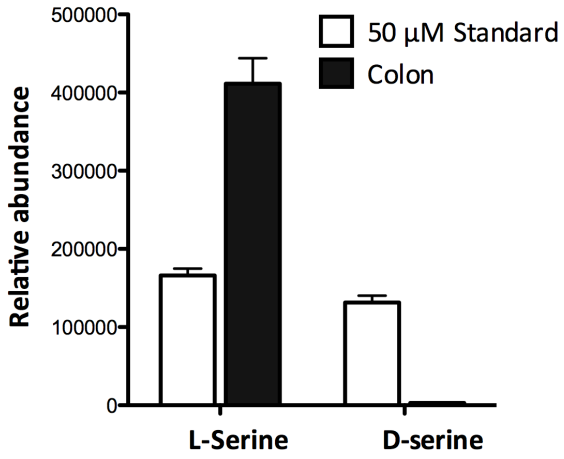

Figure 4-19 Relative quantification of serine levels in the colon. Colon tissue from five BALB/c mice were used to extract available metabolites. The samples were analysed by HPLC for L and D-serine abundance relative to 50 µM standards. L-serine was present at an abundance of ~125 µM whereas D-serine was detectable at a concentration of ~1 µM.

4.3.8 *D-serine modulates known regulators of the LEE*

The discovery that LEE repression and SOS induction in response to D-serine were mechanistically independent meant that this effect on virulence was controlled in an unknown way. LEE regulation is a complex, multi-layered process with numerous regulatory systems co-ordinating to control the expression of the T3SS and associated effectors (described in detail in section 1.4). The RNA-seq data was therefore re-examined to identify if any known regulators of the LEE were differentially expressed in response to D-serine. Three transcriptional regulators from two independent regulatory systems of the LEE were affected in response to D-serine suggesting a potentially novel mechanism of LEE regulation involving these two systems (Table 4-4). The IHF alpha

subunit *ihfA* was significantly downregulated whereas two GAD regulators, *yhiF* and *gadX,* were both significantly upregulated, as summarised in table. Previous work has shown that IHF binds the LEE1 promoter and is required for LEE expression (Friedberg *et al.,* 1999). YhiF is a member of the LuxR type transcriptional regulators and has been shown to negatively regulate the LEE2/3 and LEE5 promoters while being thought to regulate acid stress (Tatsuno *et al.,* 2003, Tree *et al.,* 2011). GadX is an AraC-like transcriptional regulator also involved in regulation of acid resistance but has recently been identified as an indirect LEE1 repressor (Branchu *et al.,* 2014). The differential expression of these regulators was verified by qRT-PCR analysis (Figure 4-20).

Table 4-4 Transcriptional regulators of the LEE differentially expressed by D-serine

Gene	Log^2 fold change	P-value	Description
ihfA	-1.285	1.83E-06	IHF alpha subunit
yhiF	0.866	3.00E-02	LuxR-like transcriptional regulator of acid resistance
gadX	1.348	1.72E-04	AraC-like transcriptional regulator of acid resistance

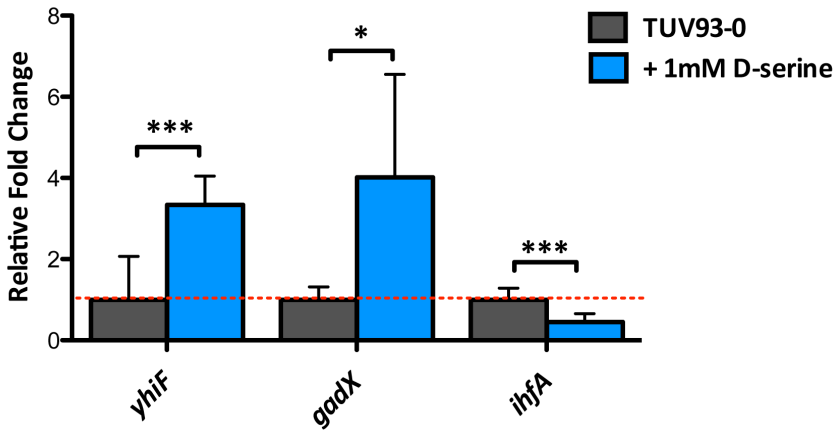

Figure 4-20 Validation of *ihfA, yhiF* and *gadX* expression levels in response to D-serine. qRT-PCR analysis of relative expression levels for *ihfA, yhiF* and *gadX* in TUV93-0 response to D-serine. The red dotted line indicates baseline expression levels in the TUV93-0 wild type alone. * and *** denote $p \leq 0.05$ and $p \leq 0.001$ respectively as calculated from three biological replicates.

The next task was to evaluate if these regulators contribute to the D-serine phenotype. It was hypothesised that *ΔyhiF* and *ΔgadX* mutants would be less sensitive to D-serine

repression as these regulators are known to negatively effect the LEE. Similarly, as D-serine exposure downregulated *ihfA,* it was predicted that constitutive complementation of *ihfA* might similarly reduce the negative effects observed. To test this the secreted protein profile of TUV93-0 was compared to each strain. *ΔyhiF* was generated elsewhere and supplied as a gift for this study (Tree *et al.,* 2011). *ΔgadX* was generated by the Lambda Red method as part of this project (Figure 4-21A) and p*ihfA* (complementation plasmid with the sequence for *ihfA* driven from the constitutively expressed *rpsM* promoter) was synthesised commercially by Genscript (Datsenko & Wanner 2000). The secreted protein profile for *ΔgadX* grown in MEM-HEPES with and without 1 mM D-serine was not observably different from that of TUV93-0 and was not carried further in this particular experiment (Figure 4-21B). Conversely, *ΔyhiF* was markedly less sensitive to D-serine repression with secreted protein levels comparable to that of TUV93-0. Complementation of *ΔyhiF* with p*yhiF* restored the ability of D-serine to reduce the secreted protein profile suggesting that YhiF is involved in regulating the response. Similarly, constitutive expression of IhfA in the TUV93-0 background also reversed the sensitivity to D-serine. All secreted protein profiles were verified by western blot analysis for Tir and EspD (Figure 4-21C). Together, these results suggest that both YhiF and IhfA are likely involved in mediating D-serine repression of the LEE through a previously unexplored mechanism.

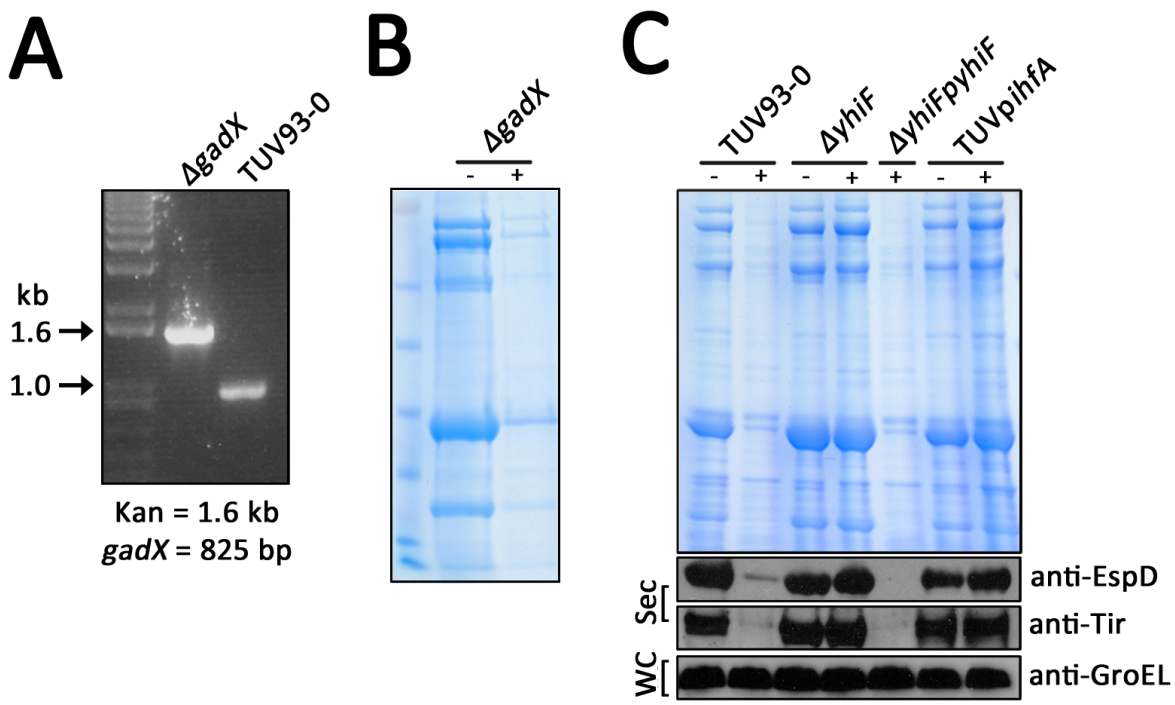

Figure 4-21 T3S regulation by *gadX*, *yhiF* and *ihfA* in response to D-serine. (A) PCR confirming the deletion of *gadX* by Lambda Red mediated mutagenesis. The TUV93-0 band (825 bp) corresponds to the *gadX* gene product whereas the *ΔgadX* band (1.6 kb) corresponds the Kan^R cassette replaced during mutagenesis. (B) Secreted protein profile screen of *ΔgadX* grown in MEM-HEPES (-) or supplemented with 1 mM D-serine (+). (C) Secreted protein profile screening of TUV93-0, *ΔyhiF, ΔyhiF*p*yhiF* and TUVp*ihfA* grown in MEM-HEPES (-) or supplemented with 1 mM D-serine (+). The SDS-PAGE results were confirmed by western blot analysis of Tir and EspD from the secreted protein fraction (Sec). GroEL is indicated as a loading control for corresponding whole cell (WC) fractions obtained at an OD_{600} of 0.7. Experiments were performed in triplicate.

At this point it was clear that YhiF and IhfA contributed to the D-serine phenotype but the mechanism was not apparent. To address if the regulators were acting via independent pathways or if they were cross regulated in some way qRT-PCR analysis was performed to investigate the expression of each regulator in the corresponding deletion or complementation background. Analysis of relative expression levels in the Δ*yhiF* background revealed that *ihfA* was no longer downregulated in response to D-serine resulting in normal LEE expression as observed in Figure 4-21. Expression levels of *espD* and *ler* still displayed a lower trend (most likely owing to transcription factors from other regulons that are actively regulating gene expression under these conditions) but this was no longer significant (Figure 4-22A), as was seen in TUV93-0 treated with D-serine (Figure 4-13). Similarly, in the TUVp*ihfA* background, *yhiF* levels were comparable with that of the wild type whereas *ihfA* levels were restored. Again, *espD* and *ler* were no longer

significantly repressed in this background when exposed to D-serine (Figure 4-22B). These data provide convincing evidence that both YhiF and IhfA work in tandem to contribute to the D-serine phenotype.

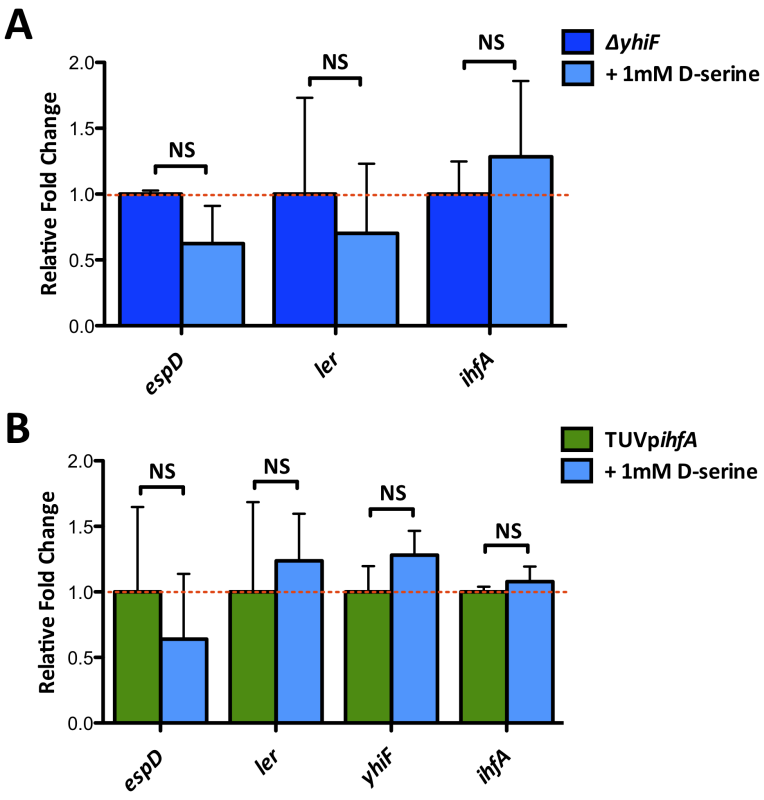

Figure 4-22 YhiF and IhfA co-ordinately regulate the LEE in response to D-serine. (A) qRT-PCR analysis of *espD, ler* and *ihfA* expression levels in the Δ*yhiF* background (blue) grown in MEM-HEPES with 1 mM D-serine (light blue) relative to Δ*yhiF* alone in MEM-HEPES. (B) qRT-PCR analysis of *espD, ler, yhiF* and *ihfA* levels in the TUV*pihfA* background (green) grown in MEM-HEPES with 1 mM D-serine relative to TUV*pihfA* alone in MEM-HEPES. NS denotes no significant difference in gene expression. The red dotted line indicates baseline expression levels of the untreated controls in each case. All experiments were performed in biological triplicate.

To determine whether IhfA and YhiF were directly regulating each other's transcription EMSA assays were performed using purified forms of these regulators. The IHF consensus binding site is well established and has been described in the literature as WATCARXXXXTTR, where W is A or T; X is A, T, C or G; R is A or G (Huo *et al.,* 2009). IHF is known to bind the LEE1 promoter and is capable of autoregulation (Aviv *et al.,* 1994, Friedberg *et al.,* 1999). The upstream DNA sequence of *yhiF* contains two IHF consensus-binding sites. *yhiF* has been established to be cotranscribed with adjacent gene *slp* but has more recently been demonstrated to be upregulated for LEE regulation without a

significant transcriptional effect on *slp* (Tree *et al.,* 2011, Tucker *et al.,* 2003). The first IHF site lies upstream of the *slp* start sequence and the second resides upstream of the *yhiF* start sequence but within the *slp* coding region. Due to the fact that *slp* was not differentially regulated in the RNA-seq data set it was proposed that IhfA might be controlling *yhiF* transcription from this second site. Indeed, a recent study on *E. coli* K-12 operon structure revealed that inter-operonic differential expression is widespread and likely conditions or input dependent (Conway *et al.,* 2014). Purified IHF and IhfA was provided by Professor Steven Goodman for use in this experiment. IHF binding was confirmed in the LEE1 and *ihfA* promoter regions but no binding was observed for the putative *yhiF* binding site. IHF is a heterodimeric complex of the IhfA and IhfB subunits. IhfA homodimers have been demonstrated in the literature to be capable of DNA binding but the conditions under which this occurs are not so well defined (Mangan *et al.,* 2006, Zulianello *et al.,* 1994). Accordingly IhfA homodimer binding was not observed to any of the promoter regions tested in this study suggesting that downregulation of *ihfA* alone is enough to affect levels of full IHF in the cell (Figure 4-24A). Addition of 1 mM D-serine to the binding reaction did not result in inhibition of IHF binding to the *ler* or *ihfA* promoter regions nor did it improve binding if IhfA to any of the sequences tested indicating that D-serine is likely not a direct signal for IHF or IhfA (Figure 4-24B).

LEE1

CTTGCCGCCTTCCTGTAACTCGAATTAAGTAGAGTATAGTGAAACGGTTCAGCTTGGTTT
TTATTCTGTTTTATTTGTTTATGCAATGAGATCTATCTTATAAAGAGAAACGCTTAACTA
AATGGAAATGCAATTATTAAAGTCGTTTGTTAACGAGATGATTTTCTTCTATATCATTGA
TTTTAAATGGATTTTAAAAATATATGATTTTTTTGTTGACATTTAATGATAATGTATTTTA
CACATTAGAAAAAAGAGAATAATAACATTTTAAGGTGGTTGTTTGATGAAATAGATGTG
TCCTAATTTGATAGATAAACGTTATCTCACATAATTTATATCATTTGATTAATTGTTGGT
CCTTCCTGATAAGGTCGCTAATAGCTTAAAATATTAAAGC**ATG**

ihfA

GTTCATCCTGAACTGGAACGTAAACTGGATCTTAACGGTCGCACTCTGGTGTTCGAACTG
GAGTGGAACAAGCTCGCAGACCGCGTGGTGCCTCAGGCGCGCGAGATTTCTCGCTTCCCG
GCGAACCGTCGTGACATCGCGGTGGTGGTCGCAGAAAACGTTCCCGCAGCGGATATTTTA
TCCGAATGTAAGAAAGTTGGCGTAAATCAGGTAGTTGGCGTAAACTTATTTGACGTGTAC
CGCGGTAAGGGTGTTGCGGAGGGGTATAAGAGCCTCGCCATAAGCCTGATCCTGCAAGAT
ACCAGCCGTACACTCGAAGAAGARGAGATTGCCGCTACCGTCGCCAAATGTGTAGAGGCA
TTAAAAGAGCGATTCCAGGCATCATTGAGGGATTGAACCT**ATG**

yhiF

GTAATCACTTTGTTACCATCCTCGGCACCATTCAGGGTGAACAACCTGGCTTTATCAATA
AAGTCCCGTATAACTTCCTGGAAGTGAATATGCAGGGCATCCAGGTGTGGCATTTGAGAG
AAGTGGTTAATACCACCTATAACCTGTGGGATTACGGCTATGGTGCATTCTGGCCGGAAC
CGGGCTGGGGTGCGCCTTACTACACCAATGCGGTGAGTCAGGTAACACCTGAGCTGGTCA
AATAACACCACCGAAAGATGCAGATGCTTTCTCAGCATCTGCATCATGCATTACATCAAA
TTAATACACAGTAAGCTAACTATTATTATTATAAGCCCTGTCCTGTTAATTACCTTTGGC
AAACTGATTATAAAGTTAATGTCCGCACCAGGARTCGGTT**ATG**

Figure 4-23 The IHF consensus binding sequence. DNA sequence of the 400 bp upstream region of *ler* (LEE1), *ihfA* and *yhiF* relative to the ATG start codon indicated in bold. The region corresponding to the IHF consensus (WATCARXXXXTTR) is highlighted in red.

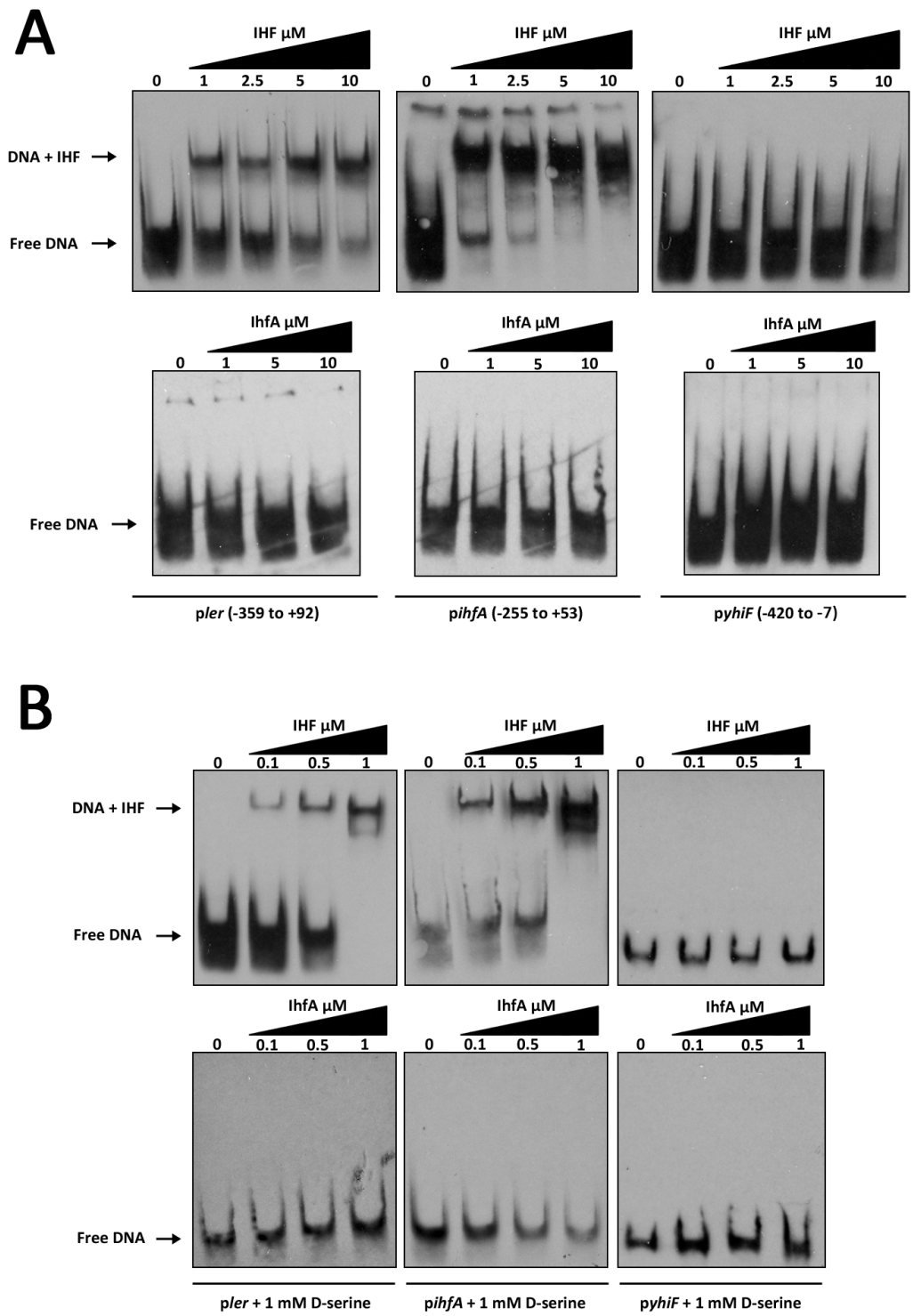

Figure 4-24 Binding of IHF and IhfA to the LEE1, *ihfA* and *yhiF* promoter regions. (A) EMSA analysis of purified IHF and IhfA binding to the LEE1 (p*ler*), *ihfA* and *yhiF* promoter regions. The position of the DNA sequence relative to the ATG start codon is indicated below. The concentration gradient of purified IHF and IhfA relative to ~30 fmol/µl of DIG labelled DNA probe is indicated above each blot. An arrow and label indicates the position of free DNA and DNA-protein complex bands. (B) Identical experiment to panel A with 1 mM D-serine added to the binding reaction. EMSAs were performed in triplicate with similar results each time.

The coding sequence for *yhiF* was cloned into a pMAL-c5x vector (New England Biolabs) allowing purification via a MBP-fusion tag, as described for the YhiF homologue GadE (Tree *et al.,* 2011). Due to time restrictions the purification process for YhiF-MBP was not optimised however a small-scale purification was performed using amylose magnetic beads. Briefly, the cloned construct was transformed into NEB-express cells optimised for MBP expression, cultured in LB with 0.2% glucose and induced with IPTG similarly to His-purification (described in chapter 3). Cells were sonicated and a pull-down was performed using the amylose beads before overexpressed protein was eluted in 10 mM maltose. Analysis of crude elute by SDS-PAGE revealed two heavy bands at ~42.5 and ~63 corresponding to the MBP tag and YhiF-MBP fusion respectively (Figure 4-25A). This crude eluate was diluted 1/10 and used for EMSAs regarding YhiF binding to target DNA. Purified YhiF-MBP showed no observable band shift of free DNA when incubated with its own promoter region probe but was capable of binding the *ihfA* and LEE1 promoter regions (Figure 4-25B). This experiment suggests that YhiF does not regulate its own expression but could potentially modulate the expression of *ihfA* and LEE1. This agrees with data from Figure 4-22 in which YhiF is required for D-serine to repress *ihfA* expression. The YhiF homologue GadE has also been shown previously to be capable of binding the LEE1 promoter although the majority of its repressive activity is mediated through LEE2/3 and LEE5 (Tree *et al.,* 2011).

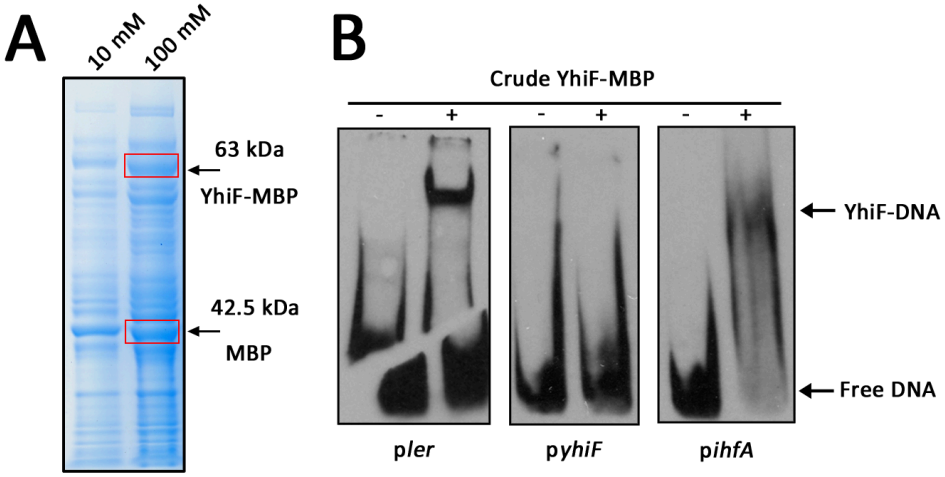

Figure 4-25 Purification and EMSA analysis of YhiF-MBP. (A) SDA-PAGE analysis of YhiF-MBP overexpressed and crudely purified using an amylose bead pulldown. The heavy bands boxed in red correlate to the sizes one would expect to see for YhiF-MBP (~63 kDa) and MBP alone (~42.5 kDa) that has not fused to the recombinant protein. (B) EMSA analysis of crude YhiF-MBP binding to the *ihfA* promoter fragment probe. DNA was incubated with a 1/10 dilution of the crude elute in the binding reaction.

EMSA assays suggested that D-serine was not directly modulating the ability of IhfA or YhiF to bind DNA. In chapter 3 YhaJ was hypothesised to regulate the LEE in response to D-serine via the YhaO membrane transporter protein. Indeed, both YhaO and YhaJ were found to be essential for full LEE expression. RNA-seq data revealed that *yhaOM* were upregulated after D-serine exposure (Figure 4-26A). Monitoring of *yhaO* transcription using a GFP tagged promoter fusion construct (p*yhaO*:GFP) confirmed that YhaO is actively upregulated ($p \leq 0.05$) in response to D-serine supplemented at 1 mM to MEM-HEPES (Figure 4-26B). As a control, L-serine did not alter *yhaO* expression. This result revealed active regulation of *yhaO* in response to D-serine, confirming the physiological relevance of YhaO implied in chapter 3. Seeing as the *yhaOM* promoter is under the control of YhaJ, the ability of YhaJ to bind and regulate the promoters of *yhiF* and *ihfA* was assessed in order to identify if YhaJ was mediating the D-serine response. YhaJ could not bind the *yhiF* promoter but weak binding was observed to the *ihfA* promoter at higher concentrations of YhaJ (Figure 4-26C). The weak binding may be a result of the conditions tested but also suggests the ability of YhaJ to modulate *ihfA* expression. It is proposed that YhaJ could inhibit *ihfA* through D-serine sensing via YhaO and thus contribute to LEE1 repression when D-serine is present. However, the mediator of *yhiF* modulation in response to D-serine remains elusive.

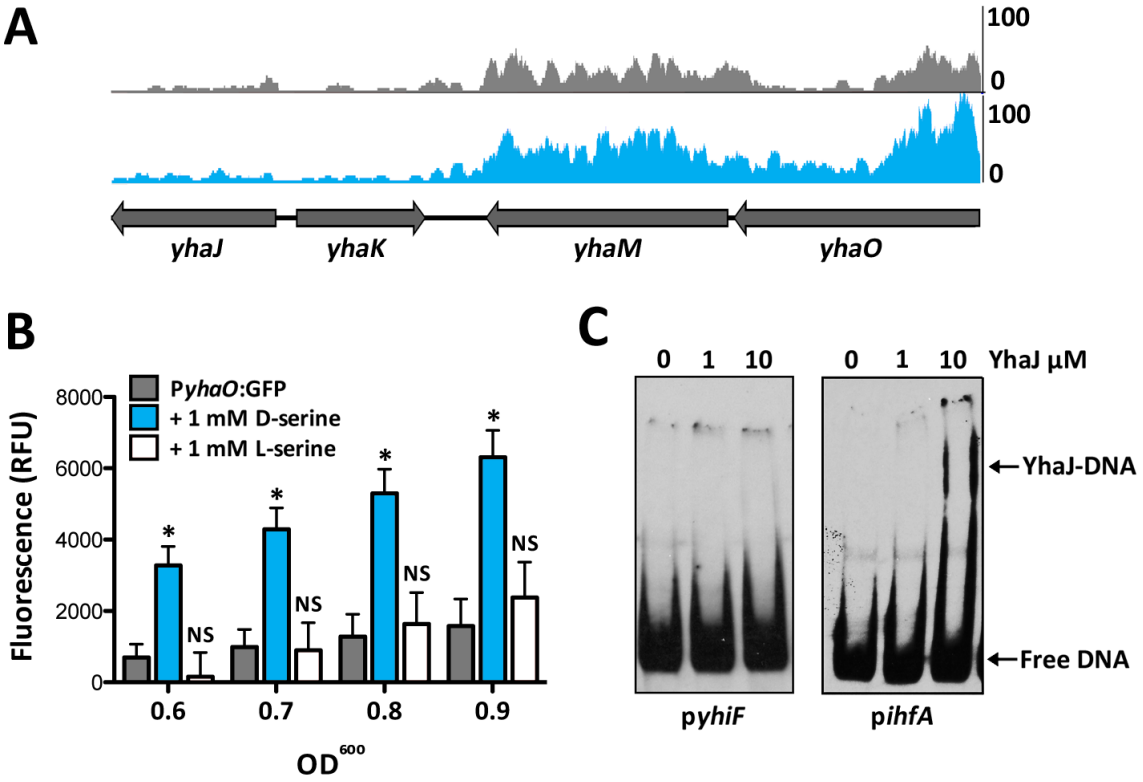

Figure 4-26 Regulation of *ihfA* by YhaJ in response to D-serine. (A) Coverage graph generated from raw RNA-seq data. The grey peaks represent TUV93-0 coverage of *yhaOMKJ* whereas the blue peaks represent coverage in response to D-serine. The shift in expression of *yhaOM* was not significant but replicates revealed a similar pattern of expression. (B) Monitoring of *yhaO* expression in TUV93-0 using the p*yhaO*:GFP construct. Samples were taken from a growth cultures in MEM-HEPES supplemented with and without 1 mM D or L-serine at various time points, measured for OD_{600} and fluorescence (RFU). * and NS denote $p \leq 0.05$ and no significance respectively as compared with the untreated p*yhaO*:GFP control. Experiments were performed in biological triplicate. (C) EMSA analysis of purified YhaJ binding ability to DIG labelled promoter fragments of *yhiF* and *ihfA.* Black arrows indicate the position of free DNA (unbound probe) and a YhaJ-DNA complex. Experiments were repeated three times with similar results gained.

4.4 Additional implications of D-serine exposure

Exposure of O157:H7 to exogenous D-serine resulted in two clear phenotypes – inhibition of the T3SS and activation of the SOS response. These responses were the result of major shifts in the transcriptional profile of the strains in question. Firstly, exposure to D-serine triggered a regulatory cascade that included modulation of the nucleoid regulator IHF and the GAD acid stress regulator YhiF that resulted in repression of the LEE PAI and subsequent attenuation of virulence. Secondly, an intracellular accumulation of D-serine resulted in SOS activation via an unknown mechanism. Intriguingly, these two responses were found to be independent of one another demonstrating the diverse effects that exposure to this metabolite can have on O157:H7.

D-AAs classically have very specific functions in bacteria, with D-glutamate and D-alanine being integral members of the PG sugar-backbone crosslinking peptides. It has been reported on several occasions in the past that NCDAAs provided exogenously at high concentrations could result in incorporation into the PG of the cell wall however only recently have physiological roles for these observations begun to emerge (Caparros *et al.,* 1992, Caparrós *et al.,* 1991). Lam *et al.* have described in detail how diverse bacteria can release numerous D-AAs into the environment and subsequently incorporate them into their PG to downregulate cell wall synthesis under times of stress, revealing a refreshed outlook on the roles of D-AAs in bacteria (Cava *et al.,* 2011a, Lam *et al.,* 2009). Interestingly, D-serine specifically has been shown by the *E. gallinarium* to incur vancomycin resistance through modification of a D-alanine vancomycin target site (Arias *et al.,* 1999). Additionally, D-serine accumulation in a *ΔdsdA* mutant revealed an unusual morphology and was hypothesised to be due to PG remodelling (Anfora *et al.,* 2007, Roesch *et al.,* 2003). Based on these studies, the final aim of this chapter was to address the possibility of D-serine acting as an exogenous PG-modifier in O157:H7.

4.4.1 *Exogenous D-serine exposure reshapes the bacterial cell*

Growth experiments of TUV93-0 in MEM-HEPES supplemented with 1 mM D-serine revealed no effects on growth rate or viability Figure 4-1. SEM analysis of bacteria from these samples however revealed unusual cell morphology. The classically rod-shaped *E.*

coli cells were now noticeably misshapen with a more rounded appearance. This unusual morphology correlated with exposure to exogenous D-serine supplemented in the growth medium.

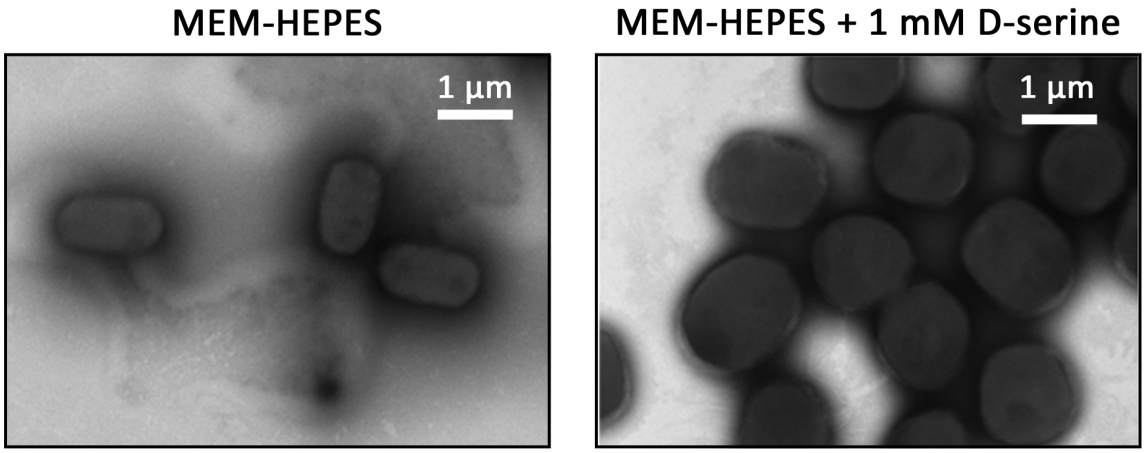

Figure 4-27 SEM images of TUV93-0 exposed to exogenous D-serine. SEM images highlighting the normal morphology (rod-shape) of *E. coli* cells grown in MEM-HEPES compared to the misshapen (swollen, rounded) appearance after addition of 1 mM D-serine to the growth medium. Multiple images were taken per sample and these were selected to highlight the overall trend of the population. A scale bar is indicated in white on the top right hand corner of each image.

In order to address whether this morphological phenotype could be attributed to changes in the PG architecture induced by exposure to D-serine, high performance liquid chromatography (HPLC) analysis of purified PG from cultures grown under the conditions described above were analysed. HPLC analysis allows a quantitative understanding the relationship between PG composition and morphology in tandem with mass spectrometry (MS) to identify novel PG components (Desmarais *et al.,* 2013). Briefly, cultures of bacteria grown in MEM-HEPES (with and without 1 mM D-serine) at 37°C were concentrated and boiled in 8% SDS to disrupt cell wall structures. SDS treated bacterial cells were then processed and HPLC-MS analysed in the lab of Professor Waldemar Vollmer (University of Newcastle). Chromatographic peaks corresponding to the retention times of different muropeptide structures were identified for TUV93-0 with and without D-serine supplemented in the growth medium. Exposure of TUV93-0 to D-serine resulted in the identification of three unique peaks containing measureable D-serine (Figure 4-28A). D-serine was found to be incorporated into position 4 of the muropeptide Tetra monomer (~10% of the total measured muropeptide), the TetraTri dimer (~4%) and the TetraTetra dimer (~8%). Novel D-serine containing peaks were not identified in the

CFT073 wild type but interestingly the *ΔdsdA* mutant could incorporate D-serine into the same positions as for TUV93-0, although to a lesser extent. Complementation of TUV93-0 with DsdA from CFT073 (TUVp*dsdA*) resulted in a marked decrease (~3 fold, ~5 fold and ~2 fold corresponding to Tetra-Serine, TetraTri-Serine and TetraTetra-Serine respectively) in the amount of D-serine containing muropeptide peaks. Furthermore, the CFT073 serine transporter mutant strain (*ΔdsdXΔcycA*) displayed no detectable D-serine containing peaks (Figure 4-28B). These data mimic the data obtained in Figure 4-17 and suggest that uptake and intracellular accumulation of D-serine is required for PG remodelling. The identification of the same unique D-serine containing peaks in two unique strains of *E. coli* (TUV93-0 and CFT073) also suggests a universal mechanism of D-serine incorporation for *E. coli.* The data obtained from these experiments and how they correlate with observations described earlier are summarised in Table 4-5.

Table 4-5 Incorporation of D-serine into PG muropeptide and the associated correlation with D-serine metabolism.

	TUV93-0		**TUVp*dsdA***		**CFT073**		***ΔdsdA***		***ΔdsdXΔcycA***	
D-serine	-	+	-	+	-	+	-	+	-	+
Muropeptide[1]										
Tetra-Ser	0	10.77	0	3.6	0	0	0	6.7	0	0
TetraTri-Ser	0	4.22	0	0.79	0	0	0	0.8	0	0
TetraTetra-Ser	0	8.52	0	4.59	0	0	0	1.89	0	0
Metabolism[2]										
D-serine accumulation		++		+		-		++		-
Growth[3]										
c-source		-		+		+		-		-

[1] % of muropeptides containing D-serine
[2] Intracellular accumulation of D-serine: ++ = abundant; + = minimal; - = none
[3] Ability to grow on D-serine as a sole carbon source: + = yes; - = no

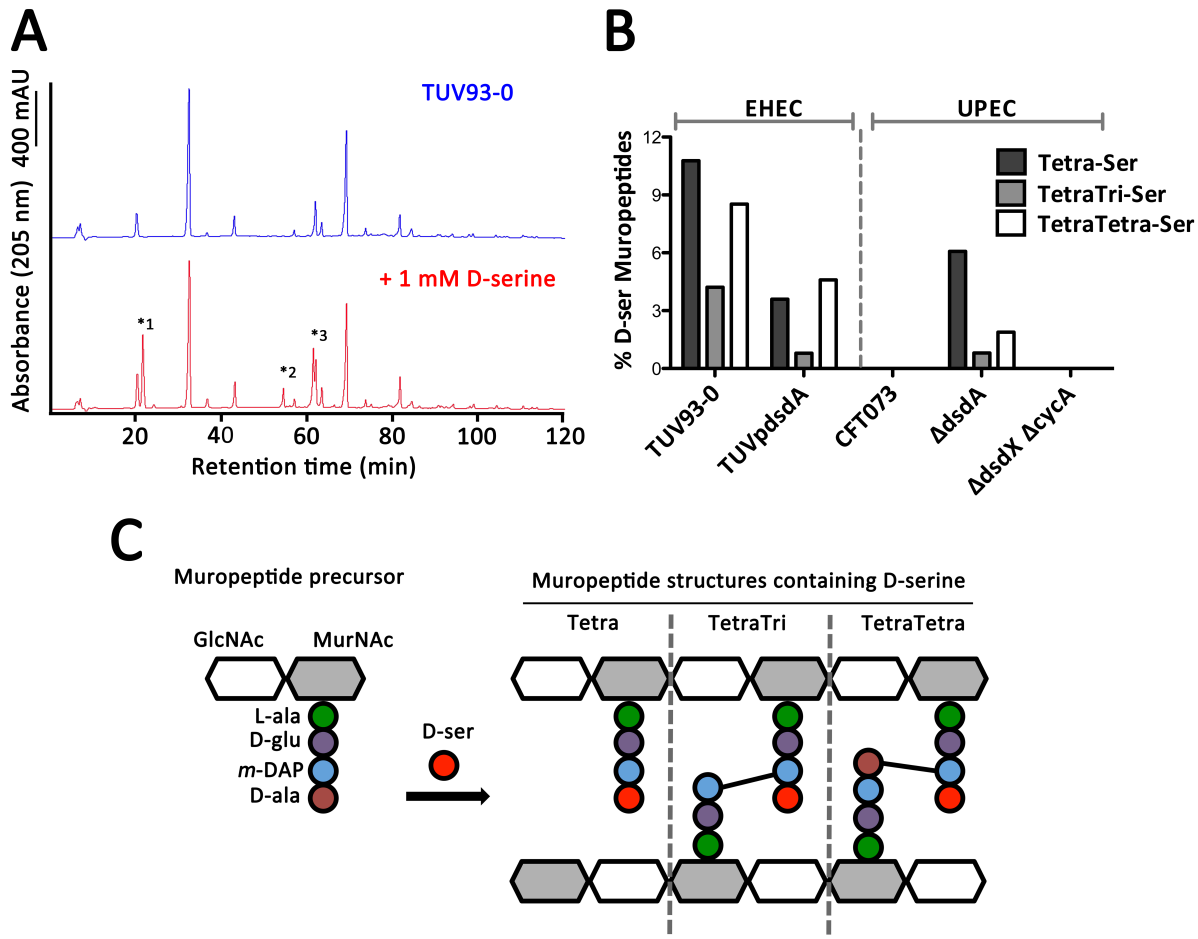

Figure 4-28. D-serine specifically restructures the *E. coli* PG layer. (A) Chromatograph of PG HPLC analysis. Peaks represent detectable muropeptides. The blue peaks correspond to TUV93-0 PG whereas the red peaks indicate the PG profile of TUV93-0 with exogenous D-serine supplied in the growth medium. A * indicates a Tetra peptide monomer peak (1), TetraTri dimer peak (2) or TetraTetra dimer peak (3) containing D-serine as identified by MS. (B) The % of D-serine containing muropeptides in EHEC (TUV93-0 and TUV*pdsdA*) and UPEC (CFT073, *ΔdsdA* and *ΔdsdXΔcycA*). Black, grey and white bars correspond to Tetra-Ser, TetraTri-Ser and TetraTetra-Ser respectively. (C) Schematic representation of a muropeptide precursor and D-serine incorporated structures as identified by HPLC-MS analysis.

The specificity of D-serine incorporation into the Tetra, TetraTri and TetraTetra muropeptide side chains was conserved between EHEC and UPEC, resulted in reshaping of the cell wall and was dependent on intracellular accumulation of D-serine in strains lacking a functional DsdA. Previous work has demonstrated that NCDAAs can be incorporated into the PG of bacteria as an alternative substrate to D-ala or D-glu via the normal biosynthetic pathway or by modification of pre-existing PG crosslinks (Cava *et al.,* 2011a). To attempt to identify a possible mechanism of action for D-serine incorporation, the RNA-seq data was revisited once again and searched for differential expression of

genes that code for known members of the cell wall biosynthetic process. Two genes from the known repertoire of PG biosynthetic genes were significantly differentially expressed in response to D-serine (Typas *et al.,* 2012). Firstly, the lytic transglycosylase (LT) coding gene *mltD* was upregulated ($p \leq 0.001$) and interestingly the second gene to be affected was the D-alanine racemase coding gene *dadX,* which was downregulated in the dataset ($p \leq 0.001$) (Figure 4-29A). MltD is a member of the LT family of PG hydrolases of which there are six other members. LTs are believed to be responsible for cleaving or terminating glycan chains, possibly for PG length control or recycling of pre-existing PG (Typas *et al.,* 2012, van Heijenoort 2011). MltD specifically has been demonstrated to be necessary for muropeptide abundance in *Helicobacter pylori,* despite the large degree of redundancy in PG biosynthetic enzymes suggesting a specific function (Chaput *et al.,* 2007). Additionally, overexpression of MltD has been linked to spheroplast formation and could potentially be responsible for more rounded cells (van Heijenoort 2011). In order to assess the role of MltD in PG incorporation of D-serine a *ΔmltD* deletion mutant was generated by the Lambda Red method (Datsenko & Wanner 2000) (Figure 4-29B). Deletion of *mltD* had no effect on cell growth when compared to TUV93-0 in a variety of media (LB, MEM-HEPES, DMEM) (Figure 4-29C). SEM analysis of cell shape indicated that *ΔmltD* had no observable effect on the morphology of the cell however addition of 1 mM D-serine to the growth medium resulted in a unique morphology. The cells did not adopt the more rounded cell shape of TUV93-0 upon exposure to D-serine but rather developed a striking pointed-polar cap (Figure 4-29D). Indeed, the polar caps of TUV93-0 exposed to D-serine also had a misshapen morphology in a small number of cells visualised but the frequency of this event was much more pronounced in *ΔmltD* where every cell visualised was elongated with pointed poles. This phenomenon has been previously identified in a penicillin binding protein 3 (PBP3) mutant and has not been linked to an LT (Taschner *et al.,* 1988). Due to time constraints the PG profile of *ΔmltD* was not analysed here leaving the role of MltD in PG restructuring unclear but it is exciting to speculate on a specific role for MltD in PG remodelling by exogenous D-serine.

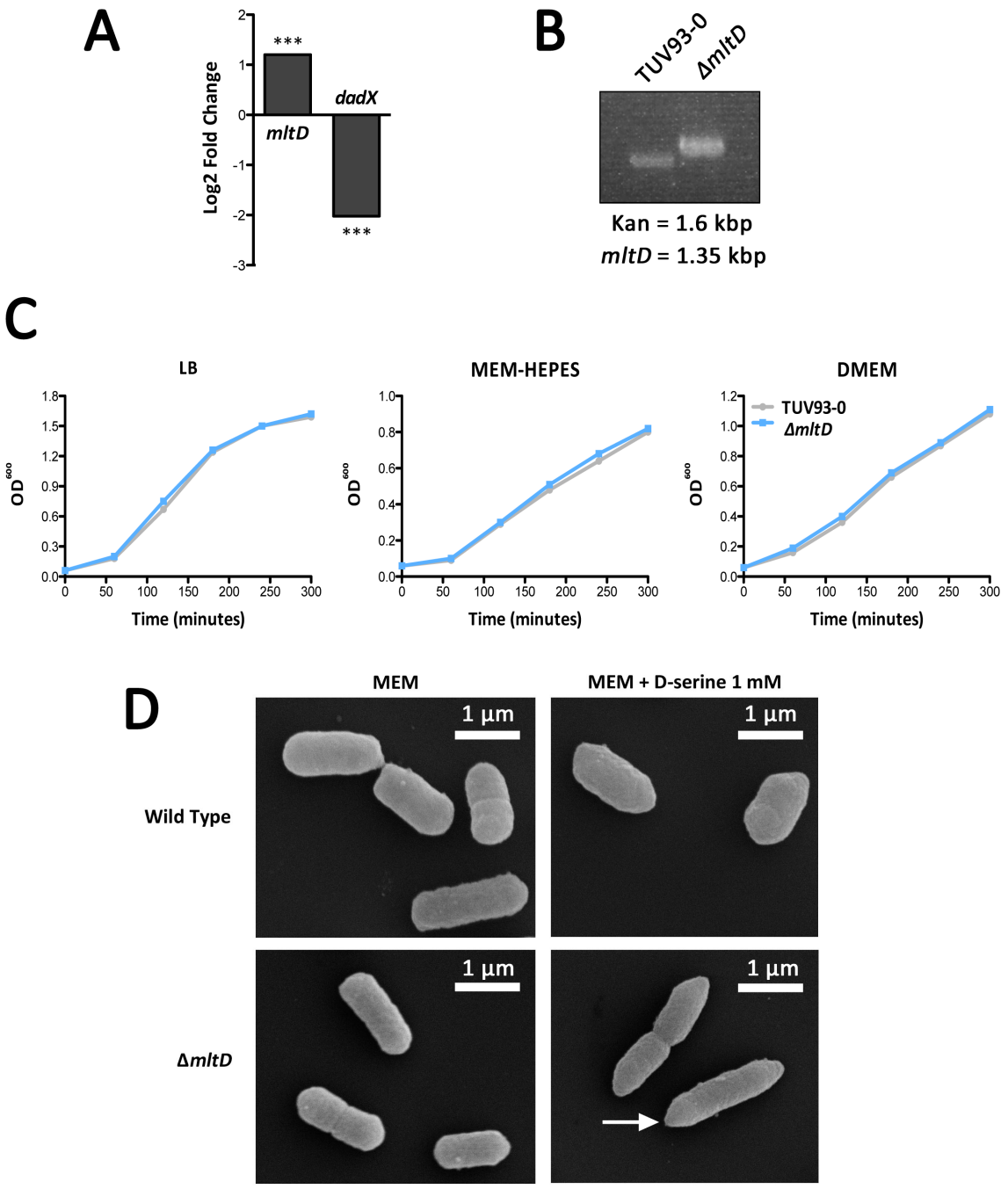

Figure 4-29 Characterisation of the role of MltD in PG remodelling by D-serine. (A) RNA-seq analysis revealed significant differential expression of two PG biosynthetic enzyme encoding genes, *mltD* and *dadX*. *** denotes $p \leq 0.001$. (B) PCR confirming replacement of the *mltD* gene (1.3 kbp) with the Kan^R cassette (1.6 kbp) by Lambda Red mutagenesis. (C) Deletion of *mltD* has no negative effects on growth in LB, MEM-HEPES or DMEM when compared with TUV93-0. Curves show OD_{600} over time and are representative of three biological replicates. (D) SEM analysis showing representative images of cell morphology of wild type TUV93-0 and *ΔmltD* after exposure to 1 mM D-serine supplied in the growth medium. A white arrow indicates the presence of elongated pointed-polar caps observed on all cells in the *ΔmltD* with D-serine sample.

4.5 Discussion

The topic of NCDAAs in bacteria has recently emerged in the literature. Classically, D-isomers of AAs (glutamate and alanine) were only thought to be involved in cell wall biogenesis, due to the inability of D-AAs to be incorporated into ribosomally produced peptides. It was demonstrated decades ago that various D-AAs could substitute D-glu and D-ala in PG incorporation but only recently has this been given physiological relevance (Caparros *et al.,* 1992, Caparrós *et al.,* 1991). Since then, it was discovered that diverse phyla of bacteria both produce NCDAAs by way of racemisation but also release these D-AAs into the surrounding environment. Furthermore, these D-AAs (which can accumulate up to mM concentrations exogenously of the cell) function as regulatory signals for bacteria that can incorporate them into their PG in place of D-ala. It is hypothesised that this event regulates PG structure and abundance, therefore coupling PG architecture with metabolism under conditions of stress (Cava *et al.,* 2011a, Lam *et al.,* 2009).

D-serine has been implicated in independent studies to be involved in cellular processes. It has been demonstrated that members of the enterococci can racemise L-serine into the D-isomer by way of a dedicated serine racemase, VanT. Conversion of the terminal D-ala-D-ala in the muropeptide side chain to D-ala-D-ser was also found to confer vancomycin resistance to the bacteria (Arias *et al.,* 1999). D-serine catabolism has also been extensively studied in *E. coli.* D-serine is toxic to *E. coli* in minimal media and requires the action of DsdA, a D-serine deaminase, to convert D-serine to non-toxic pyruvate and ammonia (Cosloy & McFall 1973). D-serine is abundant in the urinary tract (> 1 mM) and was demonstrated previously to be used both as a carbon source and a regulator of virulence gene expression in the UPEC strain CFT073, therefore contributing to overall fitness and control of virulence in the urinary tract (Anfora *et al.,* 2007, Roesch *et al.,* 2003).

The aim of this chapter was to explore the concept of D-serine modulating gene expression for the InPEC strain *E. coli* O157:H7. Chapter 3 characterised a novel regulatory system in O157:H7 involved in regulation of the LEE. The signal for this system is hypothesised to be D-serine. Environmental sensing of the host physiology is emerging as key factor on how bacteria colonise a particular niche. Cellular accumulation of D-

serine was previously proposed to modulate the expression of numerous virulence associated genes in a CFT073 *ΔdsdA* mutant and more recently in the uropathogen *S. saprophyticus* (Haugen *et al.,* 2007, Korte-Berwanger *et al.,* 2013). O157:H7 does not carry a functional *dsdCXA* system and therefore is susceptible to D-serine toxicity (Moritz & Welch 2006). However, the ability of D-serine to modulate gene expression in O157:H7 has not been explored.

4.5.1 *D-serine represses the LEE without inhibiting growth*

Studying T3S in EHEC requires growth conditions that mimic that of the host. Tissue culture minimal media, DMEM and MEM-HEPES, has long been used to induce LEE transcription *in vitro* and analyse the effects of gene deletions or compounds on expression of the T3SS (Kenny *et al.,* 1997a). In this study it was found that TUV93-0 was able to tolerate the addition of 1 mM D-serine to this growth medium without any effect on growth or viability of the cells. This was in contrast to M9 minimal media in which D-serine inhibited growth, an observation made previously for a K-12 *dsdA* mutant (Cosloy & McFall 1973). Growth in MEM-HEPES at 37°C mimics the gastrointestinal tract and thus conditions favourable by EHEC strains for colonisation. Although growth was not affected, analysis of LEE expression and T3S of associated effector proteins showed both were markedly inhibited by exposure to D-serine. The inhibition occurred at the transcriptional level with significant repression in the LEE1 through LEE5 operons. Due to the master regulator Ler being encoded on the first ORF of LEE1 and its subsequent responsibility for expression of the entire LEE it was proposed that the negative effects were occurring through an inhibition of Ler transcription (Deng *et al.,* 2004, Elliott *et al.,* 2000, Haack *et al.,* 2003, Mellies *et al.,* 1999, Sperandio *et al.,* 2000). D-serine levels in the host are likely to fluctuate and so concentrations as high as 1 mM may not necessarily be reached outside of the urinary tract. Analysis of LEE1 expression over a range of D-serine concentrations confirmed that as low as 100 µM was capable of inhibiting LEE1 expression significantly. Seeing as Ler is encoded on LEE1 this would likely repress the entire LEE in a similar manner. RNA-seq and qRT-PCR was used to analyse the transcriptome of TUV93-0 in response to D-serine. The experiment revealed that 26 ORFs from the LEE, located across operons LEE1 to LEE5 and inter-operonic regions, were significantly downregulated in this environment. The master regulator Ler was indeed found to be expressed much less than in the wild type scenario suggesting that D-serine

repression is occurring through *ler* and subsequently downregulates the remainder of the LEE. Ler is known to be essential for expression of the LEE and thus the ability to form a T3SS (Deng *et al.,* 2004). Therefore, inhibition at the level of Ler represents a direct and powerful method of inhibiting T3S in LEE positive pathogens. Finally, it was shown that addition of D-serine to the growth media during an *in vitro* infection model significantly reduced host cell colonisation and A/E lesion formation. This result demonstrated that not only can D-serine modulate expression of the LEE but it can also inhibit the ability of TUV93-0 to colonise host cells in a T3SS dependent manner. These results confirm that D-serine can modulate expression of virulence factors in contrasting pathogens irrespective of their ability to use it as a carbon source (Anfora *et al.,* 2007) further bolstering the theory that it is involved in environmental signalling. As discussed in chapter 3, InPEC strains have replaced the *dsdCXA* locus with the *cscRAKB* locus for sucrose utilisation and this is believed to have occurred through adaptation to the gastrointestinal niche, an environment more rich in sugars as carbon sources (Jahreis *et al.,* 2002, Moritz & Welch 2006). However, the ability to respond to environmental D-serine is clearly an important trait and is maintained by O157:H7, exemplified by the data presented here.

Although D-serine has been discussed specifically in the context of *E. coli* virulence, other D-AAs have emerged as playing roles in diverse bacteria. D-methionine and D-tyrosine are known to be produced by *V. cholerae* and *B. subtilis* respectively whereas D-glutamate and D-alanine are produced by all bacteria for PG biosynthesis (Cava *et al.,* 2011a, Lam *et al.,* 2009). The D-AAs tested here had minimal effects on growth and no repression of the LEE was observed, with the exception of 1 mM D-alanine. The effect was not as strong as observed for D-serine but still significant. Structurally D-serine and D-alanine are quite similar. Mechanistically, D-serine is believed to inhibit biosynthetic pathways involving β-alanine due to the structural similarity between the two molecules (Cosloy & McFall 1973, Maas & Davis 1950). Therefore, it could be hypothesised that if D-serine is acting as a regulatory signal for gene expression then this could be mimicked by a high concentration of D-alanine unless the production of exogenous D-alanine plays regulatory roles itself, the latter of which remains to be determined.

Additional to its role in regulating the LEE, D-serine increased motility in TUV93-0. Motility is normally repressed in tissue culture media, most likely as these are conditions that promote expression of the LEE and therefore mimic an environment more suited to

colonisation, however under certain inducing signals such as SCFAs motility can be activated (Tobe *et al.,* 2011). L-serine is known to be a strong chemoattractant for *E. coli* representing a positive nutrient readily utilised by the bacteria (Deloney-marino *et al.,* 2003, Mesibov & Adler 1972). D-serine was recently reported to have chemotactic properties although the data obtained were not significant. In fact, the study tested many D-AAs available as potential nutrients in urine none of which were significantly chemotactic. Certain L-AAs on the other hand tested positive for chemotaxis (Raterman & Welch 2013). Experiments performed here suggested that D-serine was not chemoattractive but may possess chemorepellant properties. This was logical considering the effects on the T3SS colonisation factor. It is also believed that flagella have adhesive properties used for initial attachment to host tissue followed by a regulatory switch between the flagella and the T3SS in a temporal manner (Beckham *et al.,* 2014, Mahajan *et al.,* 2009). It makes little sense for a cell to express both in tandem and creates a plausible scenario whereby colonising cells might sense metabolites such as D-serine, downregulate attachment mechanisms and become more motile in order to evade a toxic niche. This result is open to interpretation however as even though intact flagella were identified, there was no significant increase in transcription of *fliC.* Additionally, RNA-seq analysis did not reveal any significant differential expression in flagella-associated genes and regulators but investigation of the raw data suggests that there is indeed some variation in transcription of flagella-associated genes. However, the lack of significance might suggest a slight shift in expression simply as a result of LEE repression. Moreover, microscopy analysis suggests that only a small proportion of the population have become flagellated in the presence of D-serine, quite opposite to the infection model that revealed the formation of significantly fewer A/E lesions on host cells in the presence of D-serine. The use of emerging single-cell transcriptomic technology could be applied in this instance to shed light transcriptional variation within mixed populations (Saliba *et al.,* 2014).

The gastrointestinal tract is a complex and highly competitive environment. O157:H7 perceives signals from the environment and responds appropriately to them by regulating its colonisation factors in search of the most favourable niche. D-serine has emerged as a key regulator of virulence in contrasting pathogens that can also be highly toxic. It is proposed that it can act as an environmental signal that can be transduced to the

genome, affecting LEE expression and influencing when and where bacteria attach to host tissues.

4.5.2 *D-serine as an unexpected activator of the SOS response*

RNA-seq analysis revealed differential expression of 411 genes in response to D-serine exposure. The pattern of expression was well conserved between replicates. Although D-serine had no negative effect on growth under these conditions the transcriptome data revealed that various genes relating to cellular stress responses were differentially expressed. Most strikingly was the activation of the bacterial SOS response. The SOS response is classically defined as a global response to DNA damage and encodes numerous genes involved in DNA repair (Baharoglu & Mazel 2014). 26 member of the SOS response were activated after exposure to D-serine and these included LexA, the SOS repressor, and RecA, the LexA antirepressor and thus positive regulator of SOS induction (Fernández De Henestrosa *et al.,* 2000). Even though *lexA* was found to be upregulated also, expression of *recA* is considered indicative of a bona fide SOS response and a previous report has suggested that both genes are continually expressed during persistent SOS activation (Shimoni *et al.,* 2009). Cellular DNA damage has many causes including UV irradiation, pressure, antibiotic treatment and oxidative stress. SOS activation by antibiotic treatment has also been observed at concentrations lower than that required to affect growth, which is interesting considering D-serine activates an SOS response but normal growth continues (Baharoglu & Mazel 2014). Surprisingly, the SOS cell division inhibitor SulA is activated but does not halt cell division (Trusca *et al.,* 1998). This may be due an unknown factor involved in the D-serine induced SOS response. Indeed, the mechanism of SOS activation by D-serine was unclear during these studies however it is known that SOS induction can overlap with other cellular stress responses to ensure cell survival under aggressive environmental conditions. Additional to the SOS response, changes were observed in multiple cellular responses including metabolism and cell wall composition but these were diverse and included many hypothetical genes making an explanation for the phenotype observed difficult to extrapolate. Its likely that activation of the SOS stress would cause a global shift in gene expression alone as the bacteria would be entering a “survivalist” state as opposed to normal exponential phase growth.

SOS induction has been linked to virulence regulation in diverse pathogens such as *V. cholerae* and *S. aureus* (Baharoglu & Mazel 2014). Specifically in the case of O157:H7, SOS induction has been shown to activate Stx expression either by spontaneous SOS activation or after exposure to sub-inhibitory concentrations of antibiotics (Fuchs *et al.,* 1999, Nassar *et al.,* 2013). In this study the O157:H7 EDL933 derivative TUV93-0 was used, which is deleted for Stx, thus the idea of D-serine induced Stx production was beyond the scope of this project but it is interesting to note that a collaborating group found no increase in Stx expression or phage lysis after exposure to D-serine (personal communication) and SOS induction here did not affect bacterial growth. Mellies *et al.* have previously shown that the SOS response can mediate LEE2/LEE3 expression in EPEC through LexA binding to the promoter of this divergent operon. The study also suggested that SOS can regulate NLE expression providing a potential global signal for effector expression across diverse cryptic prophage (Mellies *et al.,* 2007b). The clear phenotype of LEE repression and SOS induction in response to D-serine therefore led to the hypothesis that the SOS response may also regulate the LEE in EHEC but in a unique way to EPEC. The SOS response and presumably D-serine toxicity was found to be entirely dependent on the presence of a functional DsdA. DsdA protects UPEC from intracellular accumulation of D-serine and thus hypermodulation of gene expression by metabolising D-serine into non-toxic by products (Anfora *et al.,* 2007). Accumulation of D-serine in a CFT073 *ΔdsdA* mutant resulted in overexpression of multiple virulence factors leading to the hypothesis that exogenous D-serine is sensed as a signal for virulence gene expression (Haugen *et al.,* 2007). Providing TUV93-0 with DsdA *in trans* resulted in deactivation of the SOS response but did not restore LEE expression, suggesting that the SOS response and repression of the LEE were likely occurring via independent mechanisms. This result was not met with disregard. The CFT073 *ΔdsdA* mutant was highly susceptible to D-serine toxicity indicated by growth perturbation and induction of the SOS response. It was apparent that intracellular accumulation of D-serine was concurrent with SOS activation and thus suggests a novel mechanism of this response in *E. coli.*

Avoiding stress is key for bacteria competing in complex environments. It is clear that diverse *E. coli* pathogens have the ability to sense and respond to this host derived signal and can therefore use this to modulate gene expression both in respect to survival and also expression of key virulence factors. The D-serine induced SOS response is particularly intriguing when one considers the role of the SOS response in bacteria. Indeed, the

classical definition of SOS activation is in response to DNA damage, either spontaneous or induced (Baharoglu & Mazel 2014, Nanda *et al.,* 2015). However, the SOS response is known to activate the lytic life cycle in chromosomally carried prophage. This creates an almost paradoxical scenario whereby a core stress response used to promote survival can ultimately lead to lysis of the cell. For this reason, temporal regulation of the SOS response is key to limiting unnecessarily prolonged SOS activation and this is an active area of research (Nanda *et al.,* 2015). LexA regulated genes are known to be expressed heterogeneously by stochastic factors and the exact levels of RecA expression required to induce a lytic SOS response are undefined (Kamenšek *et al.,* 2010, Shimoni *et al.,* 2009). Various reports have attempted to quantify this but the results gained are largely dependent on the system and conditions used (Cárdenas *et al.,* 2012, Casaregola *et al.,* 1982). Here, an ~10 fold increase in *recA* expression was observed (relative to the growth conditions used) after D-serine exposure but also a ~6 fold increase in *lexA* expression. A persistent expression of *lexA* may mask the effects that RecA has on other phage specific repressors by saturating the substrate range of RecA. The RNA-seq dataset also revealed that the RecA filament inhibitors, DinD and RecX, involved in RecA activity control are induced more strongly than RecA (Cárdenas *et al.,* 2012, Massoni *et al.,* 2012, Uranga *et al.,* 2011). It is possible that a previously undefined post-translational regulation of the SOS response may be inhibiting its normal function and thus allowing normal growth. This may explain why D-serine exposed cells induce an unexpected SOS response but do not undergo lysis or growth inhibition. Additionally, single cell analysis has previously revealed that individuals within large populations are capable of activating the SOS response in shared environments. It is possible that a detectable increase in RecA may be coming from a proportion of the population but the overall growth dynamics are maintained. A recent study by Xu et al proposed a mechanism in O157:H7 whereby spontaneous lysis of a small subpopulation, resulted in Stx release which subsequently primed intestinal epithelial cells for colonisation by the remainder of the population in a T3SS dependent manner, thus SOS mediated lysis by Stx phage actually promotes O157:H7 colonisation and potentially creates a selective pressure for phage acquisition (Xu *et al.,* 2012). Indeed Stx positive EHEC are known to encode more cryptic prophage encoded effector genes than EPEC (Deng *et al.,* 2012, Tobe *et al.,* 2006). In this study, the SOS response occurred in tandem with LEE repression. Although the responses were found to be independent of one another, it cannot be ruled out that the SOS response may play some role in deactivation of colonisation under stressful conditions. It is exciting

to postulate that developing a sub-lytic method of SOS activation could potentially provide a novel method of reducing colonisation of pathogenic bacteria.

4.5.3 *Hijacking of transcriptional regulatory networks by D-serine*

LEE repression and the SOS response represented at least two major shifts in gene expression as a result of D-serine exposure. However, with > 400 genes being differentially expressed it is clear that there is not a universal system controlling these responses. The more likely scenario is that numerous regulatory networks are working in parallel to elicit such responses. Indeed, a recent report identified >50 genes belonging to the Ler regulon that are not LEE or NLE-associated. These data were growth phase dependent, which highlights the fine balance that must be achieved between PAIs and the core genome in temporal regulation of virulence factors (Bingle *et al.,* 2014). From the RNA-seq data, 30 transcriptional regulators were identified as being differentially expressed (16 upregulated and 14 downregulated) in response to D-serine. Not all of these have been characterised in the literature but the diverse functions that the remainder encode are suggestive of the significant impact that exposure to a toxic metabolite can have on how a cell functions irrespective of virulence. In this study, particular interest was given to the small number of regulators identified that are known to be involved LEE regulation.

Two transcriptional regulators of the GAD acid stress response, YhiF and GadX, showed increased expression in response to D-serine. GadX did not appear to mediate the D-serine phenotype but YhiF was found to be necessary for repression of the LEE in response to D-serine by way of a *ΔyhiF* mutant that was incapable of significant LEE downregulation, although the transcript levels of *ler* and *espD* were not fully restored suggesting other factors involved in this response. GAD regulators are known to be involved in mediating LEE repression (Branchu *et al.,* 2014, Tatsuno *et al.,* 2003). This is believed to allow passage through the stomach, a very acidic environment, without unnecessary expression of colonisation factors such as the T3SS. However, only two members of the acid fitness island, a network of genes involved in battling acid stress, were significantly activated in response to D-serine presumably as a by product of YhiF and GadX activation (Tramonti *et al.,* 2008). This makes sense, as D-serine does not alter the pH of the growth media. The GAD regulators are not innately responsible for

virulence regulation but rather this function has been adapted upon acquisition of the LEE. Recently, the prophage encoded secretion regulators (Psr) were identified on NLE encoding cryptic prophages in O157:H7. These indirect repressors of the LEE were shown to hijack the GAD regulators YhiF and GadE, mediating LEE repression. The mechanism was hypothesised to promote competitive delivery of NLEs to the T3SS apparatus (Tree *et al.,* 2011). In an independent study, GadX and GadE were demonstrated to mediate LEE repression in response to environmental NO stress. This complex study also identified the NO sensitive repressor, NsrR, as being an activator of the LEE in the absence of NO (Branchu *et al.,* 2014). Interestingly, NsrR was also upregulated in response to D-serine although this seems counterintuitive given its role in LEE activation. Nonetheless, it is clear that the GAD regulators are emerging as an adaptable system capable of being regulated in diverse ways to elicit control over the LEE. This is highlighted by a report that identified the extensive promoter of GadE as a "sensory integration region" involved in sensing of diverse signals from the environment (Sayed & Foster 2009). In addition to YhiF, analysis of the RNA-seq data revealed the downregulation of the IHF alpha subunit, IhfA. IHF is a global regulator of numerous cellular processes (Dillon & Dorman 2010). IhfB was not affected by D-serine but it is known that IhfA and IhfB are capable of functioning as both homodimers and heterodimers (Mangan *et al.,* 2006). A recent study revealed that IHF was capable of binding ~30% of all operons in *E. coli* K-12 and ~10% exhibited differential expression in an IHF mutant under the conditions tested (Prieto *et al.,* 2012). More specifically, IHF plays a critical role in regulation of the LEE1 operon and thus LEE2 through LEE5 (Friedberg *et al.,* 1999). Similarly to YhiF, IhfA was found to be necessary for mediating D-serine repression of the LEE. A plasmid borne copy of *ihfA* expressed constitutively was capable of restoring wild type expression levels of LEE1 expression even in the presence of D-serine.

Based on this it was hypothesised that YhiF and IhfA might be acting in tandem to respond to D-serine. Interestingly, analysis of *yhiF* and *ihfA* expression levels in the corresponding deletion/complementation background revealed that in the *ΔyhiF* background *ihfA* was no longer repressed in the presence of D-serine. Similarly, in the TUVp*ihfA* strain *yhiF* was no longer activated. EMSA analysis performed also demonstrated that YhiF was capable of binding the *ihfA* promoter, suggesting that in the presence of D-serine YhiF can repress *ihfA* and thus LEE1 expression. Expression analysis did reveal that deletion of *yhiF* could restore *ihfA* expression and reverse the response to

D-serine but not fully. Indeed, YhiF has been previously proposed to regulate the LEE, independently of Ler, through LEE2/3 and LEE5 (Tatsuno *et al.,* 2003, Tree *et al.,* 2011). It is possible that YhiF enhances the repression of the LEE by blocking *ihfA* transcription and inhibiting LEE2/3/5 but is not absolutely required. Neither IHF or IhfA were found to bind the predicted *yhiF* promoter region under the conditions tested however this does not rule out the possibility that IHF might repress *yhiF* under normal conditions as *yhiF* has been predicted to be co-transcribed with an adjacent gene *slp* (Tucker *et al.,* 2003).

Although it was convincing that a regulatory interplay between YhiF and IhfA was controlling the response to D-serine, the data obtained suggest that this is rather indirect regulation. D-serine did not directly affect IHF or YhiF binding to DNA and it is therefore likely that another system responds to exogenous D-serine and signals into the proposed regulatory mechanism. As discussed in chapter 3, the putative *yhaOMKJ* D-serine sensing system was found to be absolutely necessary for LEE activation via Ler. It was therefore hypothesised that this system might initially sense D-serine and therefore hijacks the IhfA-YhiF regulatory loop to strengthen its response. YhaJ was hypothesised to control *yhaO* transcription and regulate steady state LEE expression but maintained constant levels in order to rapidly adapt to signals from the environment. In contrast, *yhaO* was actively upregulated by D-serine providing an important link between *yhaOMKJ* and the LEE. It was therefore proposed that YhaJ could sense D-serine via YhaO and subsequently transmit signals to the LEE via the YhiF/IhfA regulatory loop. Purified YhaJ was not able to bind the putative *yhiF* promoter region but weak binding to the *ihfA* and LEE1 upstream regions was observed by EMSA. It is unclear whether YhaJ would therefore activate or repress IhfA activity but perhaps it is capable of both. The binding was indeed weak and not enhanced by addition of D-serine to the binding reaction but this theory assumes that D-serine is the direct cofactor required for differential regulation, which is unlikely. YhaK was hypothesised previously to be a potential cofactor for YhaJ transcriptional regulation and therefore may be required to mediate such responses (Gurmu *et al.,* 2009). The binding of YhaJ to the *ihfA* promoter could therefore be weak or transient in the absence of the correct cofactor. Due to the requirement of YhaJ and IHF for LEE expression it is therefore proposed that YhaJ positively influences *ihfA* in the absence of the repressive signal D-serine.

This system has a huge level of complexity. It is clear that numerous networks have been hijacked in order to regulate the response to D-serine but exactly how is unclear. YhaJ and IhfA are likely linked as discussed above but the mechanism of YhiF activation is less apparent. YhiF shares homology (23% identity) with GadE and is known to be part of the regulons responding to acid stress but its precise role has not been described (Tatsuno *et al.,* 2003). Indeed, multiple studies have reported that YhiF is actually dispensable for the acid stress response meaning that it must harbour alternative roles (Kailasan Vanaja *et al.,* 2009, Masuda & Church 2002, Tucker *et al.,* 2002, 2003). Nonetheless, this is the third report to identify LEE control by the GAD regulators independently of acid stress further bolstering the theory that this is a highly adaptable system (Branchu *et al.,* 2014, Tree *et al.,* 2011). YhaJ was proposed in chapter 3 to be an *E. coli* specific regulator of the LEE and YhiF is known to be unique to species of *E. coli* and *Shigella* (Sayed & Foster 2009). Transcriptional regulation of the LEE is complex and multifactorial. In UPEC the *dsdCXA* regulator DsdC has been proposed to regulate numerous genes including those involved in virulence (Anfora *et al.,* 2007). One could speculate that InPEC strains that have lost *dsdCX* and thus the ability to metabolise D-serine therefore overriding pre-existing regulatory networks to efficiently regulate virulence in response to non-favourable environmental stress. The overlap between GAD and NO stress regulation is particularly intriguing with respect to the D-serine response when one considers that YhaO is not only a member of the Ler regulon but also part of the NO response network and that YhaK has been reported to be upregulated by NO donors and human urine, a concentrated source of D-serine (Bingle *et al.,* 2014, Flatley *et al.,* 2005, Gurmu *et al.,* 2009). There is an abundance of literature available describing the role of D-serine in the brain as a signalling coagonist and recently studies have revealed cross-talk between NO, D-serine and associated biosynthetic enzymes however no such associations have been made between the two in prokaryotic systems (Darra *et al.,* 2009, Wolosker *et al.,* 2008). This study has therefore uncovered novel links between three independent stresses and virulence regulation in *E. coli* highlighting the adaptability and efficiency of bacterial pathogens. The data presented here build on those presented in chapter 3 and therefore allow the proposal of a new working model for LEE regulation in response to D-serine.

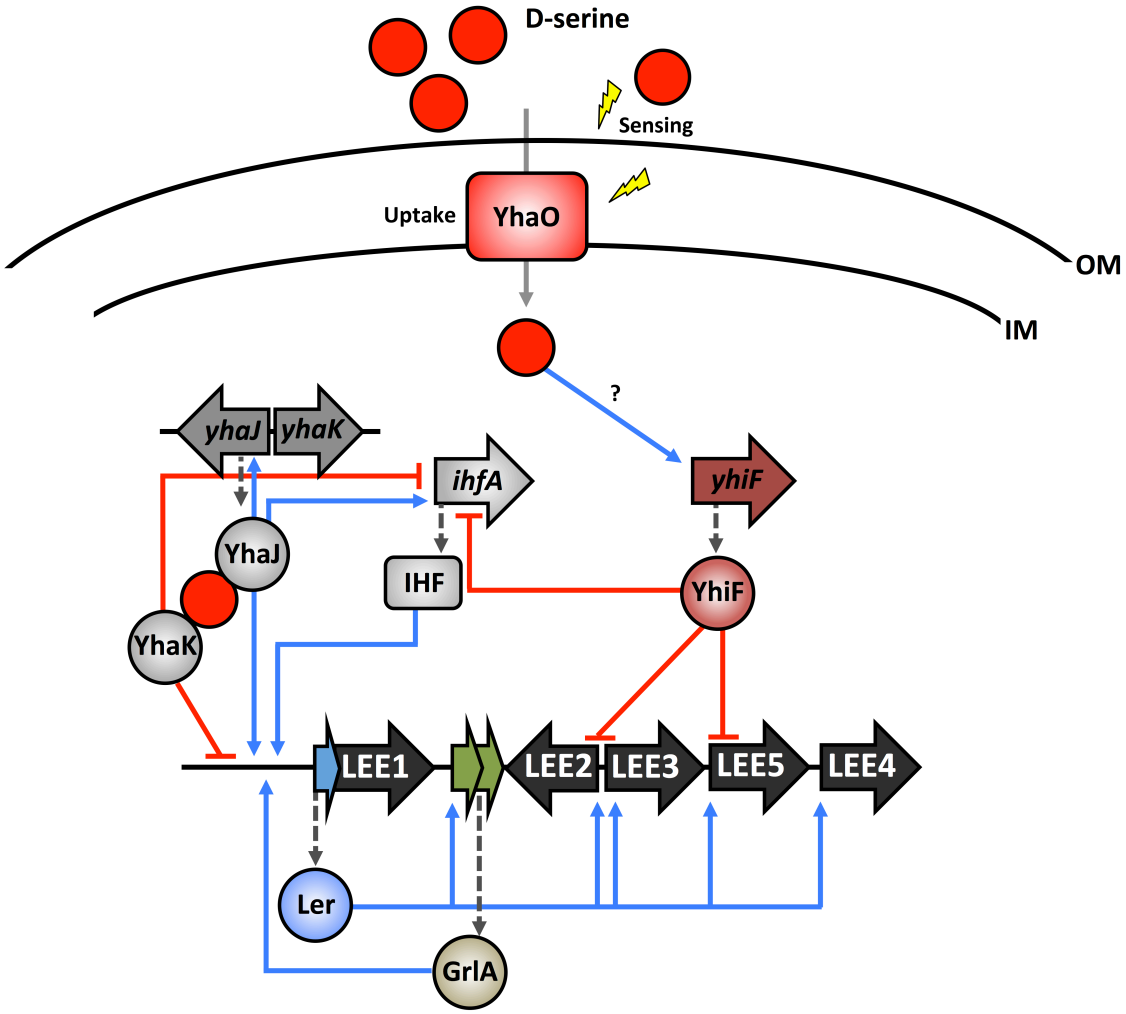

Figure 4-30 Model of transcriptional hijacking by D-serine. Schematic model proposing the mechanism of D-serine sensing and subsequent hijacking of three transcriptional regulatory systems in *E. coli* O157:H7. IHF promotes LEE1 expression and thus full LEE activation through the Ler-GrlRA regulatory feedback loop. YhiF is an adaptable negative regulator of LEE2/3/5. YhiF is upregulated in the presence of D-serine by an unknown mechanism and subsequently inhibits *ihfA* transcription as well as inhibiting LEE2/3/5 expression. YhaJ is postulated to regulate steady state LEE expression and act as an environmental sensor system for D-serine, in combination with YhaO and the cofactor YhaK. This is proposed to assist in differential LEE regulation by inhibiting both YhaJ and IHF mediating activation of LEE1. Blue and red arrows represent positive and negative transcriptional regulation respectively. Broken grey arrows represent translation of the indicated regulator. The red circle represents D-serine as a direct or indirect signalling molecule.

4.5.4 *D-serine reorganises cell wall architecture*

The role of NCDAAs in bacteria is emerging as an important theme in contemporary bacteriology, particularly in regulated PG reorganisation (Cava *et al.,* 2011b). Previous reports have identified D-serine as an alternative substrate for the terminal D-alanine on PG peptide cross-bridges in the *Enterococci* (Arias *et al.,* 1999). It was therefore hypothesised that O157:H7 exposed to exogenous D-serine may be susceptible to PG restructuring by this D-AA. Similar to a report by Roesch *et al.* using a CFT073 *ΔdsdA* mutant, it was found in this study that TUV93-0 exposed to D-serine resulted in unusual cell morphology (Roesch *et al.,* 2003). Cells no longer exhibited the classical rod-shaped appearance of *E. coli* but rather appeared more rounded. It was hypothesised for CFT073 *ΔdsdA* that the intracellular accumulation of D-serine was responsible for PG restructuring although it was not investigated further (Anfora *et al.,* 2007). In this study TUV93-0 was confirmed to accumulate intracellular D-serine and this result was entirely dependent on the lack of a functional DsdA. Thus, it appears that accumulation of intracellular D-serine in either EHEC or UPEC can result in unusual cellular morphology. Furthermore, analysis of the PG structure between the conditions and strains tested indicated a distinct pattern of PG architecture that was consistent with D-serine accumulation. D-serine replaced the terminal D-alanine at position four of three distinct PG side chain types. Incorporation of NCDAAs in *V. cholerae* and *B. subtilis* was recently reported to occur both by the modification of pre-existing PG in the periplasm and by the normal biosynthetic pathway (Cava *et al.,* 2011a). The data presented here revealed that for *E. coli* (either EHEC or UPEC) PG remodelling by D-serine, intracellular accumulation was absolutely required suggesting that D-serine replaces D-alanine as the substrate for the normal biosynthetic pathway of PG precursors. The positions of incorporated PG were also consistent between both EHEC and UPEC suggesting a common mechanism and not merely random insertion. These data also further bolster the theory that due to physical similarities in the structure between D-serine and D-alanine, an abnormal concentration of either might mimic the effect of the other D-AA. It was interesting to note that despite PG remodelling, the abundance of muropeptide was not largely altered in *E. coli.* Indeed, TUV93-0 did not exhibit any growth defects in MEM-HEPES supplemented with D-serine suggesting that PG reorganisation may have a more regulatory role as hypothesised for other diverse bacteria, however CFT073 *ΔdsdA* was more susceptible to growth inhibition by D-serine in MEM-HEPES, an observation made previously for growth in human urine

(Roesch *et al.,* 2003). The release of D-AAs and subsequent PG reorganisation in *V. cholerae* occurred in tandem with entry into stationary phase as it was hypothesised to signal metabolic slowing in a stressful condition (Lam *et al.,* 2009). For D-serine, it appears incorporation is not growth phase dependent but this does not rule out physiological relevance of the event.

Cell wall biosynthesis is a complex process. Numerous enzymes are employed at various stages to maintain efficient turnover of this integral cell component (Typas *et al.,* 2012). To explore the idea of mechanistic reasoning behind D-serine incorporation the RNA-seq data was investigated for genes known to encode for PG biosynthetic enzymes. Two genes were differentially regulated in the dataset. Firstly *dadX,* encoding one of the two known alanine racemases in *E. coli,* was downregulated (Kang *et al.,* 2011). D-serine replaces D-alanine as a substrate for PG precursors in the study presented. It is tantalising to imagine that oversaturation of the cell with D-serine may cause a feedback to DadX to downregulate its expression. Secondly *mltD,* a gene encoding one of the seven LTs in *E. coli,* was significantly enhanced in its expression in response to D-serine. LTs are one of the families of hydrolases in *E. coli.* Due to the fact that the number of hydrolases outnumbers the number of enzymes required for PG biosynthesis, LTs are considered to have a high degree of redundancy with multiple deletions being required to have even small implications on PG structure. However, one must assume that they are maintained in such abundance for a specific purpose (van Heijenoort 2011). LTs are involved in cleavage of synthesised and processed glycan chains for the purpose of cell separation and efficient peptidoglycan recycling (Typas *et al.,* 2012). In a *ΔmltD* mutant strain exposed to D-serine a unique morphology was observed. The cells exhibited an elongated shape with pointed poles at either end, a morphology previously only observed in a PBP3 mutant strain (Taschner *et al.,* 1988). This result suggested a specific role for MltD in response to D-serine. As mentioned, LT's are involved in PG recycling. Perhaps MltD is upregulated as an emergency mechanism to attempt to turnover PG faster in response to D-serine incorporation. Indeed, the cells are elongated in the *ΔmltD* strain but only when exposed to D-serine. Little is known about how LT's are regulated but overexpression of MltD has been observed to result in spheroplast formation resulting in cell lysis (van Heijenoort 2011). It is possible that an increase in MltD in order to "detoxify" the cell wall from D-serine results in over cleavage of glycan chains and thus a more rounded cell morphology. These data are preliminary but provide fascinating insights into the roles of

D-AAs in PG architecture and cell morphology as well as potentially novel mechanisms of cell wall related enzymes.

Making the association between intracellular accumulation of D-serine and PG reorganisation was an important observation. It was discussed earlier how an intracellular accumulation of D-serine was concurrent with activation of the SOS response in both EHEC and UPEC and that this was independent of growth inhibition. Classically the SOS response is activated after damage to DNA by stresses such as pressure, UV or reactive oxygen species (Baharoglu & Mazel 2014). Antibiotics that target DNA or associated replication machinery are therefore also known activators of the SOS response. Interestingly, β-lactam antibiotics have emerged as unexpected activators of the SOS response in *E. coli* and *S. aureus.* β-lactams function by interacting with PBPs and inhibiting cell wall synthesis. PBP3 was found to be the mediator of this response and signals to the genome through a two-component regulatory system called DpiAB. Increased expression of *dpiAB* by β-lactam antibiotics results in activation of the SOS response in a RecA/LexA dependent manner, halting of cell division and protection against antibiotic lethality (Maiques *et al.,* 2006, Miller *et al.,* 2004). No increase in *dpiAB* expression was observed in response to D-serine but this work revealed a direct link between the cell wall and SOS activation. Additionally, several antibiotics have been found to activate the SOS response at concentrations below those needed to inhibit bacterial growth (Baharoglu & Mazel 2014). This is interesting as D-serine activated the SOS response in TUV93-0 without affecting growth or viability of the population significantly but this activation was indeed concurrent with D-serine accumulation and incorporation into the cell wall.

4.5.5 *Conclusion*

The results presented in chapter 4 reveal D-serine as an inhibitor of the LEE encoded T3SS in O157:H7 and a novel activator of the SOS response in diverse *E. coli.* The mechanism behind the latter is unclear but is potentially linked to cell wall reorganisation by this D-AA. Maintaining normal structure and function of the bacterial cell wall is of the utmost importance for withstanding osmotic pressures and acting as a physical barrier from the environment. Reorganisation of the cell wall as a regulatory mechanism linked to cellular metabolism is also emerging as a physiologically relevant process in diverse bacteria

(Cava *et al.,* 2011b). Therefore, D-serine incorporation into the PG may not be detrimental for growth *per se* but could potentially occur to activate the SOS response and signal a shift in the gene expression profile of the cell in response to this otherwise toxic D-AA. As this occurs in tandem with a downregulation of the T3SS, it is proposed that *E. coli* could sense D-serine as a signal of when and where to attach to host tissues. D-serine is known to be a host-derived metabolite found at high concentrations in both the urinary tract and the brain with reports indicating its presence in plasma and gastric juices also at varying concentrations (Nagata *et al.,* 1992, 2007). Exogenous sources of D-serine for humans however are a less explored area. There are reports that indicate excessive alkali or heat treatment during food processing can result in racemisation of AAs present (Man & Bada 1987). Specifically, Finley reported that D-serine was found in food formulated for weight loss to contain as much as 50% D-serine which could potentially pose a risk to both the human and the microbiota (Richardson & Finley 1986). Although this is an extreme example, the ingestion of D-serine is a likely scenario at some point in time meaning that bacteria found in the gut that are sensitive to its toxic nature must be able to respond to it appropriately. One must also appreciate that *E. coli* O157:H7 is a natural coloniser of cattle and its prevalence as a human pathogen is somewhat accidental. The ability to respond to diverse environmental challenges is therefore understandable as these bacteria primarily colonise the terminal rectum of the ruminant host and are therefore at almost immediate exposure to the environment before even reaching the human host (McNeilly *et al.,* 2008). Together, these findings provide novel insights into the host metabolite D-serine as both a regulatory signal controlling virulence and a modulator of the SOS response in *E. coli.*

5 Investigating the incompatibility between carriage of the LEE PAI and the ability to tolerate D-serine via the *dsdCXA* locus

5.1 Introduction

In chapter 3, a novel regulatory system of the LEE in O157:H7 was proposed. This system senses the environment and signals to the genome altering the expression of virulence factors in response to certain stresses. The substrate of the system, D-serine, was investigated in chapter 4 and found to have profound effects on the global transcriptomic network. The toxic metabolite that is derived from both the host and ingested through food sources downregulates the LEE and activates the SOS response concurrently by independent mechanisms. It was therefore hypothesised that InPEC strains that have lost the ability to metabolise D-serine via *dsdCXA* maintain an ability to respond to this signal presumably to secure a competitive stance in complex and diverse environments.

The ability to utilise D-serine as a carbon source and how this relates to *E. coli* genotypes has been explored previously. The *dsdCXA* locus for D-serine tolerance was found to be intact in ExPEC strains but truncated in InPEC strains resulting in an inability to utilise D-serine catabolically. The *dsdCX* truncation was also found to occur at a conserved site and the genes were replaced by the *cscRAKB* locus for sucrose utilisation in many InPEC strains. This replacement occurs at the *argW* tRNA gene, a recombination hotspot for integration of bacteriophage and associated genetic elements. Indeed, additional surrounding genes were found to vary between *E. coli* isolates but the presence of *cscRAKB* was a common theme. Correspondingly the researchers performed a phenotypic characterisation of the ability to utilise D-serine and sucrose as carbon sources among a collection of InPEC and ExPEC strains. Unsurprisingly, 97.5% of ExPEC strains tested could utilise D-serine whereas only 5% of InPEC strains had this ability, a trait intriguingly limited to a small number of STEC O111 strains (Jahreis *et al.,* 2002, Moritz & Welch 2006). The ability for InPEC strains to selectively utilise sucrose provides an obvious and tangible advantage for colonisation of the gut, as does the reciprocal for ExPEC strains as they disseminate into sugar-poor niches. However, why this genetic replacement occurs always at the *argW* site resulting in this phenotypic switch in nutrient utilisation is obscure.

The ability of D-serine to downregulate the LEE in O157:H7 was found in chapter 4 to be independent of whether a functional DsdA was present or not (Figure 4-18). This

sensitivity to D-serine would therefore limit LEE-positive pathogens to environments in which D-serine is at low concentrations, irrespective of the ability to metabolise it. One hypothesis is that acquisition of the LEE is linked to the carriage of a truncated *dsdCX* as ExPEC strains, which encode an intact *dsdCXA,* are known to not encode the LEE. In this chapter, a comparative genomics approach was applied to test the hypothesis of LEE and *dsdCXA* incompatibility across the entire *E. coli* phylogeny. This work was performed in collaboration with Dr Robert Goldstone and Professor David Smith who performed the computational analyses and assisted with interpretation of the data. Selected experiments from this chapter have been published recently in the ISME journal as presented in section 9 (Connolly *et al.,* 2015).

5.2 LEE and *dsdCXA* prevalence across the *E.coli* phylogeny

E. coli have been broadly grouped into seven phylogroups (A, B1, B2, C, D, E and F) based on the distribution of target genes used for evolutionary analyses but this generally accepted structure of the *E. coli* phylogeny does not take into account virulence genes (Clermont *et al.,* 2013, Goldstone *et al.,* 2014). Interestingly, a general correlation between pathotypes and phylogroups can be seen. For instance, phylogroup E houses O157:H7 strains, the highlight of this thesis. Other, InPEC and commensal *E. coli* are found usually within groups A, B1 and C whereas ExPEC strains are predominantly located within phylogroups B2, D and F (Goldstone *et al.,* 2014, Ogura *et al.,* 2009, Sims & Kim 2011, Touchon *et al.,* 2009). This idea of pathotype/phylogroup correlation is not without its exceptions however. The well studied O127:H6 str. E2348/69, which is a member of the EPEC family of LEE-positive InPEC, is found in phylogroup B2 along with diverse UPEC pathogens such as CFT073 that were discussed in this thesis (Elliott *et al.,* 1998, Iguchi *et al.,* 2009). Using this framework, results presented here describe how the LEE and *dsdCXA* locus are carried across the *E. coli* phylogeny. One would assume a strong correlation between LEE acquisition and InPEC strains but the acquisition/loss of *dsdCXA* is much less predictable given its rather indispensible requirement for ExPEC but obscure role in commensals and InPEC.

5.2.1 *Carriage of both the LEE and dsdCXA is extremely rare*

In order to assess the carriage of the LEE and *dsdCXA* across the *E. coli* phylogeny, a phylogenomic tree was first constructed as per Goldstone *et al.* (2014). Briefly, the genome sequences of 1591 *E. coli* and *Shigella* strains were obtained from publically available databases (as per 3/6/2014). From these sequences, 159 core genes that were found to have high sequence identity and a low occurrence of recombination were used to generate a maximum likelihood tree of the *E. coli* phylogeny (Goldstone *et al.,* 2014). This tree was highly consistent with the structure of previously published works on *E. coli* evolution (Clermont *et al.,* 2013, Ogura *et al.,* 2009, Sims & Kim 2011, Touchon *et al.,* 2009). Next, the amino acid sequences of every ORF encoded within the LEE from EDL933 and DsdC/DsdX/DsdA from CFT073 were iteratively aligned to the 1581 genomes in question. In general, the cutoff for calling a gene "present" was if the protein coding sequence aligned at >70% identity over >80% of the coding sequence in question. To account for variations in the sequences encoded on the LEE between strains and the splitting of sequences across contigs for unfinished genomes, the entire LEE was considered "present" if a genome possessed 21 or more of the LEE encoded ORFs. Seeing as it is well known that the *dsdCX* truncation is highly conserved and results in phenotype switching, the *dsdCXA* locus was only considered "present" if all three genes were identified (Moritz & Welch 2006).

Overlaying the results of the gene-presence identification protocol onto the phylogenomic tree of *E. coli* produced a striking pattern of LEE/*dsdCXA* carriage. There was a clear trend of LEE positive strains being grouped predominantly in phylogroups E and B1, with less frequencies of LEE carriage occurring in groups B2, A and D. This is not surprising and is in agreement with the literature as O157 EHECs and non-O157 EHECs/EPECs are known to cluster in groups E and B1 respectively (Hazen *et al.,* 2013). The clustering of certain EPEC strains, including E2348/69 as discussed above, has been known to occur in phylogroup B2, which is largely composed of ExPEC strains (Goldstone *et al.,* 2014, Hazen *et al.,* 2013). The research by Hazen et al further characterised the diverse nature of *E. coli* pathotypes by clustering a small number of unclassified A/E causing pathogens in the closely linked phylogroups D and F, a similar observation to that presented here (Hazen *et al.,* 2013). Interestingly, the occurrence of *dsdCXA* across the *E. coli* phylogeny was largely reciprocal to the pattern of LEE carriage. Phylogroup B2

displayed an almost ubiquitous trend of strains being *dsdCXA* positive, as did phylogroup A, suggesting that carriage of *dsdCXA* is not solely associated with ExPEC strains. Many strains in the analysis carried neither the LEE nor *dsdCXA,* particularly apparent in groups B1, C and D. The *Shigella* species was included in the analysis due to their close relatedness to *E. coli. Shigella* are not A/E pathogens and thus do not carry the LEE. Most *Shigella* in the analysis were *dsdCXA* negative. A phylogenomic tree highlighting the *E. coli* phylogroups and corresponding carriage of the LEE and *dsdCXA* is shown in Figure 5-1.

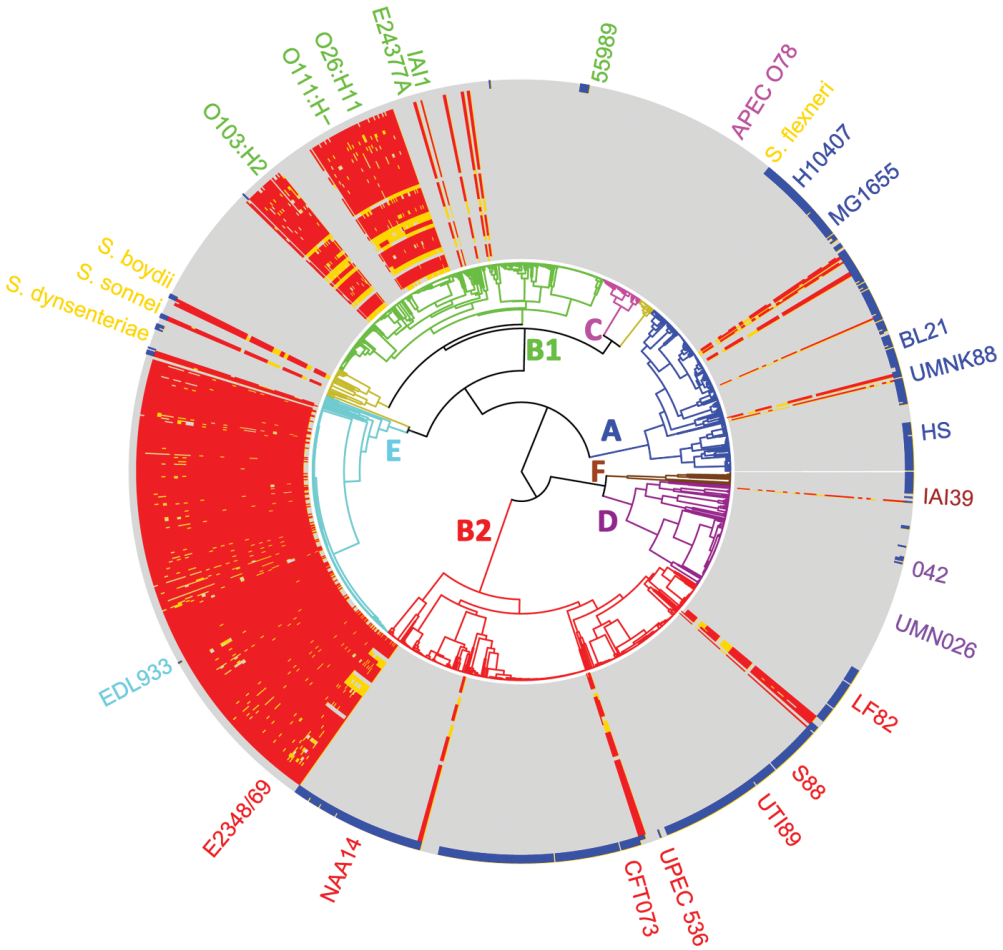

Figure 5-1 Carriage of the LEE and the *dsdCXA* locus across the *E. coli* phylogeny. Circularised phylogenomic tree of 1591 *E. coli* and *Shigella* isolates overlaid with the gene carriage of the LEE PAI (red blocks indicate >70% identity over >80% of the coding sequence) and the *dsdCXA* locus as defined by the presence of intact *dsdCX* (blue blocks indicate >70% identity over >80% of the coding sequence). To account for gene sequences that are divergent but may still be considered "present" a lower threshold was applied also and indicated by yellow blocks (>50% over >50% of the coding sequence). Phylogenetic sub-grouping is indicated on the circularized dendrogram by branch color-coding as follows: A = blue; B1 = green; B2 = red; C = magenta; D = purple; E = cyan; F = brown; *Shigella* = gold (these isolates do not strictly form a phylogroup but were highlighted in gold for the purpose of this analysis). Prototypical strains used to identify phylogroups are indicated outside according to the color code of the phylogroup they belong to.

Data presented in chapter 3 describe how exposure to D-serine downregulates the LEE in TUV93-0. TUV93-0 harbours the *dsdCX* truncation thus rendering its remaining DsdA useless. Complementation of TUV93-0 with a plasmid expressing a functional DsdA from UPEC CFT073 allowed growth on D-serine as a sole carbon source but the effects on the LEE were not reversed. It was concluded that the ability to metabolise D-serine was independent of virulence gene regulation and thus the presence of a functional DsdA has no bearing on whether a strain can use the LEE or not. This suggests that the LEE and *dsdCXA* are functionally incompatible, as the ability to metabolise D-serine via *dsdCXA* would strongly select for dissemination into a D-serine rich environment where the LEE would be useless. It was hypothesised that this selective pressure or incompatibility forces LEE positive strains to lose *dsdCX*. To further bolster this hypothesis the largely ExPEC phylogroup B2 was examined closely. The majority of group B2 strains (344/427; 82%) carried *dsdCX.* This group also contains a number of EPEC isolates, which are well characterised A/E pathotypes only capable of infecting the gut. Their pathogenesis is defined by and therefore reliant on the LEE. The data shows clearly that even though these EPEC strains are genetically more similar to other members of the B2 phylogeny, acquisition of the LEE consistently results in truncation of *dsdCX* (Figure 5-2). Furthermore, it is clear from Figure 5-1 that this LEE acquisition/*dsdCX* truncation event has occurred at numerous times on the evolutionary timeline, irrespective of phylogroup, providing further evidence for this hypothesis.

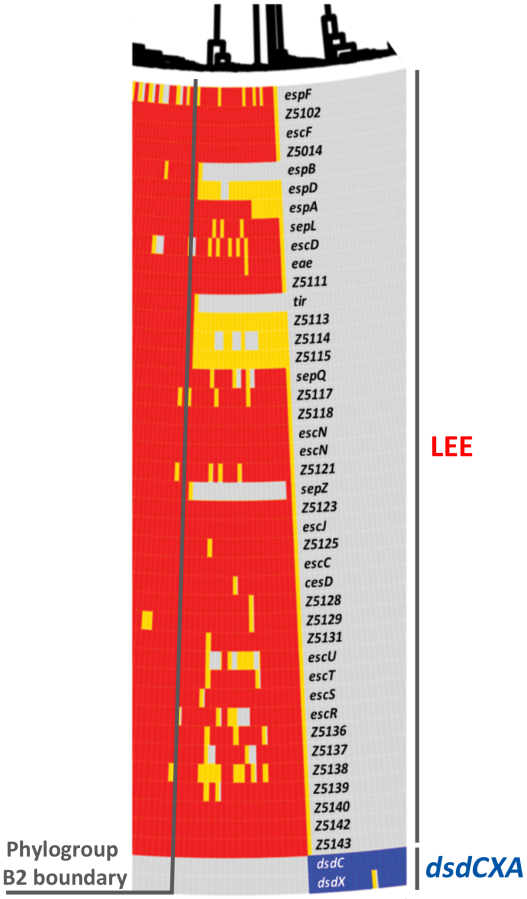

Figure 5-2 Phylogroup B2 highlights the LEE/*dsdCXA* incompatibility. Expansion of a selected section of the B2 phylogeny from Figure 5-1. LEE positive EPEC strains from B2 cluster together and display a consistent pattern of *dsdCX* loss exemplifying the clear distinction between carriage of the two loci irrespective of phylogroup. The annotations for genes encoded on the LEE (red) and *dsdCXA* (blue) are labelled next to the block that they represent. Note that *dsdA* is not included as the *dsdCX* truncation confers an inability to metabolise D-serine (Moritz & Welch 2006). The B2 boundary is indicated by a grey line. Z numbers for LEE encoded genes were obtained from NCBI.

5.2.2 *The dsdCXA locus is ancestral to E. coli*

The LEE is a horizontally acquired PAI that is an accessory to the core *E. coli* genome. Therefore an important point to address was the origin of the *dsdCXA* locus. One could assume that given the role of D-serine metabolism in extraintestinal pathogenesis and the probable lack of extensive D-serine concentrations in the gut, the *dsdCXA* locus could have been acquired allowing dissemination to new niches. On the other hand, the intact *dsdCXA* is predominantly found in phylogroup B2 and its closer neighbour phylogroup A. Furthermore, recent evolutionary studies have suggested that the B2 group were the first to diverge closely followed by phylogroup D. It is believed that *E. coli* commensalism is a

more recent attribute in the evolutionary timeline and that the progenitor may have been classified as an opportunistic pathogen, as is the case for UPECs (Hazen *et al.,* 2013). In any scenario, the *dsdCX* truncation appears to be mechanistically conserved and might be as a result of the pressure from LEE acquisition as discussed above. To test this the *dsdA* sequence was extracted from the genomes in the analysis above. A maximum likelihood phylogenetic tree was then constructed based on the divergence of the *dsdA* sequence. It was found that *dsdA* diverged in a pattern comparable to that of the *E. coli* core genome and is likely to have evolved according to the same phylogroup structure and therefore is proposed to have been encoded on the progenitor genome. Seeing as *dsdA* is maintained after the *dsdCX* truncation (>98% of strains in this study carried *dsdA;* 1561 of 1591), it was concluded the ability to metabolise D-serine via *dsdCXA* was likely a capacity of the progenitor and a common trait among early diverging *E. coli*, for instance phylogroup B2.

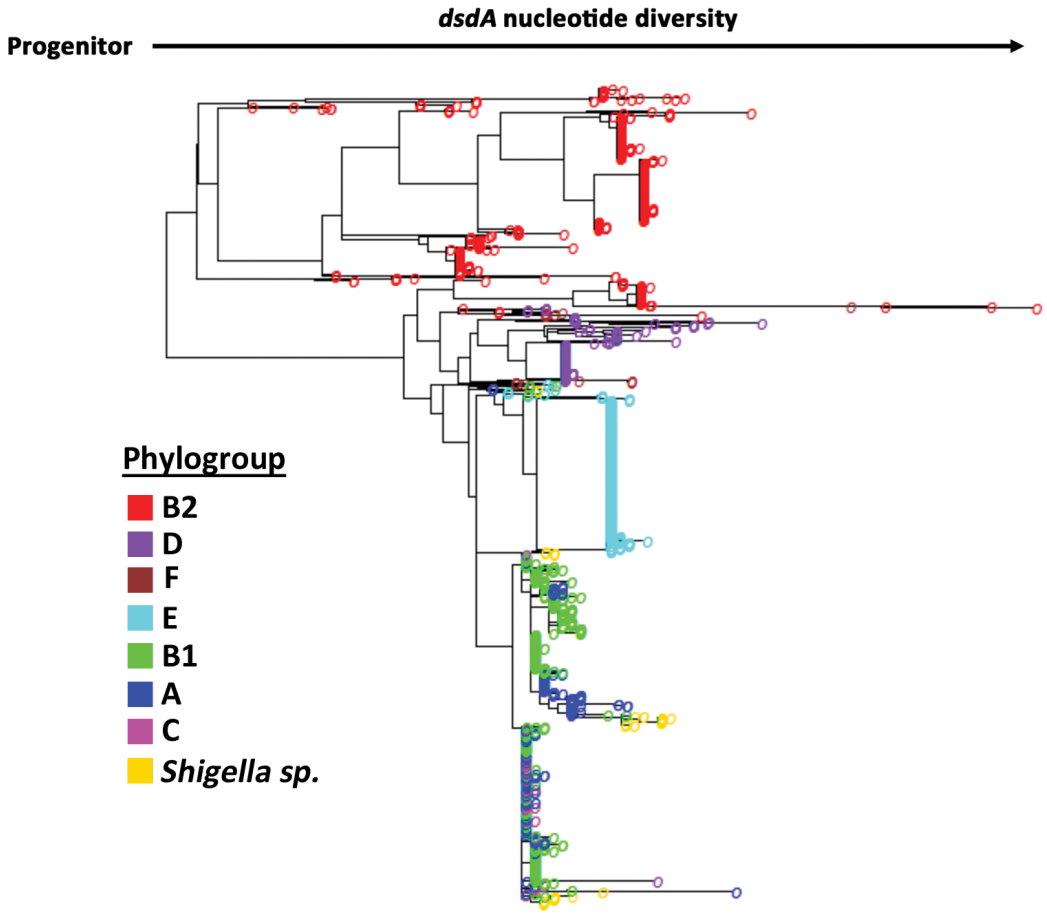

Figure 5-3 *dsdA* is ancestral to *E. coli*. Maximum likelihood phylogenomic tree generated from the diversity in the *dsdA* nucleotide sequence among genomes analysed in this study. A coloured circle represents an individual isolate and is colour coded according to the phylogroup with which it belongs. The phylogroup colour coding is as follows: A = blue; B1 = green; B2 = red; C = magenta; D = purple; E = cyan; F = brown; *Shigella* = gold.

5.2.3 *Acquisition of the LEE and loss of dsdCXA is a significant event*

The data presented in Figure 5-1 show a clear pattern of carriage for either the LEE or *dsdCXA* with a striking correspondence for the absence of the reciprocal locus. Nonetheless, it was still apparent that a small number of cases, albeit rare, existed whereby a strain carried both elements. Specifically, only 26 isolates out of the 1591 analysed were positive for both LEE and *dsdCXA,* occurring in phylogroups B2, D, B1, A and E (Figure 5-4A). Five isolates from phylogroup B2 were absent for *dsdX* only (Figure 5-4B). These exceptions are not extensively characterised in the literature, were predominantly faecal isolates and none are known to have been sourced from ExPEC infections (Table 8-3).

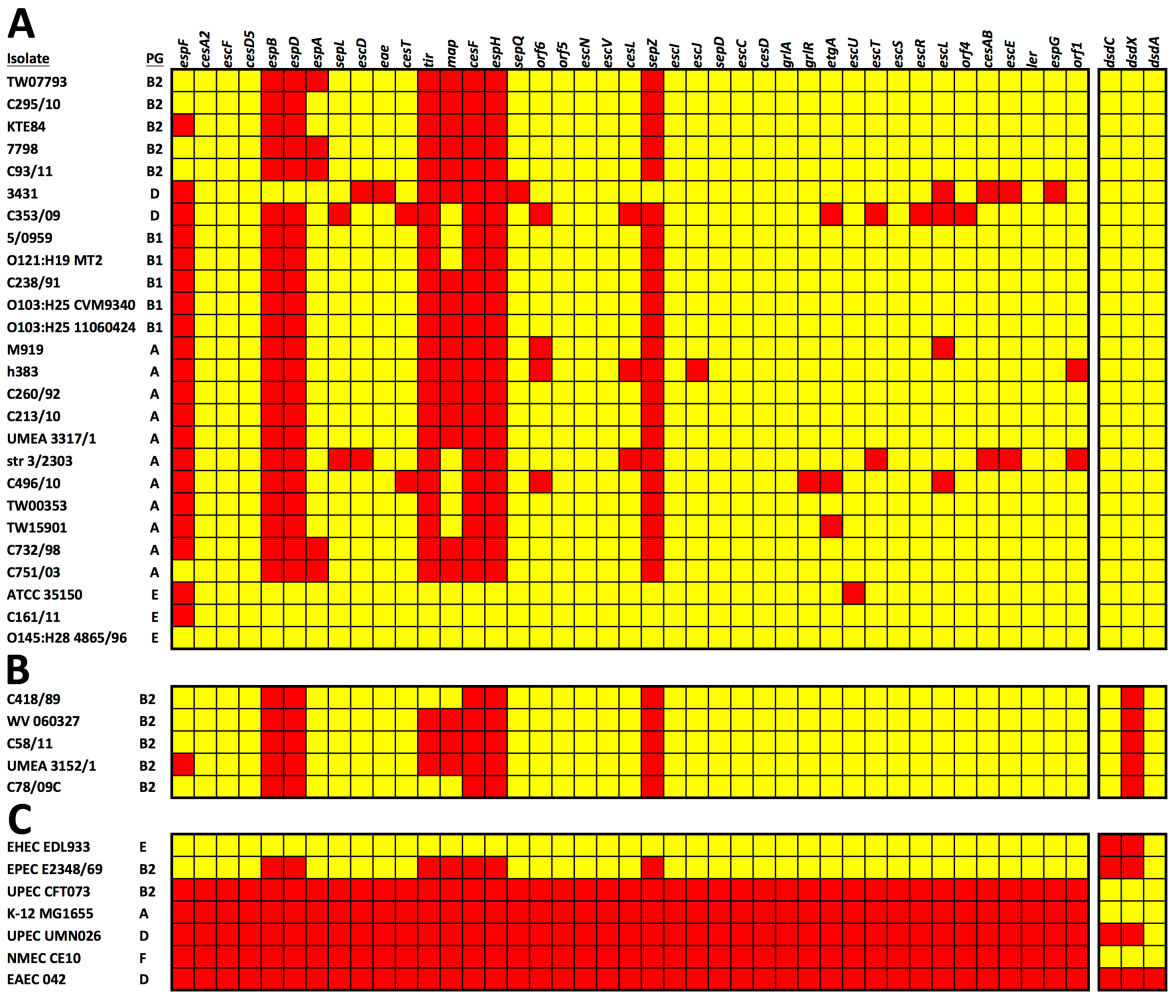

Figure 5-4 Compilation of LEE and *dsd* positive isolates. Colour-coded binary presence/absence matrix of isolates from Figure 5-1 testing positive for both the LEE and *dsdCXA* (A), LEE and *dsdCA* only (B) and a selection of prototypical *E. coli* strains as a comparator (C). Yellow blocks are considered present (>70% identity over >80% of the coding sequence) whereas red blocks are considered absent. Phylogroups are labelled under PG.

It is not surprising that there are some exceptions to the hypothesis of LEE/*dsdCXA* incompatibility. Indeed, Figure 5-3 demonstrates that *dsdA* and likely the *dsdCXA* locus are ancestral to *E. coli.* The LEE on the other hand is a known horizontally acquired PAI that confers the A/E phenotype onto the accepting strain (Elliott *et al.,* 1999, McDaniel *et al.,* 1995). Therefore, as the LEE is not replacing *dsdCX* at that particular point in the chromosome there will likely be an adaptation period before selective pressure for colonisation of the gut and against niche dissemination force the deletion of *dsdCX* from its genome. Nonetheless, the pattern of reciprocal locus carriage was a statistically significant event ($p < 0.0001$; Fishers exact test). 428 isolates (26.9%) in this study were LEE positive/*dsdCXA* negative with an odds ratio of 16.49 (that is, 16.49 times more likely for an isolate to be LEE positive/*dsdCXA* negative than LEE positive/*dsdCXA* positive). 569 isolates (35.8%) tested *dsdCXA* positive/LEE negative with an odds ratio of 21.88. In contrast, among LEE negative strains, there was no correspondence with presence or absence of the *dsdCXA* locus (odds ratio of 0.99) with a corollary in *dsdCXA* negative isolates that are LEE negative (1.37). In other words, LEE negative isolates were equally as likely to have retained *dsdCXA* as they were to have lost it, whereas isolates carrying either the LEE or *dsdCXA* displayed a statistically significant correspondence to absence of the reciprocal locus (Figure 5-5).

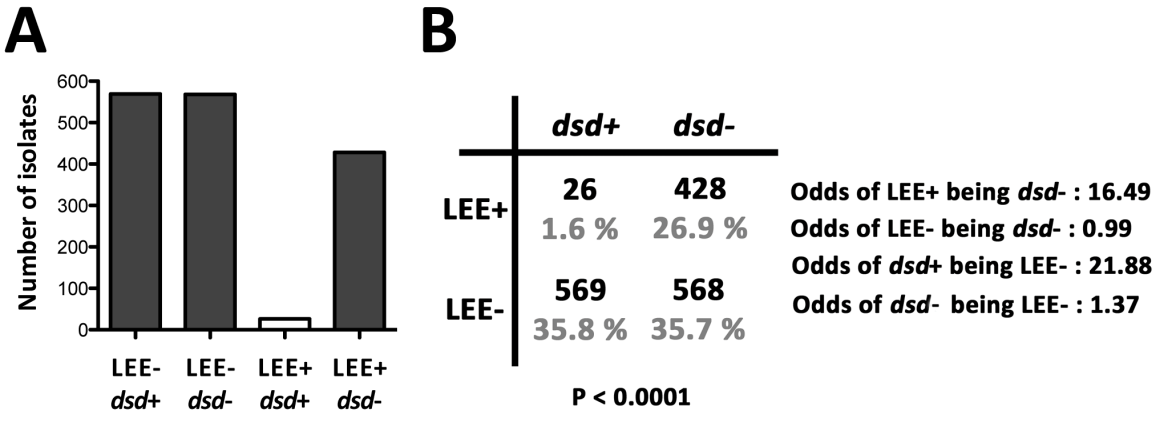

Figure 5-5 Reciprocal carriage of the LEE and *dsdCXA* locus in *E. coli.* (A) Bar chart visually representing the distribution of LEE or *dsdCXA* carriage among 1591 *E. coli* strains analysed. A + denotes presence whereas a – denotes absence. (B) 2 x 2 contingency matrix showing the relationship between LEE+, LEE-, *dsd*+ and *dsd*- isolates. The number of strains for each scenario as well as their percentage distribution among the total 1591 isolates tested are indicated in black and grey respectively. Statistical significance was determined using a Fishers exact test. The odds ratios calculated for each scenario are indicated also.

5.2.4 *The prevalence of cscRAKB within the E. coli phylogeny*

The ability to metabolise D-serine is an important trait for *E. coli* pathogens both in terms of niche fitness and virulence gene regulation. A previous observation that the sucrose utilisation locus, *cscRAKB*, replaced *dsdCX* at the *argW* recombination site formed the basis of a niche optimisation model whereby ExPEC strains retained *dsdCXA* to allow dissemination and competitive fitness outside of the gastrointestinal tract. Accordingly, InPEC strains acquired the ability to utilise sucrose presumably to improve fitness in the gut where sugar sources are more readily available (Jahreis *et al.,* 2002, Moritz & Welch 2006). To address the role of sucrose adaptation, the prevalence of *cscRAKB* was assessed using the above presence/absence criteria. 421 (26.4%) isolates were LEE and *cscRAKB* positive, with an odds ratio of 12.76. Conversely, the LEE/*cscRAKB* negative group (622 isolates; 39.1%) had an odds ratio of 1.21 indicating an equal likelihood that a LEE negative strain was carrying *cscRAKB* or not. This result mirrored that obtained for LEE/*dsdCXA* negative isolates in Figure 5-5. Accordingly, *cscRAKB* negative isolates were 18.85 times more likely to be LEE negative rather than LEE positive. In other words, a strain lacking *cscRAKB* was also likely to lack the LEE but a *cscRAKB* positive strain was equally as likely to be LEE positive or negative (Figure 5-6AB). Unsurprisingly, a reciprocal pattern of D-serine versus sucrose utilisation gene carriage was observed. 568 (35.7%) and 909 (57.1%) of isolates were *cscRAKB* negative/*dsdCXA* positive and *cscRAKB* positive/*dsdCXA* negative respectively. 87 (5.5%) isolates were negative for both traits and only 27 (1.7%) were positive for both. Thus, the likely scenario is for an isolate to carry one or the other and rarely ever both or neither (Figure 5-6C).

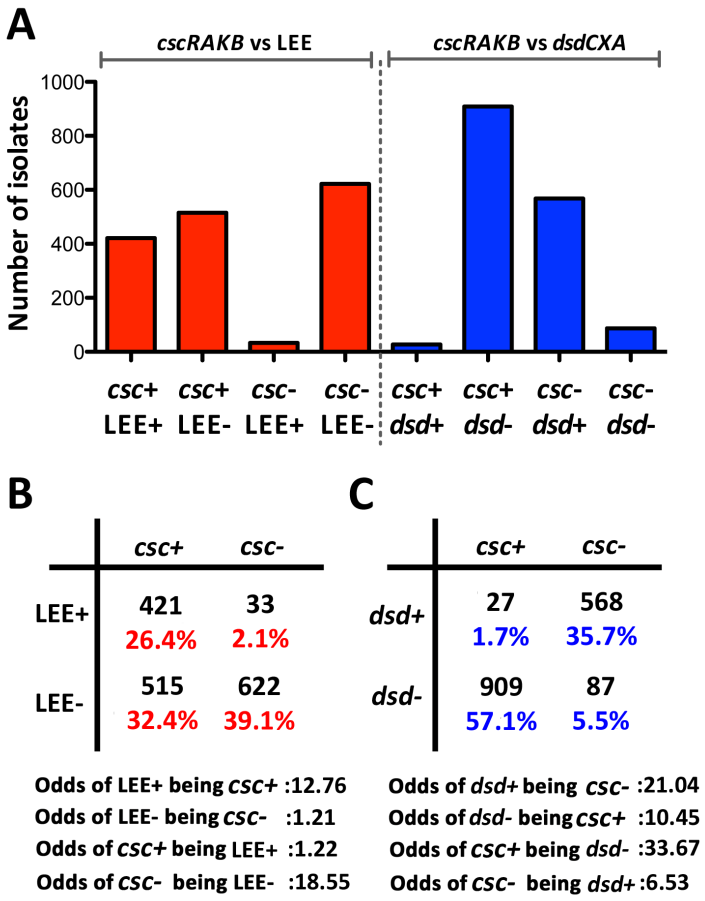

Figure 5-6 Carriage of the *cscRAKB* locus in *E. coli.* (A) Bar chart visually representing the distribution of LEE, *cscRAKB* or *dsdCXA* carriage among 1591 *E. coli* strains, in the context of *cscRAKB* vs. LEE and *cscRAKB* vs. *dsdCXA* indicated in red and blue respectively. A + denotes presence whereas a – denotes absence. 2 x 2 contingency matrices showing the relationship between LEE+, LEE-, *csc*+ or *csc*- isolates (B) and *dsd*+, *dsd*-, *csc*+ or *csc*- isolates (C). The number of strains for each scenario as well as their percentage distribution among the total 1591 isolates tested are indicated in black, red and blue, with the latter two corresponding to panel A respectively. The odds ratios calculated for each scenario are indicated also.

These data are restricted by the fact that they report trends based on presence and absence and do not offer insight into gene functionality, however the results are all in keeping with the hypothesis that *E. coli* strains acquiring the LEE horizontally are much more likely to lose *dsdCX* and replace it with *cscRAKB.* Little is known about how the *cscRAKB* locus is regulated or even its precise role other than as a metabolic accessory. Indeed, a closer look at the presence/absence data reveals that the genomic region between *argW* and the *cscRAKB* integration site is not well conserved and considered hypervariable (Moritz & Welch 2006). The functionality of an incomplete *cscRAKB* locus is also unknown. It is also interesting to note that sucrose utilisation does not seem to be

required for intestinal pathogenesis. The prototypical EAEC strain O42 was found to be LEE/*dsdCXA*/*cscRAKB* negative (Moritz & Welch 2006, Okhuysen & Dupont 2010). Indeed, closer inspection of its genome reveals notable degeneracy when compared to other prototypical *E. coli* strains at the *argW* integration site (Chaudhuri *et al.,* 2010). To assess if sucrose metabolism plays an unsuspecting role in T3S, a pLEE1:GFP reporter assay was performed in MEM-HEPES supplemented with increasing concentrations of sucrose. No significant shift in LEE1 expression was observed in the presence of sucrose and normal housekeeping gene expression (monitored by p*rpsM*:GFP) was observed also (Figure 5-7). This suggests that the acquisition of *cscRAKB* is concurrent with LEE acquisition but not a requirement for A/E mediated pathogenesis.

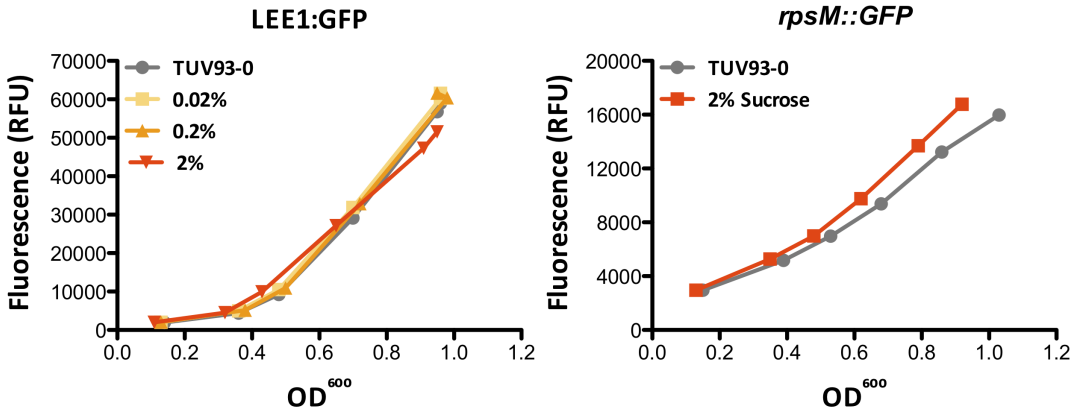

Figure 5-7 Sucrose does not affect T3S in TUV93-0. Transcriptional reporter assays using LEE1 promoter and *rpsM* promoter GFP-fusion constructs to monitor expression as RFU. Bacteria were cultured in MEM-HEPES supplemented with 0.02%, 0.2% or 2% sucrose and monitored for fluorescence against OD_{600}. Experiments were performed in triplicate with similar results recorded each time.

5.3 The prevalence of *yhaOKJ* across the *E. coli* phylogeny

Chapter 3 discussed the proposal of the *yhaOMKJ* locus to be involved in sensing of D-serine and gene regulation in response to the environment. Data presented in chapter 4 and 5 together propose that D-serine metabolism is not required and the context of niche specification is actually detrimental for expression of the LEE. However, the ability to sense and respond to D-serine is of the utmost importance, especially for InPEC strains

carrying the LEE, and therefore would be expected to be a well-conserved trait. Indeed, D-serine has been proposed to act as both a positive fitness trait and a signal transduced for regulation of virulence factors in UPEC. Furthermore, the regulator of the D-serine tolerance locus, DsdC, was found to not play a direct role in modulation of virulence gene expression in response to D-serine (Anfora *et al.,* 2007). This suggests that other systems are also capable of responding to this metabolite to assist in controlling expression of key virulence genes. One such system is *yhaOMKJ.* Several lines of evidence support the hypothesis that genes of this locus are both regulated by and involved in the response to D-serine exposure and also nitric oxide stress. However, the system is not well characterised and the prevalence of these genes is currently undefined. Therefore, the presence/absence approach was applied to the sequences of YhaO, YhaK and YhaJ in order to assess the carriage of these genes and potential role as a ubiquitous sensory system in *E. coli.* YhaM was not included in this analysis as its physiological role is unclear and it was demonstrated in Figure 3-2 to be co-transcribed with *yhaO*.

It was immediately apparent that *yhaOKJ* is highly conserved across the *E. coli* phylogeny. Of the 1591 strains analysed only 23 (1.4%) tested absent for some or all three of the *yha* genes (Table 8-4). Four, two and seven isolates were missing *yhaJ, yhaK* and *yhaO* respectively, a further two were missing *yhaJ* and *yhaK*, and only 8 isolates in total (0.5%) tested absent for all three genes. The distribution of these isolates across the phylogenomic tree was quite undefined with absence being reported in groups A, B1. B2, C, D and E. It is interesting to note that although absence of *yhaOKJ* occurred on a small number of occasions in phylogroups E, B1 and B2 (the predominant EHEC, EPEC and EPEC-containing ExPEC groups respectively), absence of these genes formed a trend in phylogroups D and A. More importantly, only two of the observed absences are concurrent with LEE positive strains (phylogroup E). Given that YhaO and YhaJ are essential for LEE expression and subsequent T3S mediated A/E pathogenesis (discussed in chapter 3) this may reflect the power of InPEC strains to retain these key genes. However, the overall distribution of *yhaOKJ* absence is so low that this observation might simply be coincidental. It is worth noting that five of the 23 isolates lacking members of *yhaOKJ* were classified as *Shigella* sp. The role of YhaO and YhaJ in *Shigella* associated virulence was not explored as part of this work.

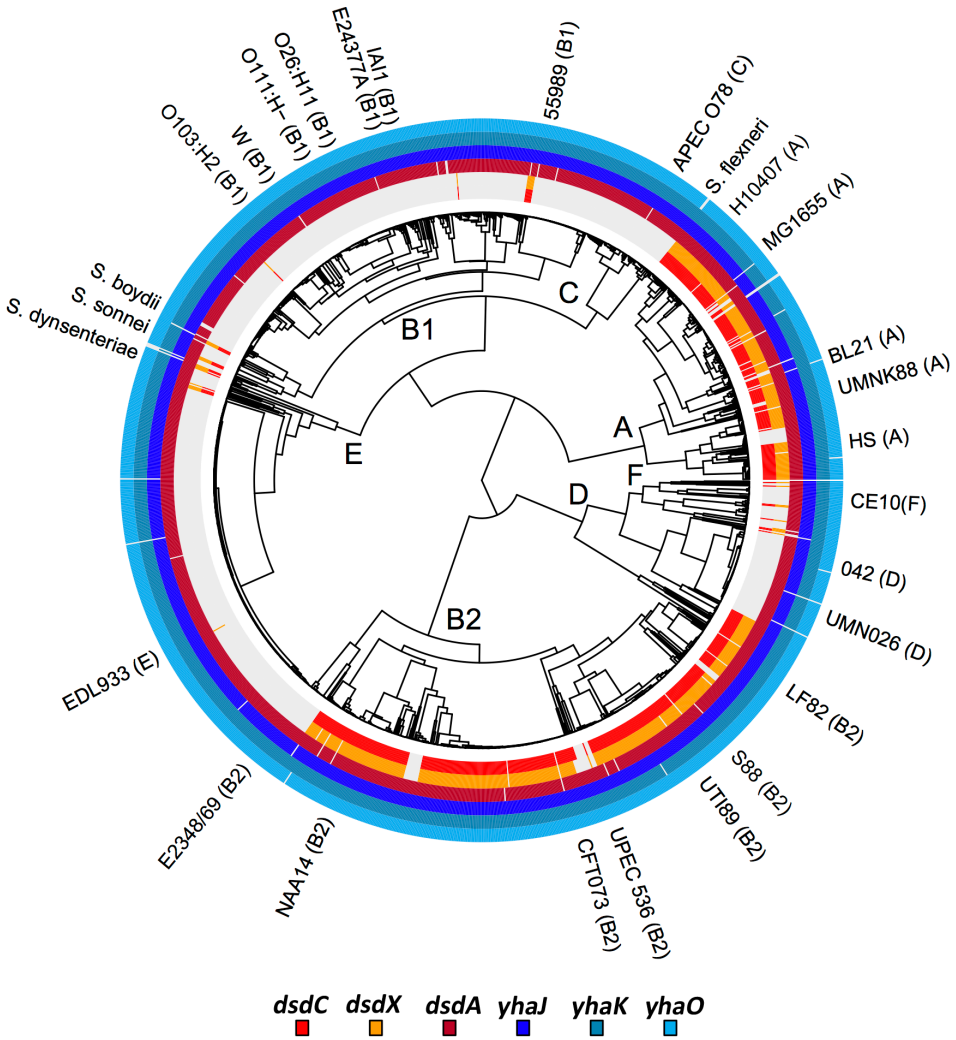

Figure 5-8 Distribution of *yhaOKJ* across the *E. coli* phylogeny. Circularised phylogenomic tree of 1591 *E. coli* and *Shigella* isolates overlaid with the gene carriage of the *yhaOKJ* locus and the *dsdCXA* locus. Phylogenetic sub-grouping A, B1, B2, C, D, E F and the location of selected *Shigella* isolates is indicated on the circularized dendrogram at the points of branch divergence. A colored block indicates that the gene in question was considered "present" (>50% identity over >50% of the coding sequence) whereas a light grey block denotes that a gene was considered "absent". Color-coding of gene blocks is as follows: *dsdC* = red; *dsdX* = yellow; *dsdA* = auburn; *yhaJ* = blue; *yhaK* = aqua; *yhaO* = cyan. Prototypical strains used to identify phylogroups are indicated outside.

5.4 D-serine and other *E. coli* pathotypes

The data presented in chapter 4 provides a detailed framework of how D-serine is sensed and used by EHEC to modulate expression of virulence factors. The implications of D-serine uptake on activation of stress responses and cell wall reorganisation were also demonstrated for both EHEC and UPEC experimentally. Furthermore, comparative genomic analysis of 1591 *E. coli* isolates from across the entire phylogeny suggest that the data presented in chapter 4 is broadly applicable to all *E. coli* irrespective of their niche, commensal or pathogenic nature. Collectively, it can be therefore hypothesised that D-serine acts as a major niche determinant capable of restricting certain pathogens to certain areas of the body by shifting gene expression either in favour or against colonisation selectively. In order to support these results the effects of D-serine on growth and virulence was investigated for two further distinct pathotypes – the LEE positive intestinal pathogen E2348/69, which belongs to the EPEC pathotype, and the recently investigated MNEC isolate CE10.

5.4.1 *D-serine represses the LEE in the EPEC strain E2348/69*

The research presented focused on the LEE encoded T3SS, a major virulence factor for both EHEC and EPEC. Indeed, the LEE is employed by both pathogens to mediate intestinal colonisation however major differences in regulation of the LEE have been observed for both EHEC and EPEC, being attributed to distinct genetic backgrounds and carriage of unique prophages (Iguchi *et al.,* 2009, Mellies *et al.,* 2007a, Perna *et al.,* 2001). The work described in chapters 3 and 4 was carried out in the context of EHEC LEE regulation however exposure to the EPEC strain E2348/69 to D-serine under LEE inducing conditions resulted in a similar phenotype observed for EHEC, downregulation of the LEE-encoded T3SS with minimal effects on growth in MEM-HEPES (LEE inducing conditions) whereas growth in M9 minimal media supplemented with D-serine inhibited growth dramatically (Figure 5-9AB). These results provide strong evidence in favour of the hypothesis that D-serine represses the LEE irrespective of pathotype or phylogroup by providing further experimental data to support the observations made from Figure 5-1.

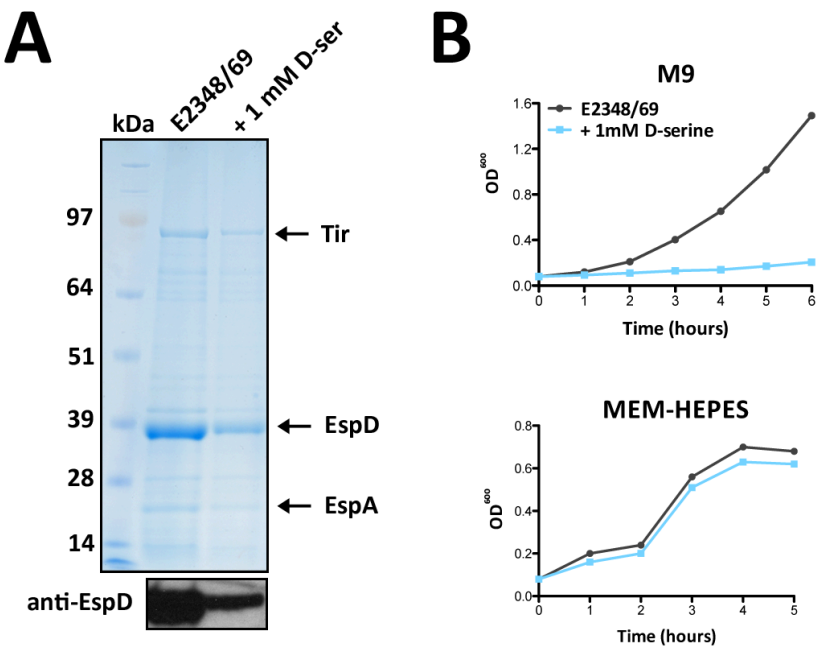

Figure 5-9 The effects of D-serine on EPEC T3S and growth. (A) SDS-PAGE analysis of the secreted protein profile for EPEC (E2348/69) grown in MEM-HEPES with and without 1 mM D-serine as a supplement. Cultures were obtained at an OD_{600} of 0.6 and the secreted proteins were precipitated from the supernatant using 10% TCA. Western blot analysis of EspD confirmed the reduction in the T3S profile for the D-serine supplemented sample. This experiment is representative of three independent replicates. (B) Growth analysis of E2348/69 in M9 and MEM-HEPES plotted as OD_{600} versus time. E2348/69 is indicated in grey and the D-serine treated sample indicated in blue. Growth experiments were performed in triplicate with similar results obtained each time.

5.4.2 *The MNEC isolate CE10 and its response to D-serine*

Another T3SS has emerged in the literature and its role in virulence is being uncovered. The ETT2 system is a T3SS that shares similarity to the SPI system of *Salmonella enterica* and was identified during sequence analysis of the first O157:H7 genome (Perna *et al.,* 2001). Since then it has been documented that the ETT2 system is present to some extent in most *E. coli* strains irrespective of pathogenicity, except those found in phylogroup B2, but has undergone widespread attrition therefore making its role in virulence somewhat difficult to characterise (Ren *et al.,* 2004). Interestingly, strains carrying incomplete ETT2 loci are still attenuated for virulence upon further mutation of the system and regulators encoded within the ETT2 locus in EHEC were shown to influence regulation of the LEE (Ideses *et al.,* 2005, Zhang *et al.,* 2004). The prototypical EAEC strain O42 was found to have a full, uninterrupted ETT2 that may play a role in colonisation. The research identified a putative master regulator of ETT2, EilA, which co-ordinately controls

expression this T3SS and proposed effectors (Sheikh *et al.,* 2006). Recently, the completed genome sequence of MNEC strain CE10 revealed the presence of a functional ETT2 system that was proposed to play a role in invasion and survival within human brain microvascular endothelial cells (Lu *et al.,* 2011, Yao *et al.,* 2009). Given that the brain is a rich source of D-serine and the knowledge that MNEC strains commonly carry two copies of *dsdCXA* and not the LEE, the role of D-serine in MNEC pathogenesis was assessed using CE10 as the model organism (Moritz & Welch 2006). Indeed, CE10 carried two copies of *dsdCXA* and was capable of growing on D-serine as a sole carbon source (Figure 5-10A). The appearance of colonies on MOPS minimal media plates supplemented with D-serine was actually more rapid than observed for UPEC CFT073. Within 12 hours colonies were appearing visible for CE10 whereas CFT073 required 18-24 hours. This suggests and enhanced ability to catabolise D-serine in strains carrying two copies of *dsdCXA.* Growth in MEM-HEPES, M9 and LB liquid culture did not indicate a growth advantage for CE10 when supplemented with D-serine in this media however MNEC strains are highly adapted pathogens and these media may not reflect their preferred environment however these results confirm an inability of supplementary D-serine to inhibit growth of CE10 (Figure 5-10B).

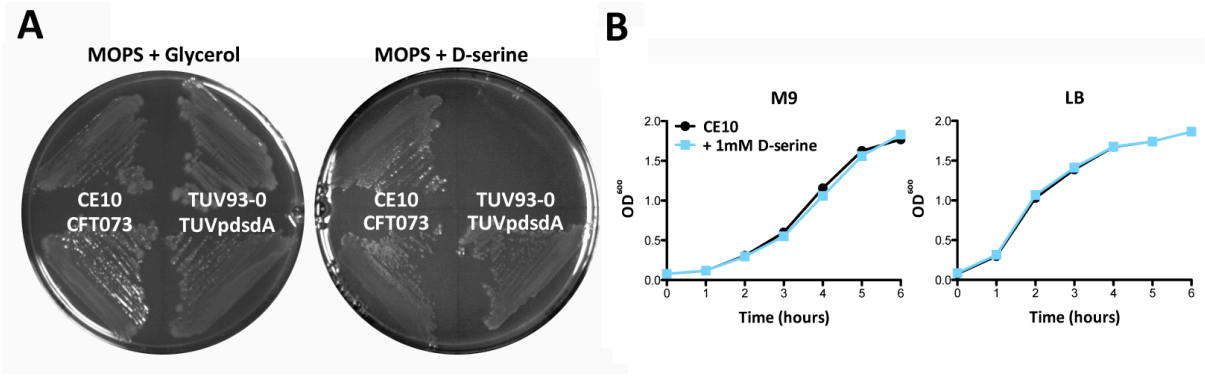

Figure 5-10 Growth profile of *E. coli* CE10 in the presence of D-serine. (A) Growth of CE10 on MOPS minimal media agar plates supplemented with Glycerol and D-serine. Plates were divided into quadrants and streaked with CE10 (MNEC), CFT073 (UPEC), TUV93-0 (EHEC) and TUVp*dsdA* as labelled. Plates were incubated at 37°C static for 24 to 48 hours. (B) Growth curves for CE10 cultured in minimal media (M9) and rich media (LB) supplemented with 1 mM D-serine. Graphs are plotted as OD_{600} over time (hours) and are representative of three biological replicates.

D-serine was capable of downregulating the LEE in both EHEC and EPEC. A role for regulation of virulence factors by D-serine in diverse uropathogens has also previously been proposed (Anfora *et al.,* 2007, Korte-Berwanger *et al.,* 2013). It was therefore postulated that the ETT2 might be capable of responding to endogenous D-serine to potentially promote colonisation of the brain by this adapted pathogen. To test this a reporter of *eilA,* the proposed ETT2 master regulator, was generated (Sheikh *et al.,* 2006). An ~500bp DNA fragment upstream of the *eilA* promoter in CE10 was amplified by PCR containing restriction flanks complimentary to the cloning site of pAJR70. This plasmid contains GFP in frame with a cloning site designed to accept gene promoter fragments and was constructed in a previous study to assess transcription of the LEE operons through fluorescence monitoring from liquid culture, as demonstrated in chapter 4 (Roe *et al.,* 2003). The *eilA* upstream fragment was then cloned into pAJR70 (with the assistance of James Mordue, an undergraduate trainee) to obtain the p*eilA*:GFP reporter plasmid, which was confirmed by PCR and sequencing. The EilA regulator has not been studied in CE10 and the conditions that promote its expression were therefore unknown. A study on *eilA* expression in EAEC O42 defined the optimal expression conditions as a 1:1 mix of LB and DMEM supplemented with 0.2% glucose. This was therefore used as a starting point to analyse if *eilA* is capable of being expressed in CE10 under these conditions. As a comparator, *eilA* expression was tested in LB alone and DMEM + 0.2% glucose also. Bacterial cultures were grown aerobically at 37°C shaking and samples were taken over the course of two days to allow analysis of *eilA* expression in both exponential and stationary phase. Aliquots were measured for OD_{600} and GFP expression, with the background fluorescence of an p*eilA*:GFP negative CE10 strain being measured also and corrected for. It was immediately apparent that *eilA* was capable of being expressed in CE10 as expression levels increased over time in all media tested (Figure 5-11). Expression was low in all conditions tested after 8 hours but prolonged stationary phase incubation resulted in increased expression of *eilA.* Culture in LB was the most favourable condition for expression of *eilA,* which is in contrast to the O42 study (Sheikh *et al.,* 2006). Supplementation of 1 mM D-serine to the growth medium did not affect *eilA* expression in LB or DMEM but a consistent trend of enhanced *eilA* activity was observed in LB:DMEM. The enhanced expression was measureable at 24 hours and increased thereafter. Although the shift in *eilA* expression was not significant, the results suggest that D-serine may enhance ETT2 associated virulence of CE10 under the appropriate conditions. It is important to reiterate that O42 is an InPEC isolate, which is in contrast to

CE10, and these conditions tested may not be optimal for expression of ExPEC virulence factors. Nonetheless, the data confirm that *eilA* is capable of being expressed in CE10 and given the importance of the ETT2 system for its pathogenesis should be the focus of future studies on regulation of this virulence factor (Yao *et al.,* 2009).

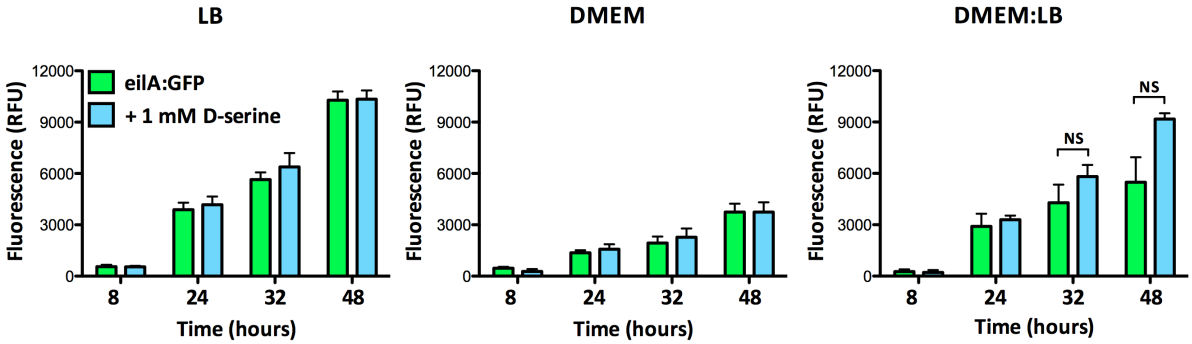

Figure 5-11 Expression of the ETT2 master regulator *eilA* in CE10. The p*eilA*:GFP transcriptional reporter construct was introduced into the *E. coli* CE10 background and monitored for its expression in LB, DMEM (+0.2% glucose) and 1:1 LB and DMEM (+0.2% glucose). Cultures were incubated at 37°C, 200 rpm and sampled at 8, 24, 32 and 48 hours for GFP expression (RFU). Expression of *eilA*:GFP in media alone or media supplemented with 1 mM D-serine is indicated by green and blue bars respectively. Experiments were performed in triplicate. NS denotes no significant difference.

5.5 Discussion

Exposure of *E. coli* O157:H7 to D-serine caused a distinct pattern of gene expression – inhibition of virulence by repression of the LEE and activation of the SOS response by a novel mechanism. It was therefore postulated that the importance of this phenotype might affect carriage of the D-serine tolerance locus (*dsdCXA*), which is truncated beyond functionality in LEE positive pathogens, for two rather distinct reasons. Firstly, supplying a constitutively expressed DsdA *in trans* to TUV93-0 did not overcome the repression exhibited on the LEE and allowed metabolism of D-serine as a carbon source, hence the T3SS is not functional in a D-serine rich environment. This suggested that carriage of the *dsdCXA* locus in full would offer little advantage for LEE positive *E. coli* pathotypes. Secondly, supplying TUV93-0 with a functional DsdA eliminated the intracellular accumulation of D-serine and thus prevented the activation of the SOS response thereby fundamentally shifting the gene expression profile of the pathogen. It was hypothesised that these two loci were therefore functionally incompatible and offer no advantage to one another. These results prompted investigation into the carriage of both the LEE and *dsdCXA* amongst a large range of sequenced *E. coli* genomes to address this incompatibility.

5.5.1 *Carriage of both the LEE and dsdCXA is a significantly rare event*

The frequency of LEE and *dsdCXA* carriage was investigated by applying a presence or absence approach to the genome sequences of 1591 *E. coli* isolates that were available publically. The results of this analysis were applied to the *E. coli* phylogeny based on diversity of the core genome and revealed a clear distinction in strains that carry the LEE PAI and those that have retained the ancestral *dsdCXA* locus in full. This was consistent with the theory that the two genetic elements are functionally incompatible and corroborated the experimental data gained for TUV93-0 (EHEC) and CFT073 (UPEC) during the course of this project. The specific loss of D-serine tolerance by way of a conserved truncation in *dsdX* has been documented previously and was believed to simply be an adaptation of InPEC strains to life in the gastrointestinal tract as the loss of *dsdCX* is concurrent with the gain of the *cscRAKB* locus for sucrose utilisation in most cases (Jahreis *et al.,* 2002, Moritz & Welch 2006). This event is independent of LEE

acquisition however, as the data presented here strongly indicate that replacement of *dsdCX* with *cscRAKB* is just as likely to occur in a LEE negative strain as in a LEE positive strain. This trend is also supported by the fact that the same likelihood can be applied to the maintenance or loss of *dsdCX* in a LEE negative background. Thus, commensal *E. coli* and non-LEE pathogens that occupy a niche within the gastrointestinal tract can maintain the ability to utilise D-serine.

Clearly the prevalence of *dsdCXA* among ExPEC isolates is an indicator of the necessity to maintain D-serine tolerance for extraintestinal niche adaptation (Moritz & Welch 2006). This is a tangible hypothesis and certainly agrees with data presented here on LEE positive strains being statistically more likely to lose *dsdCX.* However, the ability to metabolise sucrose is an accessory to *E. coli* and therefore not essential for life in the gastrointestinal tract *per se*. Indeed, an isolate that did not carry the LEE was just as likely to lose *dsdCX* irrespective of whether they were commensal or pathogenic. This was exemplified here by the enhanced ability of non-pathogenic *E. coli* K-12, which carries the full *dsdCXA* locus, to grow in the presence of D-serine. The presence/absence phylogenomic tree presented here illustrates that loss of *dsdCXA* is sporadic but is nearly always concurrent with LEE acquisition, irrespective of phylogroup. The most striking example of this was the appearance of LEE positive EPEC strains in the largely ExPEC dominated group B2. Phylogroup B2 members are evolutionarily distinct from the largely LEE positive groups E and B1 but the acquisition of the LEE even in this genetic background results in the generation of dedicated InPEC isolates, such as the well studied EPEC strain E2348/69 (Iguchi *et al.,* 2009). A/E associated pathogens have been recently documented to occur at multiple positions in the *E. coli* phylogeny, being found to some extent in nearly every phylogroup (Hazen *et al.,* 2013). Data from the study presented here corroborates this but provides strong evidence that this acquisition of the LEE nearly always results in subsequent loss of *dsdCX.* It seems that the ability to metabolise either D-serine or sucrose is dispensable depending on an individual's particular niche. All *E. coli* pathogens must successfully colonise that gastrointestinal tract in a competitive manner (Meador *et al.,* 2014). Indeed, UPEC strains are known to be extremely competitive in the intestine prior to dissemination into the urinary tract (Chen *et al.,* 2013). This argues for the maintenance of *dsdCXA.* Exposure to sucrose had no effect on expression of the LEE but D-serine was found to be a potent negative regulator of this seemingly niche specific virulence factor. Therefore, a hypothesis was proposed whereby the selective pressure

against expression of the LEE in a D-serine rich environment restricts LEE-positive pathogens to the gastrointestinal tract and also selects against ExPEC strains acquiring the LEE.

There are some rare exceptions to this hypothesis that were identified in this work. Some 26 of the 1591 isolates analysed (1.6%) tested positive for both the LEE and *dsdCXA,* falling into various phylogroups (A, B1, B2, D and E). These isolates are not well known and largely of faecal origin and none were isolated form ExPEC associated infections. The interpretation of these isolates based on the working hypothesis proposed in this study predicts that these strains have acquired the LEE without opportunity for subsequent selective pressures to result in the loss of *dsdCX.* Furthermore, five more uncharacterised B2 isolates were identified as being LEE positive but *dsdX* negative indicating isolates that may be in the process of adapting their genomes to carriage of the LEE.

Taken together, the work presented in chapter 5 provides significant evidence for the hypothesis of a functional incompatibility between the LEE and the *dsdCXA* locus as first identified in chapter 4. This work provides novel insights into niche specification in *E. coli* and why certain pathotype are restricted to certain areas within the host, irrespective of their ability to scavenge nutrients from that niche.

5.5.2 *The yhaOMKJ locus is highly conserved in E. coli*

The role of YhaO and YhaJ was explored in chapter 3 of this thesis. YhaO, a D-serine transporter, was found to be essential for full activation of the LEE in O157:H7 as was the LTTR YhaJ. YhaO is transcriptionally under the control of YhaJ and the two formed part of an operon, along with the regulatory cofactor YhaK and a deaminase YhaM, which was predicted to be involved in D-serine sensing. Historically, this system may have been involved in D-serine metabolism however even though YhaO can successfully internalise D-serine, YhaM cannot metabolise it effectively resulting in intracellular accumulation and modulation of gene expression. This is consistent with the fact that O157:H7 cannot grow on D-serine as a sole carbon source. However, work described here provide evidence for an alternative role of the *yhaOMKJ* system in regulation of virulence, suggesting an adaptation of the pathogen to utilise otherwise redundant genes located on its chromosome. Due to its ability to at least transport D-serine and the knowledge

that D-serine is capable of regulating the LEE, D-serine was proposed to be a substrate for this system allowing the pathogen to monitor its environment and respond accordingly.

Carriage of the genes *yhaO, yhaK* and *yhaJ* was highly conserved across the *E. coli* phylogenomic tree presented here. Less than 1.5% of the 1591 isolates examined are lacking one of the three genes and only 0.5% is absent for all of them. It seems counter intuitive for *E. coli* to maintain a second D-serine metabolic operon considering the conserved presence/absence genotype observed for the *dsdCXA* system, further bolstering the theory that this operon has evolved and adapted to have alternative roles in the cell. This agrees with the "use it or lose it" approach of bacterial genome minimalism (Moran 2002). The idea of gene adaptation is elegant and offers an efficient alternative to gene deletion or functional mutation. The precise function of YhaO and YhaJ is still under investigation however their role in regulation of the LEE is unquestionable. These genes have not been characterised prior to this work (with the exception of a crystal structure for YhaK) but have appeared in the literature in independent studies as being actively regulated by NO donors, human urine, extracytoplasmic stress and by Ler in EPEC (Bingle *et al.,* 2014, Bury-Moné *et al.,* 2009, Gurmu *et al.,* 2009, Hancock & Klemm 2007, Pullan *et al.,* 2007). Furthermore, these studies have also involved non-LEE associated *E. coli* strains suggesting that regulation of the LEE by YhaJ is an EHEC and/or EPEC specific adaptation and that other commensals or pathotypes have evolved numerous roles for these genes. Interestingly, a large study aimed at identifying the most conserved genome segments across Bacteria, Archaea and the Eukarya revealed *yhaO* as falling within the criteria they proposed (Isenbarger *et al.,* 2008). Other genes identified were amino acid metabolic genes and, unsurprisingly, ribosomal RNA encoding genes. Seeing as serine is considered a non-essential amino acid and that D-serine specifically is toxic to a wide range of bacteria but also a physiologically relevant metabolite it seems appropriate to conserve an ability to respond to this D-AA. Recently, two studies investigating pseudogene accumulation in the independent *S. enterica* serovars Typhi and Paratyphi A identified *yhaO* in their analyses (Holt *et al.,* 2009, McClelland *et al.,* 2004). It is interesting to note that these serovars have evolved to cause systemic disease whereas *Salmonella* species that cause more classical gastroenteritis have retained *yhaO.* This thesis proposes D-serine as a key niche determinant for *E. coli,* restricting dedicated enteric pathotypes to the intestine, and that YhaO is employed in D-serine sensing. It is intriguing to propose that a pathogen that has

evolved to cause a more systemic disease, such as Typhi or Paratyphi A, may have lost the necessity to sense metabolites such as D-serine through YhaO, thus minimising their genomes accordingly.

The specific role for *yhaOMKJ* is still under investigation but contribution of this system to the regulation of virulence in O157:H7 is without question. This work has begun to unravel how the LEE is controlled by YhaJ and provides a strong framework for further study on this topic. Understanding how virulence factors are regulated naturally is critical to allow rational design of novel inhibitors that can be used to limit bacterial infections, an area of particular relevance for O157:H7 as antibiotic use employed to tackle this pathogen is controversial due to increased risk of Stx expression (McGannon *et al.,* 2010). Given that half of all current drug targets are membrane proteins, YhaO is therefore an attractive target for such investigation due to its conserved nature and apparent role in pathogenesis (Drews 2000). Targeting virulence factors allows "disarming" of pathogens rather than killing by antibiotics and thus promotes a natural clearance from the body, a strategy thought to reduce resistance to therapeutics (Allen *et al.,* 2014, Keyser *et al.,* 2008).

5.5.3 *D-serine and other diverse pathotypes*

Collectively, this thesis provides a convincing hypothesis that the host metabolite D-serine has significant implications for the regulation of virulence and stress responses in O157:H7. Furthermore, comparative genomic analysis suggests that these effects apply to the entire *E. coli* phylogeny, irrespective of origin or niche. It is therefore important to address the effects of D-serine on other pathotypes not discussed earlier so as to bolster this theory further.

The negative effect of D-serine on regulation of the LEE in EHEC has been investigated thoroughly in chapter 4. Chapter 5 suggested that members of the EPEC pathotype are subject to the same selective pressure of D-serine and therefore have undergone the same pattern of LEE gain in conjunction with *dsdCX* loss. The LEE of EHEC and EPEC are incredibly similar with subtle differences in the regulation of this PAI being as a result of different genetic backgrounds. However, this is not unsurprising as multiple regulatory systems coordinate to control the LEE in response to diverse inputs (Mellies *et al.,* 2007a).

The resulting pathogenesis via A/E lesions however is highly comparable with downstream host cell subversion largely being attributable to NLE input, which is known to differ greatly between the two pathotypes (Tobe *et al.,* 2006). Therefore, it can be assumed that irrespective of these differences D-serine is still capable of acting as a negative environmental signal restricting EPEC pathogens to the intestine also. Indeed, the effects of D-serine on EPEC *in vitro* were indistinguishable from that observed for EHEC. The EPEC strain E2348/69 displayed similar growth characteristics to TUV93-0 in media supplemented with D-serine, an inability to grow on D-serine as a sole carbon source and inhibition of the LEE encoded T3SS. This provides strong evidence that the effects of D-serine are applicable to distinct A/E pathogens. It is likely that each pathotype will regulate the response differently depending on their genetic background. For instance, YhaJ is predicted to regulate the LEE directly through an EHEC specific promoter sequence but indirect regulation of the LEE in response to D-serine is also apparent through conserved transcriptional regulators that are hijacked to elicit the response, for instance IHF and the GAD regulators (Friedberg *et al.,* 1999, Tree *et al.,* 2011).

It was clear that diverse A/E pathogens were influenced heavily by D-serine as an environmental signal. However, the role of D-serine in the regulation of uropathogens is a different story, whereby D-serine acts as a positive fitness trait and regulator of virulence in both UPEC and *S. saprophyticus* (Anfora *et al.,* 2007, Korte-Berwanger *et al.,* 2013). What is less explored however is the regulation of virulence for alternative ExPEC strains. MNEC is capable of penetrating the BBB and disseminating to the brain where it causes infection. MNEC isolates do not encode a LEE T3SS, a theme in keeping with other ExPEC, however they do encode the ETT2 system (Lu *et al.,* 2011). Recently the role of ETT2 in virulence of diverse *E. coli* pathotypes has been explored and the system has specifically been identified as being intact and important for host cell invasion in the MNEC isolate O7:K1 CE10 (Yao *et al.,* 2009). On this basis, one would expect that the ETT2 system is functional in the presence of D-serine and regulated very differently to the LEE encoded T3SS. In this way, MNEC strains have adapted to both tolerate and metabolise D-serine whilst still retaining a functional T3SS for delivery of effector proteins into host cells. CE10 was capable of growing on D-serine as a sole carbon source effectively in a manner similar to UPEC isolates. This trait was attributed to the two copies of *dsdCXA* encoded on its genome (Lu *et al.,* 2011, Moritz & Welch 2006). Additionally to D-serine tolerance, the expression of the ETT2 master regulator *eilA* was not negatively affected by exposure to

this D-AA. This is in stark contrast to the effects characterised for Ler, the LEE master regulator. Characterisation of *eilA* expression conditions were indeed preliminary but supplementation of D-serine into LB:DMEM, previously described as being the optimal conditions for ETT2 expression in *E. coli* O42, resulted in an increase in *eilA* expression (Sheikh *et al.,* 2006). This trend was consistent although not significant, likely due to the conditions being sub-optimal for *eilA* expression in a CE10 specific background. This does however suggest that D-serine may be capable of enhancing pathogenesis of CE10 through positive modulation of the ETT2 system. Indeed, positive modulation of virulence factors by D-serine exposure has been characterised in the uropathogen *S. saprophyticus* (Korte-Berwanger *et al.,* 2013). More recently the prototypical MNEC strain RS218 was demonstrated to have impaired interactions with human brain microvascular endothelial cells in a mutant deleted for D-serine catabolism, a phenotype similar to mutants deleted for known MNEC adhesins, invasins and toxins, further demonstrating the absolute importance of D-serine tolerance for ExPEC isolates in causing disease (Yousuf *et al.,* 2014).

6 Final discussion and outlook

The LEE encoded T3SS is an essential colonisation factor for the A/E pathogens EHEC, EPEC and *C. rodentium* (McDaniel *et al.,* 1995). This system, which is encoded on an ~35 kb horizontally acquired PAI, has been integrated into the chromosome of these pathogens and is indispensible in their pathogenesis. The core role of the LEE and its 41 ORFs is largely well characterised, encoding all the necessary machinery to assemble a functional T3SS as well as the primary secreted effector proteins and their cellular chaperones, allowing LEE encoding bacterial pathogens to intimately attach to and colonise host cells all the while subverting their normal function (Büttner 2012). The LEE is also regulated in the first instance by two systems encoded within this PAI – Ler and GrlRA (Deng *et al.,* 2004, Mellies *et al.,* 1999). However, transcriptional control of the LEE is extremely complex and is regulated by numerous systems of broad and narrow spectrum function within the cell. Conserved chromosomally encoded regulators of the nucleoid, stress response systems and global transcriptional regulation have essential roles in helping to modulate the LEE accordingly. Additionally, numerous plasmid and prophage encoded regulators that are more pathotype specific have emerged as playing equally important roles in LEE regulation (Mellies *et al.,* 2007a, Tree *et al.,* 2009). These complex layers of regulatory input to the LEE have been cumulatively studied for nearly two decades painting a picture of how diverse regulatory systems adapt to elicit control over horizontally acquired elements. Furthermore, how the cell specifically responds at the molecular level to physiologically relevant signals perceived from the environment is now a more recognised theme within the field of LEE regulation, allowing a deeper understanding of how functional genetic elements relate to behaviour within a desirable niche.

The emerging role that NCDAAs play in diverse bacteria is an under-researched area. It was recently discovered that unusual D-AAs are actively produced and released into the surrounding environment for a specific purpose, in particular cell wall regulation in response to growth phase by *V. cholerae* and *B. subtilis* (Cava *et al.,* 2011a, Lam *et al.,* 2009). This is logical considering the integral part that D-alanine and D-glutamate play in formation of PG (Typas *et al.,* 2012). What are less understood are the roles that D-AAs can play in other cellular processes that do not directly relate to cell wall biogenesis and regulation. D-serine is known be involved in cell wall modification to facilitate vancomycin resistance in species of the *Enterococcus* but exogenous D-serine encountered in the environment has also been implicated to have specific functions in bacteria (Arias *et al.,*

1999). D-serine is a metabolite of both exogenous and endogenous source to the host body. Studies on processing of food sources have revealed widespread racemisation of D-AAs with D-serine being reported to be found in high concentrations in certain foods providing potentially abundant exogenous sources of this D-AA depending on the diet (Man & Bada 1987). Conversely, D-serine is known to be produced within the brain where it acts as an essential co-agonist of neurotransmitters (Mothet *et al.,* 2000, Wolosker *et al.,* 2008). Human or animal urine is also an abundant source of D-serine with reported concentrations in the mM range and it is pathogens of the urinary tract niche that have been studied in relation to D-serine (Anfora *et al.,* 2007). D-serine catabolism has been extensively described as a positive trait of fitness for the uropathogens UPEC and *S. saprophyticus.* Furthermore, the metabolite was also demonstrated to affect the regulation of virulence factors in these pathogens (Anfora *et al.,* 2007, Haugen *et al.,* 2007, Korte-Berwanger *et al.,* 2013, Roesch *et al.,* 2003). Seeing as the mechanisms of virulence used by these bacteria are distinct, it was therefore hypothesised that D-serine may act a regulator of virulence in many bacterial species that may encounter D-serine at numerous sites within the body, not only uropathogens.

The work presented in this thesis shows for the first time that the host metabolite D-serine is capable of modulating the virulence potential of an intestinal pathogen, *E. coli* O157:H7. This InPEC strain utilises a LEE encoded T3SS to colonise the human intestinal epithelium and establish dangerous and potentially fatal infection (Croxen & Finlay 2010). It was revealed that exogenous D-serine affects gene regulation on a global scale without any loss of bacterial fitness or viability. Indeed, D-serine has previously been described as being bacteriostatic to *E. coli* that lack a fully functional *dsdCXA* system, including O157:H7, in minimal media but this work revealed that growth is not inhibited in more physiologically relevant conditions and that D-serine can actually be used by O157:H7 as a carbon source upon acquisition of a functional DsdA *in trans* (Connolly *et al.,* 2015, Cosloy & McFall 1973, Moritz & Welch 2006). The most important finding of this work however was that irrespective of D-serine metabolism and associated fitness, this metabolite was capable of inhibiting expression of the LEE at the transcriptional level subsequently alleviating the ability to form A/E lesions on host cells (Connolly *et al.,* 2015). This revealed the role of D-serine as an environmental regulator of virulence in diverse pathotypes.

What is the physiological relevance of this finding? It is unsurprising that InPEC pathotypes such as O157:H7 have lost the ability to metabolise D-serine via the ancestrally encoded *dsdCXA* system, as the intestinal niche is a more sugar-rich environment. However, genomic analysis performed as part of this work revealed that the *dsdCXA* locus is retained in numerous commensal and pathogenic isolates not associated with extraintestinal disease. This means that simply being part of the intestinal flora, forming either a commensal or pathogenic association, is not enough to abolish D-serine tolerance. Additionally, dissemination to extraintestinal sites rarely results in loss of this ability but an almost ubiquitous trend across the *E. coli* phylogeny is the loss of *dsdCX* in isolates that have acquired the LEE. The LEE is a virulence factor that is specific to intestinal pathogens and not ever utilised by ExPEC isolates. Thus there is a direct association between carriage of the LEE and maintenance of *dsdCXA.* Given that the ability to metabolise D-serine has no bearing on its role in LEE repression and that the concentration of D-serine in the intestinal tract is significantly lower than the urinary tract, it is therefore proposed that D-serine acts as a niche specific signal capable of determining when and where *E. coli* pathotypes should colonise within the host.

ExPEC pathogens must first passage through the gut before disseminating to their preferred niche but they are known to do this in a highly competitive manner (Chen *et al.,* 2013). Recent work has described how *E. coli* are in strict competition with one another and form distinct niches within the gut (Meador *et al.,* 2014). The acquisition of the LEE therefore offers an advantage to EHEC and EPEC pathogens as they have a colonisation specific factor to aid their intimate attachment to host cells. Moreover, it is now known that EHEC have developed a specific system for temporally responding to and utilising fucose within the intestine to limit competition for scarce resources (Pacheco *et al.,* 2012). One can assume that because all *E. coli* must passage through the intestine, they at least have the potential to encounter the urethral opening. It can therefore be hypothesised that D-serine acts as a negative signal for LEE positive pathogens limiting them from evolving out of the intestinal niche and subsequently forcing the loss of *dsdCX*. ExPEC pathotypes on the other hand have evolved to utilise D-serine positively without the need to acquire the LEE. The data described in this thesis bolster this hypothesis strongly for multiple *E. coli* pathotypes and provide novel insights to how host physiology and metabolism can work in parallel to define bacterial niche specificity.

The mechanism underlying D-serine repression of the LEE is complex and involves multi-layered transcriptional regulation. It was revealed that a novel D-serine sensing system involving the previously uncharacterised YhaO membrane transporter and the LTTR YhaJ was actively regulated in response to this host metabolite and required for full LEE expression. Additionally, the system was proposed to modulate pre-existing transcriptional regulators of the LEE to enhance its control. YhaO and YhaJ are highly conserved across the *E. coli* phylogeny and D-serine tolerance has also been proposed here to be an ancestral trait of the progenitor strain. Therefore, this system has been adapted to a new function in a manner that relates directly to its original but now redundant feature. Loss of *dsdCXA* in InPEC strains is a sensible adaptation to their niche but maintaining an alternate ability to respond to the common signal D-serine provides the bacteria an efficient method of environmental sensing by separating metabolism and virulence regulation. It is also a perfect example of cellular economics, such that genes can be reused and adapted to new functions rather than be deleted or acquired horizontally. This flexibility also allows pathogen specific adaptations dependent on particular niches within which they reside. *C. rodentium* is a good comparator for this. Both EHEC and *C. rodentium* are LEE positive pathogens but how they regulate this PAI is quite unique, which is unsurprising considering they infect different hosts (Collins *et al.,* 2014, Yang *et al.,* 2010). YhaO and YhaJ were found to be dispensable for pathogenesis of *C. rodentium*, which relies on the LEE. Indeed, this pathogen too has adapted regulators of the LEE that are not found in EHEC. RegA for instance is a global regulator of transcription in *C. rodentium* in both pathogenic and commensal isolates but has adapted itself to become a key regulator of the LEE (Hart *et al.,* 2008, Tan *et al.,* 2015, Tauschek *et al.,* 2010). RegA is not found in EHEC suggesting that although highly adapted for LEE regulation is not intrinsically required. The same can be said of YhaJ although it is interesting to note that this is a chromosomally encoded regulator and most of the distinct regulators between EHEC, EPEC and *C. rodentium* appear to be horizontally acquired (Mellies *et al.,* 2007a, Tan *et al.,* 2015, Tree *et al.,* 2011). RNA-seq has been used in this thesis to reveal novel regulatory networks consisting of pre-existing, chromosomally encoded genes. ChIP-seq technology is now being used in tandem with transcriptomics to map global regulatory networks of DNA binding proteins (Park 2009). Future studies on YhaJ and its global regulatory profile will reveal the full extent of its adapted functions not just on the LEE but also across the genome. It is also exciting to

hypothesise that YhaJ has been adapted by other pathotypes to serve the cell considering how highly conserved this gene is across the *E. coli* phylogeny.

This is not the whole story. Aside from modulation of the LEE, D-serine induced a global shift in gene expression and revealed several effects independent of LEE repression. Accumulation of intracellular D-serine in strains lacking a functional DsdA resulted in activation of the SOS response by an unknown mechanism. This result was puzzling as SOS activation is classically associated with DNA damage or exposure to β-lactam antibiotics (Baharoglu & Mazel 2014, Miller *et al.,* 2004). SOS activation was entirely dependent on intracellular accumulation of D-serine, which occurred in tandem with PG restructuring. It is tantalising to hypothesise that D-serine incorporation into the bacterial cell wall may activate the SOS response to deal with this "stress" in an unprecedented but mechanistically obscure manner. The SOS response is of particular importance for STEC infections. Antibiotic treatment in this instance is considered unsafe as an SOS response induced by such compounds leads to activation of lysogenic phages and upregulation of Stx (McGannon *et al.,* 2010). D-serine however results in neither phage lysis nor Stx activation. This suggests that either RecA and its associated regulon is either being post-translationally controlled in this instance or there could be a "tipping point" in SOS activation within the cell, whereby RecA levels must reach a certain amount to successfully activate lysogenic phages and those levels are not achieved by D-serine accumulation. This is an extremely important point and should be considered in future research aimed at developing treatment strategies for STEC infections.

The work presented in this thesis has revealed novel insights into how host metabolism and niche specific signals impact bacterial pathogens in previously uncharacterised ways. This is a complex and rapidly expanding field that could potentially be used in the development of novel intervention strategies against bacterial pathogens. Antibiotic resistance is not an emerging theme but can no longer be viewed as a future problem (Davies & Davies 2010). A recent report by the World Health Organisation in 2014 addressed the issue of antimicrobial resistance using data from 114 countries, confirming there is currently a global crisis concerned with the continued development of antibiotic resistance (www.who.int). Although new research has provided hope in the form of newly identified antibiotics this issue cannot be overlooked (Ling *et al.,* 2015). New drugs must be regulated and implemented accordingly from day one to ensure a long and

effective lifespan of use. Additionally, there must be emphasis placed on the development of novel intervention strategies that do not rely on antibiotics. The field of anti-virulence has attempted to address this in recent years by the discovery of compounds that inhibit virulence factors specifically rather than acting in a bactericidal or bacteriostatic manner. This effectively "disarms" bacterial pathogens allowing the host to naturally clear the unwanted pathogen (Beckham & Roe 2014, Rasko & Sperandio 2010, Zambelloni *et al.,* 2015). In theory, this approach has the potential to minimise resistance to such compounds as disarming the pathogen can be done in such a way that does not kill the bacteria or have any gross effects on their proliferation, hence removing the strong selective pressure that one faces with antibiotics (Allen *et al.,* 2014). Although this approach is controversial it is promising and should be considered further. Detailed investigation of virulence regulation can also aid this cause by providing a "bottom up" approach to the development of target specific inhibitors (Beckham & Roe 2014). This has been applied recently to the RegA regulator of *C. rodentium* in a study that specifically identified a chemical inhibitor of RegA DNA binding that was capable of significantly limiting infection (Yang *et al.,* 2013). YhaJ could be considered a similar target for drug development, as could YhaO considering its status as a membrane protein. One could further envisage a dual therapy targeting both proteins simultaneously. Given the highly conserved but seemingly specific nature of YhaO and YhaJ function also, these are attractive candidates for compound targeting.

The power of individual metabolites to influence expression of virulence factors is a fascinating discovery that reveals insights into the process of niche selection by bacteria (Connolly *et al.,* 2015, Meador *et al.,* 2014, Nakanishi *et al.,* 2009, Pacheco *et al.,* 2012). Furthermore, recent work using metabolomics has identified the potential of non-toxic natural metabolites as treatment strategies. It was proposed that modulating the metabolome could effectively alter the susceptibility of antibiotic resistant bacterial pathogens to drug treatment, bypassing the need for novel drug design (Peng *et al.,* 2015). It can also be hypothesised that researchers could exploit the knowledge of how non-toxic metabolites regulate virulence with the aim of developing supplementary treatments to use in parallel with more conventional strategies, an approach that has proven useful in the treatment of diarrheal illness with supplementary zinc (Bolick *et al.,* 2014, Crane *et al.,* 2007, Mellies *et al.,* 2012, Sazawal *et al.,* 1995). The idea of modulating diet to limit EHEC infection based on our knowledge of how nutrients and metabolites

regulate the LEE is fast emerging as a plausible strategy to limit infection with studies showing biotin and acetate enriched feeds being capable of limiting EHEC pathogenesis *in vivo* (Fukuda *et al.,* 2011, Yang *et al.,* 2015). This is a largely unexplored field but has great potential considering what has been learned about the crosstalk of metabolism and virulence. It is exciting to postulate that simple dietary supplements could affect the potential of pathogens to compete and subsequently infect in the body.

In conclusion, this thesis describes novel findings in relation to *E. coli* pathogenesis and how host metabolism contributes to its evolution. The work has characterised multiple previously undescribed findings and the results provide a strong framework for much future research. This has implications not only for a deeper understanding of *E. coli* pathogenesis at the molecular level but also in the potential exploitation of such knowledge for development of new and urgent treatment strategies for bacterial infections.

7 References

Abby SS, Rocha EPC. 2012. The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet.* 8(9):e1002983

Abe H, Miyahara A, Oshima T, Tashiro K, Ogura Y, *et al.* 2008. Global regulation by horizontally transferred regulators establishes the pathogenicity of *Escherichia coli*. *DNA Res.* 15:13–23

Adler J. 1975. Chemotaxis in bacteria. *Annu. Rev. Biochem.* 44:341–56

Allen RC, Popat R, Diggle SP, Brown SP. 2014. Targeting virulence: can we make evolution-proof drugs? *Nat. Rev. Microbiol.* 12(4):300–308

Allsopp LP, Beloin C, Ulett GC, Valle J, Totsika M, *et al.* 2012. Molecular characterization of UpaB and UpaC, two new autotransporter proteins of uropathogenic *Escherichia coli* CFT073. *Infect. Immun.* 80(1):321–32

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–10

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11(10):R106

Anfora AT, Haugen BJ, Roesch P, Redford P, Welch RA. 2007. Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by *Escherichia coli* CFT073. *Infect. Immun.* 75(11):5298–5304

Anfora AT, Welch RA. 2006. DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. *J. Bacteriol.* 188(18):6622–28

Arias CA, Martin-Martinez M, Blundell TL, Arthur M, Courvalin P, Reynolds PE. 1999. Characterization and modelling of VanT : a novel , membrane-bound , serine racemase from vancomycin- resistant *Enterococcus gallinarum* BM4174. *Mol. Microbiol.* 31:1653–64

Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, *et al.* 2006. Large-scale identification of protein – protein interaction of *Escherichia coli* K-12. *Genome Res.* 16:686–91

Asakura S, Konno R. 1997. Origin of D-serine present in urine of mutant mice lacking D-amino acid oxidase activity. *Amino Acids*. 12(3-4):213–23

Atlung T, Ingmer H. 1997. H-NS: a modulator of environmentally regulated gene expression. *Mol. Microbiol.* 24(1):7–17

Aviv M, Giladi H, Schreiber G, Oppenheim AB, Glaser G. 1994. Expression of the genes coding for the *Escherichia coli* integration host factor are controlled by growth phase, rpoS, ppGpp and by autoregulation. *Mol. Microbiol.* 14(5):1021–31

Baharoglu Z, Mazel D. 2014. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.* 38:1126–45

Barba J, Bustamante VH, Flores-Valdez MA, Deng W, Finlay BB, Puente JL. 2005. A positive regulatory loop controls expression of the locus of enterocyte effacement-encoded regulators Ler and GrlA. *J. Bacteriol.* 187(23):7918–30

Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65:189–213

Beckham KSH, Connolly JPR, Ritchie JM, Wang D, Gawthorne J a, *et al.* 2014. The Metabolic enzyme AdhE controls the virulence of *Escherichia coli* O157:H7. *Mol. Microbiol.* 93(1):199–211

Beckham KSH, Roe AJ. 2014. From screen to target: insights and approaches for the development of anti-virulence compounds. *Front. Cell. Infect. Microbiol.* 4:139

Bentley R. 2010. Chiral: a confusing etymology. *Chirality*. 22(1):1–2

Berdichevsky T, Friedberg D, Nadler C, Rokney A, Oppenheim A, Rosenshine I. 2005. Ler is a negative autoregulator of the LEE1 operon in enteropathogenic *Escherichia coli*. *J. Bacteriol.* 187(1):349–57

Bertin Y, Deval C, de la Foye A, Masson L, Gannon V, *et al.* 2014. The gluconeogenesis pathway is involved in maintenance of enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. *PLoS One*. 9(6):e98367

Bertin Y, Girardeau JP, Chaucheyras-Durand F, Lyan B, Pujos-Guillot E, *et al.* 2011. Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. *Environ. Microbiol.* 13(2):365–77

Bhakdi S, Mackman N, Menestrina G, Gray L, Hugo F, *et al.* 1988. The hemolysin of *Escherichia coli*. *Eur. J. Epidemiol.* 4(2):135–43

Bidet P, Mahjoub-Messai F, Blanco J, Blanco J, Dehem M, *et al.* 2007. Combined multilocus sequence typing and O serogrouping distinguishes *Escherichia coli* subtypes associated with infant urosepsis and/or meningitis. *J. Infect. Dis.* 196(2):297–303

Bingen E, Picard B, Brahimi N, Mathy S, Desjardins P, *et al.* 1998. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J. Infect. Dis.* 177(3):642–50

Bingle LEH, Constantinidou C, Shaw RK, Islam MS, Patel M, *et al.* 2014. Microarray analysis of the Ler regulon in enteropathogenic and enterohaemorrhagic *Escherichia coli* strains. *PLoS One*. 9(1):e80160

Bloom FR, McFall E, Young MC, Carothers a M. 1975. Positive control in the D-serine deaminase system of *Escherichia coli* K-12. *J. Bacteriol.* 121(3):1092–1101

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. *Infect. Immun.* 62(2):606–14

Bolick DT, Kolling GL, Moore JH, de Oliveira LA, Tung K, *et al.* 2014. Zinc deficiency alters host response and pathogen virulence in a mouse model of enteroaggregative *Escherichia coli*-induced diarrhea. *Gut Microbes*. 5(5):618–27

Branchu P, Matrat S, Vareille M, Garrivier A, Durand A, *et al.* 2014. NsrR, GadE, and GadX interplay in repressing expression of the *Escherichia coli* O157:H7 LEE pathogenicity island in response to nitric oxide. *PLoS Pathog.* 10(1):e1003874

Brunder W, Schmidt H, Karch H. 1997. EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* 24(4):767–78

Brzuszkiewicz E, Brüggemann H, Liesegang H, Emmerth M, Olschläger T, *et al.* 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc. Natl. Acad. Sci. U. S. A.* 103(34):12879–84

Burghout P, van Boxtel R, Van Gelder P, Ringler P, Müller SA, *et al.* 2004. Structure and electrophysiological properties of the YscC secretin from the type III secretion system of *Yersinia enterocolitica*. *J. Bacteriol.* 186(14):4645–54

Burland V. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res.* 26(18):4196–4204

Burns JL, Griffith A, Barry JJ, Jonas M, Chi EY. 2001. Transcytosis of gastrointestinal epithelial cells by *Escherichia coli* K1. *Pediatr. Res.* 49(1):30–37

Bury-Moné S, Nomane Y, Reymond N, Barbet R, Jacquet E, *et al.* 2009. Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet.* 5(9):e1000651

Büttner D. 2012. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 76(2):262–310

Campellone KG, Robbins D, Leong JM. 2004. EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Dev. Cell*. 7(2):217–28

Caparros M, Pisabarro AG, de Pedro MA. 1992. Effect of D-amino acids on structure and synthesis of peptidoglycan in *Escherichia coli*. *J. Bacteriol.* 174(17):5549–59

Caparrós M, Torrecuadrada JL, de Pedro MA. 1991. Effect of D-amino acids on *Escherichia coli* strains with impaired penicillin-binding proteins. *Res. Microbiol.* 142(2-3):345–50

Cárdenas PP, Carrasco B, Defeu Soufo C, César CE, Herr K, *et al.* 2012. RecX facilitates homologous recombination by modulating RecA activities. *PLoS Genet.* 8(12):e1003126

Casaregola S, D'Ari R, Huisman O. 1982. Quantitative evaluation of recA gene expression in *Escherichia coli*. *MGG Mol. Gen. Genet.* 185(3):430–39

Cava F, de Pedro M a, Lam H, Davis BM, Waldor MK. 2011a. Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *EMBO J.* 30(16):3442–53

Cava F, Lam H, de Pedro M a, Waldor MK. 2011b. Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cell. Mol. Life Sci.* 68(5):817–31

Chang D-E, Smalley DJ, Tucker DL, Leatham MP, Norris WE, *et al.* 2004. Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl. Acad. Sci. U. S. A.* 101(19):7427–32

Chaput C, Labigne A, Boneca IG. 2007. Characterization of *Helicobacter pylori* lytic transglycosylases Slt and MltD. *J. Bacteriol.* 189(2):422–29

Chase-Topping ME, McKendrick IJ, Pearce MC, MacDonald P, Matthews L, *et al.* 2007. Risk factors for the presence of high-level shedders of *Escherichia coli* O157 on Scottish farms. *J. Clin. Microbiol.* 45(5):1594–1603

Chaudhuri RR, Khan AM, Pallen MJ. 2004. coliBASE: an online database for *Escherichia coli, Shigella* and *Salmonella* comparative genomics. *Nucleic Acids Res.* 32(Database issue):D296–99

Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, *et al.* 2010. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One*. 5(1):e8801

Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* 40:641–45

Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, *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. *Sci. Transl. Med.* 5(184):184ra60

Chevance FF V, Hughes KT. 2008. Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* 6(6):455–65

Cho B-K, Barrett CL, Knight EM, Park YS, Palsson BØ. 2008. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 105(49):19462–67

Chugani SA, Parsek MR, Chakrabarty AM. 1998. Transcriptional repression mediated by LysR-type regulator CatR bound at multiple binding sites. *J. Bacteriol.* 180(9):2367–72

Clarke MB, Hughes DT, Zhu C, Boedeker EC, Sperandio V. 2006. The QseC sensor kinase: a bacterial adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* 103(27):10420–25

Clarke MB, Sperandio V. 2005. Transcriptional regulation of flhDC by QseBC and sigma (FliA) in enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.* 57(6):1734–49

Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* 5(1):58–65

Collins JW, Keeney KM, Crepin VF, Rathinam V a K, Fitzgerald K a, *et al.* 2014. *Citrobacter rodentium*: infection, inflammation and the microbiota. *Nat. Rev. Microbiol.* 12(9):612–23

Connolly JP, Goldstone RJ, Burgess K, Cogdell RJ, Beatson SA, *et al.* 2015. The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *ISME J.* 9:1039–1051

Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, *et al.* 2014. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio*. 5(4):e01442–14

Cornelis GR. 2006. The type III secretion injectisome. *Nat. Rev. Microbiol.* 4(11):811–25

Cosloy SD, McFall E. 1973. Metabolism of D-serine in *Escherichia coli* K-12: mechanism of growth inhibition. *J. Bacteriol.* 114(2):685–94

Cowan ST. 1954. A review of names for coliform organisms. *Int. Bull. Bacteriol. Nomencl. Taxon.* 4(2):119–24

Crane JK, Naeher TM, Shulgina I, Zhu C, Boedeker EC. 2007. Effect of zinc in enteropathogenic *Escherichia coli* infection. *Infect. Immun.* 75(12):5974–84

Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KE V. 2011. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: Improved metabolite identification by retention time prediction. *Anal. Chem.* 83:8703–10

Crepin VF, Prasannan S, Shaw RK, Wilson RK, Creasey E, *et al.* 2005. Structural and functional studies of the enteropathogenic *Escherichia coli* type III needle complex protein EscJ. *Mol. Microbiol.* 55(6):1658–70

Croxen MA, Finlay BB. 2010. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* 8(1):26–38

Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. 2013. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* 26(4):822–80

Cummings JH, Macfarlane GT. 1991. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* 70(6):443–59

Darra E, Ebner FH, Shoji K, Suzuki H, Mariotto S. 2009. Dual cross-talk between nitric oxide and D-serine in astrocytes and neurons in the brain. *Cent. Nerv. Syst. Agents Med. Chem.* 9(4):289–94

Darwin AJ. 2005. The phage-shock-protein response. *Mol. Microbiol.* 57(3):621–28

Datsenko K a, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97(12):6640–45

Datta S, Costantino N, Court DL. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene*. 379:109–15

Davies J, Davies D. 2010. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74(3):417–33

De Biase D, Tramonti a, Bossa F, Visca P. 1999. The response to stationary-phase stress conditions in *Escherichia coli*: role and regulation of the glutamic acid decarboxylase system. *Mol. Microbiol.* 32(6):1198–1211

De Louvois J. 1994. Acute bacterial meningitis in the newborn. *J. Antimicrob. Chemother.* 34:61–73

Dean P, Kenny B. 2004. Intestinal barrier dysfunction by enteropathogenic *Escherichia coli* is mediated by two effector molecules and a bacterial surface protein. *Mol. Microbiol.* 54(3):665–75

Dean P, Mühlen S, Quitard S, Kenny B. 2010. The bacterial effectors EspG and EspG2 induce a destructive calpain activity that is kept in check by the co-delivered Tir effector. *Cell. Microbiol.* 12(9):1308–21

Dean-Nystrom EA, Bosworth BT, Moon HW, O'Brien AD. 1998. *Escherichia coli* O157:H7 requires Intimin for enteropathogenicity in calves. *Infect. Immun.* 66(9):4560–63

Deloney-marino CR, Wolfe AJ, Visick KL. 2003. Chemoattraction of *Vibrio fischeri* to serine, sucleosides, and N-acetylneuraminic acid, a component of squid light-organ mucus. *Appl. Environ. Microbiol.* 69(12):7527–30

Deng W, Li Y, Hardwidge PR, Frey EA, Pfuetzner RA, *et al.* 2005. Regulation of type III secretion hierarchy of translocators and effectors in attaching and effacing bacterial pathogens. *Infect. Immun.* 73(4):2135–46

Deng W, Li Y, Vallance BA, Finlay BB. 2001. Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect. Immun.* 69(10):6323–35

Deng W, Puente L, Gruenheid S, Li Y, Vallance BA, *et al.* 2004. Dissecting virulence : Systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. U. S. A.* 101(10):3597–3602

Deng W, Yu HB, de Hoog CL, Stoynov N, Li Y, *et al.* 2012. Quantitative proteomic analysis of type III secretome of enteropathogenic *Escherichia coli* reveals an expanded effector repertoire for attaching/effacing bacterial pathogens. *Mol. Cell. Proteomics*. 11(9):692–709

Desai MJ, Armstrong DW. 2004. Analysis of native amino acid and peptide enantiomers by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *J. Mass Spectrom.* 39(2):177–87

Desmarais SM, De Pedro M a, Cava F, Huang KC. 2013. Peptidoglycan at its peaks: how chromatographic analyses can reveal bacterial cell wall structure and assembly. *Mol. Microbiol.* 89(1):1–13

Dietzman DE, Fischer GW, Schoenknecht FD. 1974. Neonatal *Escherichia coli* septicemia—bacterial counts in blood. *J. Pediatr.* 85(1):128–30

Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8(3):185–95

Dong N, Liu L, Shao F. 2010. A bacterial effector targets host DH-PH domain RhoGEFs and antagonizes macrophage phagocytosis. *EMBO J.* 29(8):1363–76

Dong T, Coombes BK, Schellhorn HE. 2009. Role of RpoS in the virulence of *Citrobacter rodentium*. *Infect. Immun.* 77(1):501–7

Dong T, Schellhorn HE. 2009a. Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933. *BMC Genomics*. 10:349

Dong T, Schellhorn HE. 2009b. Control of RpoS in global gene expression of *Escherichia coli* in minimal media. *Mol. Genet. Genomics*. 281(1):19–33

Drews J. 2000. Drug Discovery: A Historical Perspective. *Science (80-.).* 287:1960–64

Durham NN, Milligan R. 1962. A mechanism of growth inhibition by D-serine in a *Flavobacterium*. *Biochem. Biophys. Res. Commun.* 7(5):342–45

Ebel W, Trempy JE. 1999. *Escherichia coli* RcsA, a positive activator of colanic acid capsular polysaccharide synthesis, functions to activate its own expression. *J. Bacteriol.* 181(2):577–84

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–97

Elliott SJ, Sperandio V, Girón JA, Mellies JL, Wainwright L, *et al.* 2000. The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 68(11):6115–26

Elliott SJ, Wainwright LA, Timothy K, Jarvis KG, Deng Y, *et al.* 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 28(1):1–4

Elliott SJ, Yu J, Kaper JB. 1999. The cloned locus of enterocyte effacement from enterohemorrhagic *Escherichia coli* O157:H7 is unable to confer the attaching and effacing phenotype upon E. coli K-12. *Infect. Immun.* 67(8):4260–63

Endo Y, Tsurugi K, Yutsudo T, Takeda Y, Ogasawara T, Igarashi K. 1988. Site of action of a Vero toxin (VT2) from *Escherichia coli* O157:H7 and of Shiga toxin on eukaryotic ribosomes. RNA N-glycosidase activity of the toxins. *Eur. J. Biochem.* 171(1-2):45–50

Erhardt M, Namba K, Hughes KT. 2010. Bacterial nanomachines: the flagellum and type III injectisome. *Cold Spring Harb. Perspect. Biol.* 2(11):a000299

Escherich T. 1885. The intestinal bacteria of the neonate and breast-fed infant. 1885. *Rev. Infect. Dis.* 11(2):352–56

Eto DS, Jones TA, Sundsbak JL, Mulvey MA. 2007. Integrin-mediated host cell invasion by type 1-piliated uropathogenic *Escherichia coli*. *PLoS Pathog.* 3(7):e100

Federiuk CS, Bayer R, Shafer JA. 1983. Characterization of the catalytic pathway for D-serine dehydratase. Evidence for variation of the rate-determining step with substrate structure. *J. Biol. Chem.* 258(9):5379–85

Fernández De Henestrosa a R, Ogi T, Aoyagi S, Chafin D, Hayes JJ, *et al.* 2000. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol. Microbiol.* 35(6):1560–72

Fischbach MA, Lin H, Liu DR, Walsh CT. 2006. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat. Chem. Biol.* 2(3):132–38

Fitzjohn RG. 2012. Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3:1084–92

Flack HD. 2009. Louis Pasteur's discovery of molecular chirality and spontaneous resolution in 1848, together with a complete review of his crystallographic and chemical work. *Acta Crystallogr. A.* 65(5):371–89

Flatley J, Barrett J, Pullan ST, Hughes MN, Green J, Poole RK. 2005. Transcriptional responses of *Escherichia coli* to S-nitrosoglutathione under defined chemostat conditions reveal major changes in methionine biosynthesis. *J. Biol. Chem.* 280(11):10065–72

Francez-Charlot A, Laugel B, Van Gemert A, Dubarry N, Wiorowski F, *et al.* 2003. RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in *Escherichia coli*. *Mol. Microbiol.* 49(3):823–32

Frankel G, Phillips a D, Rosenshine I, Dougan G, Kaper JB, Knutton S. 1998. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol. Microbiol.* 30(5):911–21

Fraser ME, Fujinaga M, Cherney MM, Melton-Celsa AR, Twiddy EM, *et al.* 2004. Structure of shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7. *J. Biol. Chem.* 279(26):27511–17

Friedberg D, Umanski T, Fang Y, Rosenshine I. 1999. Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 34(5):941–52

Frye J, Karlinsey JE, Felise HR, Marzolf B, Dowidar N, *et al.* 2006. Identification of new flagellar genes of *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 188(6):2233–43

Fuchs S, Mühldorfer I, Donohue-Rolfe A, Kerényi M, Emödy L, *et al.* 1999. Influence of RecA on *in vivo* virulence and Shiga toxin 2 production in *Escherichia coli* pathogens. *Microb. Pathog.* 27(1):13–23

Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, *et al.,* 2011. *Bifidobacteria* can protect from enteropathogenic infection through production of acetate. *Nature.* 469(7331):543–547

Gal J. 2008. The discovery of biological enantioselectivity: Louis Pasteur and the fermentation of tartaric acid, 1857-a review and analysis 150 yr later. *Chirality*. 20(1):5–19

Garcia EC, Brumbaugh AR, Mobley HLT. 2011. Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infect. Immun.* 79(3):1225–35

García-Angulo VA, Martínez-Santos VI, Villaseñor T, Santana FJ, Huerta-Saquero A, *et al.* 2012. A distinct regulatory sequence is essential for the expression of a subset of nle genes in attaching and effacing *Escherichia coli*. *J. Bacteriol.* 194(20):5589–5603

Gauthier A, Puente JL, Finlay BB. 2003. Secretin of the enteropathogenic *Escherichia coli* type III secretion system requires components of the type III apparatus for assembly and localization. *Infect. Immun.* 71(6):3310–19

Giangrossi M, Zattoni S, Tramonti A, De Biase D, Falconi M. 2005. Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in *Escherichia coli*. *J. Biol. Chem.* 280(22):21498–505

Goldberg MD, Johnson M, Hinton JCD, Williams PH. 2001. Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 41(3):549–59

Goldstone RJ, Popat R, Schuberth H-J, Sandra O, Sheldon IM, Smith DG. 2014. Genomic characterisation of an endometrial pathogenic *Escherichia coli* strain reveals the acquisition of genetic elements associated with extra-intestinal pathogenicity. *BMC Genomics*. 15(1):1075

Gonyar LA, Kendall MM. 2014. Ethanolamine and choline promote expression of putative and characterized fimbriae in enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 82(1):193–201

Gophna U, Ron EZ, Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene*. 312:151–63

Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, *et al.* 2012. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc. Natl. Acad. Sci. U. S. A.* 109:3065–70

Grainger DC, Hurd D, Goldberg MD, Busby SJW. 2006. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* 34(16):4642–52

Grula MM, Grula EA. 1962. Cell division in a species of *Erwinia* IV. Metabolic blocks in panothenate biosynthesis and their relationship to inhibition of cell division. *J. Bacteriol.* 83:989–97

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704

Gurmu D, Lu J, Johnson K a, Nordlund P, Holmgren A, Erlandsen H. 2009. The crystal structure of the protein YhaK from *Escherichia coli* reveals a new subclass of redox sensitive enterobacterial bicupins. *Proteins*. 74(1):18–31

Haack KR, Robinson CL, Miller KJ, Fowlkes JW, Mellies JL. 2003. Interaction of Ler at the LEE5 (tir) operon of enteropathogenic *Escherichia coli*. *Infect. Immun.* 71(1):384–92

Hale CR, Scallan E, Cronquist AB, Dunn J, Smith K, *et al.* 2012. Estimates of enteric illness attributable to contact with animals and their environments in the United States. *Clin. Infect. Dis.* 54(5):S472–79

Hancock V, Klemm P. 2007. Global gene expression profiling of asymptomatic bacteriuria *Escherichia coli* during biofilm growth in human urine. *Infect. Immun.* 75(2):966–76

Hansen A-M, Kaper JB. 2009. Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.* 73(3):446–65

Hart E, Yang J, Tauschek M, Kelly M, Wakefield MJ, *et al.* 2008. RegA, an AraC-like protein, is a global transcriptional regulator that controls virulence gene expression in *Citrobacter rodentium*. *Infect. Immun.* 76(11):5247–56

Hartleib S, Prager R, Hedenström I, Löfdahl S, Tschäpe H. 2003. Prevalence of the new, SPI1-like, pathogenicity island ETT2 among *Escherichia coli*. *Int. J. Med. Microbiol.* 292(7-8):487–93

Haugen BJ, Pellett S, Redford P, Hamilton HL, Roesch PL, Welch R a. 2007. *In vivo* gene expression analysis identifies genes required for enhanced colonization of the mouse urinary tract by uropathogenic *Escherichia coli* strain CFT073 dsdA. *Infect. Immun.* 75(1):278–89

Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, *et al.* 2001. Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8(1):11–22

Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, *et al.* 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*. 410(6832):1099–1103

Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA. 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 110(31):12810–15

Heincz MC, McFall E. 1978. Role of the DsdC activator in regulation of D-serine deaminase synthesis. *J. Bacteriol.* 136(1):96–103

Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* 68(4):692–744

Holmes A, Lindestam Arlehamn CS, Wang D, Mitchell TJ, Evans TJ, Roe AJ. 2012. Expression and regulation of the *Escherichia coli* O157:H7 effector proteins NleH1 and NleH2. *PLoS One*. 7(3):e33408

Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, *et al.,* 2009. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics*. 10:36

Honda N, Iyoda S, Yamamoto S, Terajima J, Watanabe H. 2009. LrhA positively controls the expression of the locus of enterocyte effacement genes in enterohemorrhagic *Escherichia coli* by differential regulation of their master regulators PchA and PchB. *Mol. Microbiol.* 74(6):1393–41

Hughes DT, Sperandio V. 2008. Inter-kingdom signalling: communication between bacteria and their hosts. *Nat. Rev. Microbiol.* 6(2):111–20

Hughes DT, Terekhova DA, Liou L, Hovde CJ, Sahl JW, *et al.* 2010. Chemical sensing in mammalian host-bacterial commensal associations. *Proc. Natl. Acad. Sci. U. S. A.* 107(21):9831–36

Huo Y-X, Zhang Y-T, Xiao Y, Zhang X, Buck M, *et al.* 2009. IHF-binding sites inhibit DNA loop formation and transcription initiation. *Nucleic Acids Res.* 37(12):3878–86

Hurley BP, Thorpe CM, Acheson DW. 2001. Shiga toxin translocation across intestinal epithelial cells is enhanced by neutrophil transmigration. *Infect. Immun.* 69(10):6148–55

Ide T, Laarmann S, Greune L, Schillers H, Oberleithner H, Schmidt MA. 2001. Characterization of translocation pores inserted into plasma membranes by type III-secreted Esp proteins of enteropathogenic *Escherichia coli*. *Cell. Microbiol.* 3(10):669–79

Ideses D, Gophna U, Paitan Y, Chaudhuri RR, Pallen MJ, Ron EZ. 2005. A degenerate type III secretion system from septicemic *Escherichia coli* contributes to pathogenesis. *J. Bacteriol.* 187(23):8164–71

Iguchi A, Thomson NR, 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. *J. Bacteriol.* 191(1):347–54

Isenbarger TA, Carr CE, Johnson SS, Finney M, Church GM, *et al.* 2008. The most conserved genome segments for life detection on Earth and other planets. *Orig. Life Evol. Biosph.* 38(6):517–33

Islam MS, Bingle LEH, Pallen MJ, Busby SJW. 2011a. Organization of the LEE1 operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Mol. Microbiol.* 79:468–83

Islam MS, Pallen MJ, Busby SJW. 2011b. A cryptic promoter in the LEE1 regulatory region of enterohaemorrhagic *Escherichia coli*: promoter specificity in AT-rich gene regulatory regions. *Biochem. J.* 436(3):681–86

Iyoda S, Koizumi N, Satou H, Lu Y, Saitoh T, *et al.* 2006. The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.* 188(16):5682–92

Iyoda S, Watanabe H. 2005. ClpXP Protease Controls Expression of the Type III Protein Secretion System through Regulation of RpoS and GrlR Levels in Enterohemorrhagic *Escherichia coli*. *J. Bacteriol.* 187(12):4086–94

Iyoda, S. and Watanabe H. 2004. Positive effects of multiple pch genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escherichia coli* O157 : H7 to HEp-2 cells. *Microbiology*. 150(7):2357–2571

Jahreis K, Bentler L, Bockmann J, Meyer A, Siepelmeyer J, *et al.* 2002. Adaptation of sucrose metabolism in the adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. *J. Bacteriol.* 184(19):5307–16

Jeong J-H, Kim H-J, Kim K-H, Shin M, Hong Y, *et al.* 2012. An unusual feature associated with LEE1 P1 promoters in enteropathogenic *Escherichia coli* (EPEC). *Mol. Microbiol.* 83(3):612–22

Johnson JR, Delavari P, Kuskowski M, Stell AL. 2001a. Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J. Infect. Dis.* 183(1):78–88

Johnson JR, Delavari P, O'Bryan TT. 2001b. *Escherichia coli* O18:K1:H7 isolates from patients with acute cystitis and neonatal meningitis exhibit common phylogenetic origins and virulence factor profiles. *J. Infect. Dis.* 183(3):425–34

Johnson JR, Russo TA. 2002. Extraintestinal pathogenic *Escherichia coli* : "The other bad *E. coli* ." *J. Lab. Clin. Med.* 139(3):155–62

Kahramanoglou C, Seshasayee ASN, Prieto AI, Ibberson D, Schmidt S, *et al.* 2011. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli. Nucleic Acids Res.* 39(6):2073–91

Kailasan Vanaja S, Bergholz TM, Whittam TS. 2009. Characterization of the *Escherichia coli* O157:H7 Sakai GadE regulon. *J. Bacteriol.* 191(6):1868–77

Kamenšek S, Podlesek Z, Gillor O, Zgur-Bertok D. 2010. Genes regulated by the *Escherichia coli* SOS repressor LexA exhibit heterogeneous expression. *BMC Microbiol.* 10(1):283

Kanamaru K, Tatsuno I, Tobe T, Sasakawa C. 2000. SdiA, an *Escherichia coli* homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohaemorrhagic Escherichia coli O157:H7. *Mol. Microbiol.* 38(4):805–16

Kang L, Shaw AC, Xu D, Xia W, Zhang J, *et al.* 2011. Upregulation of MetC is essential for D-alanine-independent growth of an alr/dadX-deficient *Escherichia coli* strain. *J. Bacteriol.* 193(5):1098–1106

Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli. Nat. Rev. Microbiol.* 2(2):123–40

Kearns DB. 2010. A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* 8(9):634–44

Kendall MM, Gruber CC, Parker CT, Sperandio V. 2012. Ethanolamine controls expression of genes encoding components involved in interkingdom signaling and virulence in enterohemorrhagic *Escherichia coli* O157:H7. *MBio*. 3(3):e00050–12

Kendall MM, Gruber CC, Rasko DA, Hughes DT, Sperandio V. 2011. Hfq virulence regulation in enterohemorrhagic *Escherichia coli* O157:H7 strain 86-24. *J. Bacteriol.* 193(24):6843–51

Kenny B, Abe A, Stein M, Finlay BB. 1997a. Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect. Immun.* 65(7):2606–12

Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB. 1997b. Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell*. 91(4):511–20

Kenny B, Jepson M. 2000. Targeting of an enteropathogenic *Escherichia coli* (EPEC) effector protein to host mitochondria. *Cell. Microbiol.* 2(6):579–90

Keyser P, Elofsson M, Rosell S, Wolf-Watz H. 2008. Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against Gram-negative bacteria. *J. Intern. Med.* 264(1):17–29

Khan NA, Kim Y, Shin S, Kim KS. 2007. FimH-mediated *Escherichia coli* K1 invasion of human brain microvascular endothelial cells. *Cell. Microbiol.* 9(1):169–78

Kim KS. 2012. Current concepts on the pathogenesis of *Escherichia coli* meningitis: implications for therapy and prevention. *Curr. Opin. Infect. Dis.* 25(3):273–78

Kitagawa R, Takaya A, Yamamoto T. 2011. Dual regulatory pathways of flagellar gene expression by ClpXP protease in enterohaemorrhagic *Escherichia coli*. *Microbiology*. 157(11):3094–3103

Korte-Berwanger M, Sakinc T, Kline K, Nielsen H V, Hultgren S, Gatermann SG. 2013. Significance of the D-serine-deaminase and D-serine metabolism of *Staphylococcus saprophyticus* for virulence. *Infect. Immun.* 81(12):4525–33

Laaberki M-H, Janabi N, Oswald E, Repoila F. 2006. Concert of regulators to switch on LEE expression in enterohemorrhagic *Escherichia coli* O157:H7: interplay between Ler, GrlA, HNS and RpoS. *Int. J. Med. Microbiol.* 296(4-5):197–210

Lam H, Oh D-C, Cava F, Takacs CN, Clardy J, *et al.* 2009. D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science*. 325(5947):1552–55

Lane MC, Mobley HLT. 2007. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney Int.* 72(1):19–25

Lathem WW, Grys TE, Witowski SE, Torres AG, Kaper JB, *et al.* 2002. StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol. Microbiol.* 45(2):277–88

Laubacher ME, Ades SE. 2008. The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. *J. Bacteriol.* 190(6):2065–74

Law D, Kelly J. 1995. Use of heme and hemoglobin by *Escherichia coli* O157 and other Shiga-like-toxin-producing E. coli serogroups. *Infect. Immun.* 63(2):700–702

Leatham-Jensen MP, Frimodt-Møller J, Adediran J, Mokszycki ME, Banner ME, *et al.* 2012. The streptomycin-treated mouse intestine selects *Escherichia coli envZ* missense mutants that interact with dense and diverse intestinal microbiota. *Infect. Immun.* 80(5):1716–27

Lehnen D, Blumer C, Polen T, Wackwitz B, Wendisch VF, Unden G. 2002. LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in *Escherichia coli. Mol. Microbiol.* 45(2):521–32

Li J, Go AC, Ward MJ, Ottemann KM. 2010. The chemical-in-plug bacterial chemotaxis assay is prone to false positive responses. *BMC Res. Notes*. 3(1):77

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, *et al.* 2015. A new antibiotic kills pathogens without detectable resistance. *Nature*. 517(7535):455–59

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*. 25(4):402–8

Low AS, Holden N, Rosser T, Roe AJ, Constantinidou C, *et al.* 2006. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environ. Microbiol.* 8(6):1033–47

Lu S, Zhang X, Zhu Y, Kim KS, Yang J, Jin Q. 2011. Complete genome sequence of the neonatal-meningitis-associated *Escherichia coli* strain CE10. *J. Bacteriol.* 193(24):7005

Luzader DH, Clark DE, Gonyar LA, Kendall MM. 2013. EutR is a direct regulator of genes that contribute to metabolism and virulence in enterohemorrhagic *Escherichia coli* O157:H7. *J. Bacteriol.* 195(21):4947–53

Maas WK, Davis BD. 1950. Pantothenate studies. I. Interference by D-serine and L-aspartic acid with pantothenate synthesis in *Escherichia coli*. *J. Bacteriol.* 60(6):733–45

Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology*. 154(12):3609–23

Madrid C, Balsalobre C, García J, Juárez A. 2007. The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. *Mol. Microbiol.* 63(1):7–14

Mahajan A, Currie CG, Mackie S, Tree J, McAteer S, *et al.* 2009. An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157:H7 with bovine intestinal epithelium. *Cell. Microbiol.* 11(1):121–37

Maiques E, Úbeda C, Campoy S, Lasa Í, Novick RP, *et al.* 2006. β-Lactam Antibiotics Induce the SOS Response and Horizontal Transfer of Virulence Factors in *Staphylococcus aureus*. *J. Bacteriol.* 188(7):2726–29

Makino S-I, Tobe T, Asakura H, Watarai M, Ikeda T, *et al.* 2003. Distribution of the secondary type III secretion system locus found in enterohemorrhagic *Escherichia coli* O157:H7 isolates among Shiga toxin-producing *E. coli* strains. *J. Clin. Microbiol.* 41(6):2341–47

Man EH, Bada JL. 1987. Dietary D-amino acids. *Annu. Rev. Nutr.* 7:209–25

Mangan MW, Lucchini S, Danino V, Cróinín TO, Hinton JCD, Dorman CJ. 2006. The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.* 59(6):1831–47

Mao Y, Doyle MP, Chen J. 2006. Role of colanic acid exopolysaccharide in the survival of enterohaemorrhagic *Escherichia coli* O157:H7 in simulated gastrointestinal fluids. *Lett. Appl. Microbiol.* 42(6):642–47

Marchès O, Batchelor M, Shaw RK, Patel A, Cummings N, *et al.* 2006. EspF of enteropathogenic *Escherichia coli* binds sorting nexin 9. *J. Bacteriol.* 188(8):3110–15

Marchès O, Covarelli V, Dahan S, Cougoule C, Bhatta P, *et al.* 2008. EspJ of enteropathogenic and enterohaemorrhagic *Escherichia coli* inhibits opsono-phagocytosis. *Cell. Microbiol.* 10(5):1104–15

Marchès O, Nougayrède JP, Boullier S, Mainil J, Charlier G, *et al.* 2000. Role of tir and intimin in the virulence of rabbit enteropathogenic *Escherichia coli* serotype O103:H2. *Infect. Immun.* 68(4):2171–82

Maroncle NM, Sivick KE, Brady R, Stokes F-E, Mobley HLT. 2006. Protease activity, secretion, cell entry, cytotoxicity, and cellular targets of secreted autotransporter toxin of uropathogenic *Escherichia coli*. *Infect. Immun.* 74(11):6124–34

Maruvada R, Argon Y, Prasadarao N V. 2008. *Escherichia coli* interaction with human brain microvascular endothelial cells induces signal transducer and activator of transcription 3 association with the C-terminal domain of Ec-gp96, the outer membrane protein A receptor for invasion. *Cell. Microbiol.* 10(11):2326–38

Massoni SC, Leeson MC, Long JE, Gemme K, Mui A, Sandler SJ. 2012. Factors limiting SOS expression in log-phase cells of *Escherichia coli*. *J. Bacteriol.* 194(19):5325–33

Masuda N, Church GM. 2002. *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. *J. Bacteriol.* 184(22):6225–34

Masuda N, Church GM. 2003. Regulatory network of acid resistance genes in *Escherichia coli. Mol. Microbiol.* 48(3):699–712

McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, *et al.* 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat. Genet.* 36(12):1268–74

McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 92(5):1664–68

McDaniel TK, Kaper JB. 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on E. coli K-12. *Mol. Microbiol.* 23(2):399–407

McGannon CM, Fuller CA, Weiss AA. 2010. Different classes of antibiotics differentially influence shiga toxin production. *Antimicrob. Agents Chemother.* 54(9):3790–98

McNeilly TN, Naylor SW, Mahajan A, Mitchell MC, McAteer S, *et al.* 2008. *Escherichia coli* O157:H7 colonization in cattle following systemic and mucosal immunization with purified H7 flagellin. *Infect. Immun.* 76(6):2594–2602

Meador JP, Caldwell ME, Cohen PS, Conway T. 2014. *Escherichia coli* pathotypes occupy distinct niches in the mouse intestine. *Infect. Immun.* 82(5):1931–38

Mellies JL, Barron AMS, Carmona AM. 2007a. Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect. Immun.* 75(9):4199–4210

Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB. 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). *Mol. Microbiol.* 33(2):296–306

Mellies JL, Haack KR, Galligan DC. 2007b. SOS regulation of the type III secretion system of enteropathogenic *Escherichia coli*. *J. Bacteriol.* 189(7):2863–72

Mellies JL, Thomas K, Turvey M, Evans NR, Crane J, *et al.* 2012. Zinc-induced envelope stress diminishes type III secretion in enteropathogenic *Escherichia coli*. *BMC Microbiol.* 12:123

Mesibov R, Adler J. 1972. Chemotaxis toward amino acids in *Escherichia coli*. *J. Bacteriol.* 112(1):315–26

Michel B. 2005. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol.* 3(7):e255

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL. 2004. *Escherichia coli* Hfq has distinct interaction surfaces for *dsrA, rpoS* and poly(A) RNAs. *Nat. Struct. Mol. Biol.* 11(12):1206–14

Miller C, Thomsen LE, Gaggero C, Mosseri R, Ingmer H, Cohen SN. 2004. SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science*. 305(5690):1629–31

Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell*. 108(5):583–86

Moritz RL, Welch RA. 2006. The *Escherichia coli argW-dsdCXA* genetic island is highly variable, and *E. coli* K1 strains commonly possess two copies of *dsdCXA*. *J. Clin. Microbiol.* 44(11):4038–48

Mothet JP, Parent AT, Wolosker H, Brady RO, Linden DJ, *et al.* 2000. D-serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor. *Proc. Natl. Acad. Sci. U. S. A.* 97(9):4926–31

Murphy KC, Campellone KG. 2003. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.* 4:11

Nagata Y, Masui R, Akino T. 1992. The presence of free D-serine, D-alanine and D-proline in human plasma. *Experientia*. 48(10):986–88

Nagata Y, Sato T, Enomoto N, Ishii Y, Sasaki K, Yamada T. 2007. High concentrations of D-amino acids in human gastric juice. *Amino Acids*. 32(1):137–40

Nakanishi N, Tashiro K, Kuhara S, Hayashi T, Sugimoto N, Tobe T. 2009. Regulation of virulence by butyrate sensing in enterohaemorrhagic *Escherichia coli*. *Microbiology*. 155(2):521–30

Nanda AM, Thormann K, Frunzke J. 2015. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* 197(3):410–19

Nassar FJ, Rahal EA, Sabra A, Matar GM. 2013. Effects of subinhibitory concentrations of antimicrobial agents on *Escherichia coli* O157:H7 Shiga toxin release and role of the SOS response. *Foodborne Pathog. Dis.* 10(9):805–12

Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli. Clin. Microbiol. Rev.* 11:142-201

Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, *et al.* 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*. 313(5784):236–38

Naylor SW, Low JC, Besser TE, Mahajan A, Gunn GJ, *et al.* 2003. Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host. *Infect. Immun.* 71:1505–12

Neely MN, Friedman DI. 1998. Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol. Microbiol.* 28(6):1255–67

Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, *et al.* 2013. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*. 502(7469):96–99

Nguyen YN, Sheng H, Dakarapu R, Falck JR, Hovde CJ, Sperandio V. 2013. The acyl-homoserine lactone synthase YenI from *Yersinia enterocolitica* modulates virulence gene expression in enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 81(11):4192–99

Njoroge JW, Nguyen Y, Curtis MM, Moreira CG, Sperandio V. 2012. Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli. MBio*. 3(5):e00280–12

Nørregaard-madsen M, Fall EMC. 1995. Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. *J. Bacteriol.* 177(22):6456–61

Ogasawara H, Yamada K, Kori A, Yamamoto K, Ishihama A. 2010. Regulation of the *Escherichia coli csgD* promoter: interplay between five transcription factors. *Microbiology*. 156(8):2470–83

Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, *et al.* 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 106(42):17939–44

Okhuysen PC, Dupont HL. 2010. Enteroaggregative *Escherichia coli* (EAEC): a cause of acute and persistent diarrhea of worldwide importance. *J. Infect. Dis.* 202(4):503–5

Oshima T, Ishikawa S, Kurokawa K, Aiba H, Ogasawara N. 2006. *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.* 13(4):141–53

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12(2):87–98

Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, *et al.* 2012. Fucose sensing regulates bacterial intestinal colonization. *Nature*. 492(7427):113–17

Padavannil A, Jobichen C, Mills E, Velazquez-Campoy A, Li M, *et al.* 2013. Structure of GrlR-GrlA complex that prevents GrlA activation of virulence genes. *Nat. Commun.* 4:2546

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. 20:289–90

Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10(10):669–80

Peng B, Su Y-B, Li H, Han Y, Guo C, *et al.* 2015. Exogenous alanine and/or glucose plus kanamycin kills antibiotic-resistant bacteria. *Cell Metab.* 21(2):249–61

Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, *et al.* 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature*. 409:529–33

Petty NK, Feltwell T, Pickard D, Clare S, Toribio AL, *et al.* 2011. *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. *PLoS Pathog.* 7(4):e1002018

Philpott DJ, Ackerley CA, Kiliaan AJ, Karmali MA, Perdue MH, Sherman PM. 1997. Translocation of verotoxin-1 across T84 monolayers: mechanism of bacterial toxin penetration of epithelium. *Am. J. Physiol.* 273(6):1349–58

Pitout JDD. 2012. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Front. Microbiol.* 3:9

Plumbridge J, Vimr E. 1999. Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by *Escherichia coli*. *J. Bacteriol.* 181(1):47–54

Porter ME, Mitchell P, Free A, Smith DGE, Gally DL. 2005. The LEE1 promoters from both Enteropathogenic and Enterohemorrhagic *Escherichia coli* can be activated by PerC-like proteins from either organism. *J. Bacteriol.* 187(2):458–72

Prasadarao N V. 2002. Identification of *Escherichia coli* outer membrane protein A receptor on human brain microvascular endothelial cells. *Infect. Immun.* 70(8):4556–63

Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee ASN, Luscombe NM. 2012. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Res.* 40(8):3524–37

Pruimboom-Brees IM, Morgan TW, Ackermann MR, Nystrom ED, Samuel JE, *et al.* 2000. Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *Proc. Natl. Acad. Sci. U. S. A.* 97(19):10325–29

Pullan ST, Gidley MD, Jones R a, Barrett J, Stevanin TM, *et al.* 2007. Nitric oxide in chemostat-cultured *Escherichia coli* is sensed by Fnr and other global regulators: unaltered methionine biosynthesis indicates lack of S nitrosation. *J. Bacteriol.* 189(5):1845–55

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. *R Found. Stat. Comput. Vienna, Austria.*

Rasko DA, Sperandio V. 2010. Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* 9(2):117–28

Raterman EL, Welch RA. 2013. Chemoreceptors of *Escherichia coli* CFT073 play redundant roles in chemotaxis toward urine. *PLoS One*. 8(1):e54133

Reading NC, Rasko DA, Torres AG, Sperandio V. 2009. The two-component system QseEF and the membrane protein QseG link adrenergic and stress sensing to bacterial pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 106(14):5889–94

Reading NC, Torres AG, Kendall MM, Hughes DT, Yamamoto K, Sperandio V. 2007. A novel two-component signaling system that activates transcription of an enterohemorrhagic *Escherichia coli* effector involved in remodeling of host actin. *J. Bacteriol.* 189(6):2468–76

Ren C-P, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, *et al.* 2004. The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition. *J. Bacteriol.* 186(11):3547–60

Richardson T, Finley J. 1986. *Chemical Changes in Food during Processing*. Boston, MA: Springer US

Riedel CU, Casey PG, Mulcahy H, O'Gara F, Gahan CGM, Hill C. 2007. Construction of p16Slux, a novel vector for improved bioluminescent labeling of gram-negative bacteria. *Appl. Environ. Microbiol.* 73(21):7092–95

Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, *et al.* 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* 308(12):681–85

Ritchie JM, Waldor MK. 2005. The locus of enterocyte effacement-encoded effector proteins all promote enterohemorrhagic *Escherichia coli* pathogenicity in infant rabbits. *Infect. Immun.* 73(3):1466–74

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26(1):139–40

Rode CK, Melkerson-Watson LJ, Johnson AT, Bloch CA. 1999. Type-specific contributions to chromosome size differences in *Escherichia coli*. *Infect. Immun.* 67(1):230–36

Roe AJ, Tysall L, Dransfield T, Wang D, Fraser-Pitt D, *et al.* 2007. Analysis of the expression, regulation and export of NleA-E in *Escherichia coli* O157:H7. *Microbiology*. 153(Pt 5):1350–60

Roe AJ, Yull H, Naylor SW, Martin J, Smith DGE, *et al.* 2003. Heterogeneous surface expression of EspA translocon filaments by *Escherichia coli* O157:H7 is controlled at the posttranscriptional level. *Infect. Immun.* 71(10):5900–5909

Roesch PL, Redford P, Batchelet S, Moritz RL, Pellett S, *et al.* 2003. Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract. *Mol. Microbiol.* 49(1):55–67

Rosenshine I, Ruschkowski S, Finlay BB. 1996. Expression of attaching/effacing activity by enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis upon contact with epithelial cells. *Infect. Immun.* 64(3):966–73

Russell RM, Sharp FC, Rasko D a, Sperandio V. 2007. QseA and GrlR/GrlA regulation of the locus of enterocyte effacement genes in enterohemorrhagic *Escherichia coli*. *J. Bacteriol.* 189(14):5387–92

Sakinç T, Michalski N, Kleine B, Gatermann SG. 2009. The uropathogenic species *Staphylococcus saprophyticus* tolerates a high concentration of D-serine. *FEMS Microbiol. Lett.* 299(1):60–64

Sal-Man N, Deng W, Finlay BB. 2012. EscI: a crucial component of the type III secretion system forms the inner rod structure in enteropathogenic *Escherichia coli*. *Biochem. J.* 442(1):119–25

Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñiz-Rascado L, *et al.* 2013. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 41:203–13

Saliba A-E, Westermann AJ, Gorski SA, Vogel J. 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42(14):8845–60

Sallee NA, Rivera GM, Dueber JE, Vasilescu D, Mullins RD, *et al.* 2008. The pathogen protein EspF(U) hijacks actin polymerization using mimicry and multivalency. *Nature*. 454(7207):1005–8

Sánchez-SanMartín C, Bustamante VH, Calva E, Puente JL. 2001. Transcriptional regulation of the orf19 gene and the *tir-cesT-eae* operon of enteropathogenic *Escherichia coli*. *J. Bacteriol.* 183(9):2823–33

Sandvig K, Garred O, Prydz K, Kozlov J V, Hansen SH, van Deurs B. 1992. Retrograde transport of endocytosed Shiga toxin to the endoplasmic reticulum. *Nature*. 358(6386):510–12

Sayed AK, Foster JW. 2009. A 750 bp sensory integration region directs global control of the *Escherichia coli* GadE acid resistance regulator. *Mol. Microbiol.* 71(6):1435–50

Sazawal S, Black RE, Bhan MK, Bhandari N, Sinha A, Jalla S. 1995. Zinc supplementation in young children with acute diarrhea in India. *N. Engl. J. Med.* 333(13):839–44

Schauer DB, Falkow S. 1993. Attaching and effacing locus of a *Citrobacter freundii* biotype that causes transmissible murine colonic hyperplasia. *Infect. Immun.* 61(6):2486–92

Schell M a. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* 47:597–626

Schmidt H, Maier E, Karch H, Benz R. 1996. Pore-forming properties of the plasmid-encoded hemolysin of enterohemorrhagic *Escherichia coli* O157:H7. *Eur. J. Biochem.* 241(2):594–601

Schüller S. 2011. Shiga toxin interaction with human intestinal epithelium. *Toxins (Basel).* 3(6):626–39

Schüller S, Heuschkel R, Torrente F, Kaper JB, Phillips AD. 2007. Shiga toxin binding in normal and inflamed human intestinal mucosa. *Microbes Infect.* 9(1):35–39

Selvaraj SK, Prasadarao N V. 2005. *Escherichia coli* K1 inhibits proinflammatory cytokine induction in monocytes by preventing NF-kappaB activation. *J. Leukoc. Biol.* 78(2):544–54

Shah DH. 2014. RNA sequencing reveals differences between the global transcriptomes of *Salmonella enterica* serovar Enteritidis strains with high and low pathogenicities. *Appl. Environ. Microbiol.* 80(3):896–906

Shakhnovich EA, Davis BM, Waldor MK. 2009. Hfq negatively regulates type III secretion in EHEC and several other pathogens. *Mol. Microbiol.* 74(2):347–63

Shames SR, Deng W, Guttman JA, de Hoog CL, Li Y, *et al.* 2010. The pathogenic *E. coli* type III effector EspZ interacts with host CD98 and facilitates host cell prosurvival signalling. *Cell. Microbiol.* 12(9):1322–39

Sharma VK, Zuerner RL. 2004. Role of *hha* and *ler* in transcriptional regulation of the *esp* operon of enterohemorrhagic *Escherichia coli* O157:H7. *J. Bacteriol.* 186(21):7290–7301

Sharp FC, Sperandio V. 2007. QseA directly activates transcription of LEE1 in enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 75(5):2432–40

Sheikh J, Dudley EG, Sui B, Tamboura B, Suleman A, Nataro JP. 2006. EilA, a HilA-like regulator in enteroaggregative *Escherichia coli*. *Mol. Microbiol.* 61(2):338–50

Shimoni Y, Altuvia S, Margalit H, Biham O. 2009. Stochastic analysis of the SOS response in *Escherichia coli*. *PLoS One*. 4(5):e5363

Shin S, Castanie-Cornet MP, Foster JW, Crawford J a, Brinkley C, Kaper JB. 2001. An activator of glutamate decarboxylase genes regulates the expression of enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, Per. *Mol. Microbiol.* 41(5):1133–50

Sims GE, Kim S-H. 2011. Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). *Proc. Natl. Acad. Sci. U. S. A.* 108(20):8329–34

Singh SS, Singh N, Bonocora RP, Fitzgerald DM, Wade JT, Grainger DC. 2014. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* 28(3):214–19

Smith WE, Kane A V, Campbell ST, Acheson DWK, Cochran BH, Thorpe CM. 2003. Shiga toxin 1 triggers a ribotoxic stress response leading to p38 and JNK activation and induction of apoptosis in intestinal epithelial cells. *Infect. Immun.* 71(3):1497–1504

Sperandio V, Li CC, Kaper JB. 2002. Quorum-sensing *Escherichia coli* regulator A: A regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic *E. coli*. *Infect. Immun.* 70:3085–93

Sperandio V, Mellies JL, Delahay RM, Frankel G, Crawford JA, *et al.* 2000. Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. *Mol. Microbiol.* 38(4):781–93

Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB. 1999. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 96(26):15196–201

Sukumaran SK, Selvaraj SK, Prasadarao N V. 2004. Inhibition of apoptosis by *Escherichia coli* K1 is accompanied by increased expression of BclXL and blockade of mitochondrial cytochrome c release in macrophages. *Infect. Immun.* 72(10):6012–22

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: A genome comparison visualizer. *Bioinformatics*. 27:1009–10

Takao M, Yen H, Tobe T. 2014. LeuO enhances butyrate-induced virulence expression through a positive regulatory loop in enterohaemorrhagic *Escherichia coli. Mol. Microbiol.* 93(6):1302–13

Tan A, Petty NK, Hocking D, Bennett-Wood V, Wakefield M, *et al.* 2015. Evolutionary adaptation of an AraC-like regulatory protein in *Citrobacter rodentium* and *Escherichia species*. *Infect. Immun.* IAI.02697-:

Taschner PE, Ypenburg N, Spratt BG, Woldringh CL. 1988. An amino acid substitution in penicillin-binding protein 3 creates pointed polar caps in *Escherichia coli*. *J. Bacteriol.* 170(10):4828–37

Tatsuno I, Nagano K, Taguchi K, Rong L, Mori H, Sasakawa C. 2003. Increased adherence to Caco-2 cells caused by disruption of the *yhiE* and *yhiF* genes in enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* 71(5):2598–2606

Tauschek M, Yang J, Hocking D, Azzopardi K, Tan A, *et al.* 2010. Transcriptional analysis of the *grlRA* virulence operon from *Citrobacter rodentium*. *J. Bacteriol.* 192(14):3722–34

Tobe T, Ando H, Ishikawa H, Abe H, Tashiro K, *et al.* 2005. Dual regulatory pathways integrating the RcsC-RcsD-RcsB signalling system control enterohaemorrhagic *Escherichia coli* pathogenicity. *Mol. Microbiol.* 58(1):320–33

Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, *et al.* 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl. Acad. Sci. U. S. A.* 103(40):14941–46

Tobe T, Nakanishi N, Sugimoto N. 2011. Activation of motility by sensing short-chain fatty acids via two steps in a flagellar gene regulatory cascade in enterohemorrhagic *Escherichia coli*. *Infect. Immun.* 79(3):1016–24

Totsika M, Gomes Moriel D, Idris A, A. Rogers B, J. Wurpel D, *et al.* 2012. Uropathogenic *Escherichia coli* mediated urinary tract infection. *Curr. Drug Targets*. 13(11):1386–99

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, *et al.* 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5(1):e1000344

Tramonti A, De Canio M, De Biase D. 2008. GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the *gadY-gadW* divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol. Microbiol.* 70(4):965–82

Tree JJ, Roe AJ, Flockhart A, McAteer SP, 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. *Mol. Microbiol.* 80(5):1349–65

Tree JJ, Wolfson EB, Wang D, Roe AJ, Gally DL. 2009. Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. *Trends Microbiol.* 17(8):361–70

Trusca D, Scott S, Thompson C, Bramhill D. 1998. Bacterial SOS checkpoint protein SulA inhibits polymerization of purified FtsZ cell division protein. *J. Bacteriol.* 180(15):3946–53

Tso WW, Adler J. 1974. Negative chemotaxis in *Escherichia coli*. *J. Bacteriol.* 118(2):560–76

Tucker DL, Tucker N, Conway T. 2002. Gene expression profiling of the pH response in *Escherichia coli*. *J. Bacteriol.* 184(23):6551–58

Tucker DL, Tucker N, Ma Z, Foster JW, Miranda RL, *et al.* 2003. Genes of the GadX-GadW Regulon in *Escherichia coli*. *J. Bacteriol.* 185(10):3190–3201

Turner L, Ryu WS, Berg HC. 2000. Real-time imaging of fluorescent flagellar filaments. *J. Bacteriol.* 182(10):2793–2801

Turner RD, Vollmer W, Foster SJ. 2014. Different walls for rods and balls: The diversity of peptidoglycan. *Mol. Microbiol.* 91(5):862–74

Typas A, Banzhaf M, Gross C a, Vollmer W. 2012. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* 10(2):123–36

Uhlich GA. 2009. KatP contributes to OxyR-regulated hydrogen peroxide resistance in *Escherichia coli* serotype O157 : H7. *Microbiology*. 155(11):3589–98

Umanski T, Rosenshine I, Friedberg D. 2002. Thermoregulated expression of virulence genes in enteropathogenic *Escherichia coli*. *Microbiology*. 148(9):2735–44

Uranga LA, Balise VD, Benally C V, Grey A, Lusetti SL. 2011. The *Escherichia coli* DinD protein modulates RecA activity by inhibiting postsynaptic RecA filaments. *J. Biol. Chem.* 286(34):29480–91

Van Elsas JD, Semenov A V, Costa R, Trevors JT. 2011. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5(2):173–83

Van Heijenoort J. 2011. Peptidoglycan hydrolases of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 75(4):636–63

Wadhams GH, Armitage JP. 2004. Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* 5(12):1024–37

Wagner PL, Acheson DW, Waldor MK. 2001. Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli. Infect. Immun.* 69(3):1934–37

Weiss SM, Ladwein M, Schmidt D, Ehinger J, Lommel S, *et al.* 2009. IRSp53 links the enterohemorrhagic *E. coli* effectors Tir and EspFU for actin pedestal formation. *Cell Host Microbe*. 5(3):244–58

Welch M, Oosawa K, Aizawa S, Eisenbach M. 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 90(19):8787–91

Welch RA, Burland V, Plunkett G, Redford P, Roesch P, *et al.* 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli. Proc. Natl. Acad. Sci. U. S. A.* 99(26):17020–24

Whitney JG, Grula EA. 1964. Incorporation of D-serine into the cell wall mucopeptide of Micrococcus lysodeikticus. *Biochem. Biophys. Res. Commun.* 14(4):375–81

Wijetunge DSS, Karunathilake KHEM, Chaudhari A, Katani R, Dudley EG, *et al.* 2014. Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated *Escherichia coli* strain RS218. *BMC Microbiol.* 14(1):203

Wiles S, Pickard KM, Peng K, MacDonald TT, Frankel G. 2006. *In vivo* bioluminescence imaging of the murine pathogen *Citrobacter rodentium*. *Infect. Immun.* 74(9):5391–96

Wiles TJ, Kulesus RR, Mulvey MA. 2008. Origins and virulence mechanisms of uropathogenic *Escherichia coli. Exp. Mol. Pathol.* 85(1):11–19

Wilson RK, Shaw RK, Daniell S, Knutton S, Frankel G. 2001. Role of EscF, a putative needle complex protein, in the type III protein translocation system of enteropathogenic *Escherichia coli*. *Cell. Microbiol.* 3(11):753–62

Wolfe AJ, Berg HC. 1989. Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci. U. S. A.* 86(18):6973–77

Wolosker H, Blackshaw S, Snyder SH. 1999. Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission. *Proc. Natl. Acad. Sci. U. S. A.* 96(23):13409–14

Wolosker H, Dumin E, Balan L, Foltyn VN. 2008. D-amino acids in the brain: D-serine in neurotransmission and neurodegeneration. *FEBS J.* 275(14):3514–26

Wong ARC, Pearson JS, Bright MD, Munera D, Robinson KS, *et al.* 2011. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol. Microbiol.* 80(6):1420–38

Wooster DG, Maruvada R, Blom AM, Prasadarao N V. 2006. Logarithmic phase *Escherichia coli* K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation. *Immunology*. 117(4):482–93

Xu X, McAteer SP, Tree JJ, Shaw DJ, Wolfson EBK, *et al.* 2012. Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli. PLoS Pathog.* 8(5):e1002672

Yabu K, Huempfner HR. 1974. Inhibition of growth of *Mycobacterium smegmatis* and of cell wall synthesis by D-serine. *Antimicrob. Agents Chemother.* 6(1):1–10

Yang B, Feng L, Wang F, Wang L. 2015. Enterohaemorrhagic *Escherichia coli* senses low biotin status in the large intestine for colonization and infection. *Nat. Comm.* 6:6592

Yang J, Dogovski C, Hocking D, Tauschek M, Perugini M, Robins-Browne RM. 2009. Bicarbonate-mediated stimulation of RegA, the global virulence regulator from *Citrobacter rodentium*. *J. Mol. Biol.* 394(4):591–99

Yang J, Hocking DM, Cheng C, Dogovski C, Perugini MA, *et al.* 2013. Disarming bacterial virulence through chemical inhibition of the DNA binding domain of an AraC-like transcriptional activator protein. *J. Biol. Chem.* 288(43):31115–26

Yang J, Tauschek M, Hart E, Hartland EL, Robins-Browne RM. 2010. Virulence regulation in *Citrobacter rodentium*: the art of timing. *Microb. Biotechnol.* 3(3):259–68

Yang T-M, Lu C-H, Huang C-R, Tsai H-H, Tsai N-W, *et al.* 2005. Clinical characteristics of adult *Escherichia coli* meningitis. *Jpn. J. Infect. Dis.* 58(3):168–70

Yao Y, Xie Y, Kim KS. 2006. Genomic comparison of *Escherichia coli* K1 strains isolated from the cerebrospinal fluid of patients with meningitis. *Infect. Immun.* 74(4):2196–2206

Yao Y, Xie Y, Perace D, Zhong Y, Lu J, *et al.* 2009. The type III secretion system is involved in the invasion and intracellular survival of *Escherichia coli* K1 in human brain microvascular endothelial cells. *FEMS Microbiol. Lett.* 300(1):18–24

Yona-Nadler C, Umanski T, Aizawa S-I, Friedberg D, Rosenshine I. 2003. Integration host factor (IHF) mediates repression of flagella in enteropathogenic and enterohaemorrhagic *Escherichia coli. Microbiology*. 149(4):877–84

Yousuf FA, Yousuf Z, Iqbal J, Siddiqui R, Khan H, Khan NA. 2014. Interactions of neuropathogenic *Escherichia coli* K1 (RS218) and its derivatives lacking genomic islands with phagocytic Acanthamoeba castellanii and nonphagocytic brain endothelial cells. *Biomed Res. Int.* 2014:265424

Zambelloni R, Marquez R, Roe AJ. 2015. Development of antivirulence compounds: a biochemical review. *Chem. Biol. Drug Des.* 85(1):43–55

Zarivach R, Vuckovic M, Deng W, Finlay BB, Strynadka NCJ. 2007. Structural analysis of a prototypical ATPase from the type III secretion system. *Nat. Struct. Mol. Biol.* 14(2):131–37

Zhang L, Chaudhuri RR, Constantinidou C, Hobman JL, Patel MD, *et al.* 2004. Regulators encoded in the *Escherichia coli* type III secretion system 2 gene cluster influence expression of genes within the locus for enterocyte effacement in enterohemorrhagic *E. coli* O157:H7. *Infect. Immun.* 72(12):7282–93

Zulianello L, de la Gorgue de Rosny E, van Ulsen P, van de Putte P, Goosen N. 1994. The HimA and HimD subunits of integration host factor can specifically bind to DNA as homodimers. *EMBO J.* 13(7):1534–40

8 Appendices

Table 8-1 Additional genes differentially expressed (P < 0.05) in the *ΔyhaO* background as identified by RNA-seq

Name	Fold change	P-value	Function
Z4462	5.61	6.42E-104	D-serine deaminase (*yhaM*)
Z0955	-3.59	1.78E-21	LpxR homologue - putative outer membrane enzyme; lipid modification
metE	2.57	1.92E-17	Methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase
mtr	-4.16	3.59E-08	Tryptophan permease
Z0957	-4.48	1.11E-07	Putative BFP - TfpB like
metF	3.95	4.97E-06	Methylenetetrahydrofolate reductase
Z4332	-2.5	1.26E-05	Putative LifA adherence factor
yfiA	-1.71	3.09E-05	Translation inhibitor protein RaiA
rcsA	-1.92	3.69E-05	Positive response regulator for ctr capsule biosynthesis
rmf	1.6	2.69E-04	Ribosome modulation factor
nadB	2.2	4.07E-04	L-aspartate oxidase
entC	1.57	5.16E-04	Isochorismate synthase
metK	1.61	8.10E-04	Methionine adenosyltransferase
moaE	1.84	8.18E-04	Molybdopterin synthase large subunit
pncB	1.61	1.17E-03	Nicotinate phosphoribosyltransferase
Z1821	-5.97	1.49E-03	Hypothetical
entE	1.48	1.89E-03	Enterobactin exporter (MFS transporter)
Z4386	1.61	2.09E-03	TonB-dependent siderophore receptor
nrdE	1.47	2.70E-03	Ribonucleotide-diphosphate reductase subunit alpha
Z6019	-4.5	2.70E-03	Putative transposase
Z1793	-2.13	3.35E-03	Hypothetical
proP	-1.51	4.08E-03	Proline/glycine betaine transporter (MFS transporter)
ycdB	1.51	4.44E-03	Peroxidase - acidic detoxification; TAT translocated
metN	1.88	4.78E-03	Methionine transporter ATP-binding subunit
gmd	-1.77	5.36E-03	NAD dependent epimerase/dehydratase family
livJ	1.62	5.44E-03	ABC-type branched-chain amino acid (leucine/isoleucine/valine) transporter
nrdH	1.58	5.92E-03	Glutaredoxin-like protein
speD	1.57	0.01	S-adenosylmethionine required for spermidine biosynthesis from putrescine
wcaG	-1.86	0.01	NAD dependent epimerase/dehydratase family
pgaA	-2.04	0.02	Adhesin required for biofilm formation
trpB	-1.54	0.02	Tryptophan synthase subunit beta
Z4849	-1.5	0.02	Predicted small periplasmic lipoprotein
moaC	1.87	0.02	Molybdenum cofactor biosynthesis protein C
ycdO	1.48	0.02	TAT translocated
Z0956	-2.6	0.02	Putative antiterminator Q protein of prophage CP-933K
hypD	-1.83	0.02	Hydrogenase maturation factor
ompA	1.38	0.02	Putative porin/phage receptor
Z1341	-1.5	0.02	Esterase/lipase

Z3297	-1.46	0.02	Putative transposase for IS629
Z3931	-1.6	0.02	Putative membrane protein - tfpB like
ompC	1.48	0.02	Outer membrane porin protein C
Z4325	-35.2	0.03	Putative integrase ISEc8
zraP	-2.31	0.03	P pilus assembly/Cpx signalling pathway,
leuO	-2.54	0.03	Activator for *leuABCD* operon
Z2082	2.76	0.03	Putative transposase within CP-933O
speA	1.39	0.03	Catalyses the formation of agmatine from arginine
htrL	1.74	0.04	Hypothetical
yhcK	-1.59	0.04	Transcriptional repressor of the nan operon (sialic acid utilisation)
ftsX	1.47	0.04	Cell division protein; putative ABC transporter
csgD	1.71	0.04	DNA binding regulator (LuxR) activates the *csgBA* and *csgDEFG* operons involved in biofilm formation
Z2351	29.04	0.04	Putative tail component of prophage CP-933R
Z3561	1.74	0.05	Hypothetical

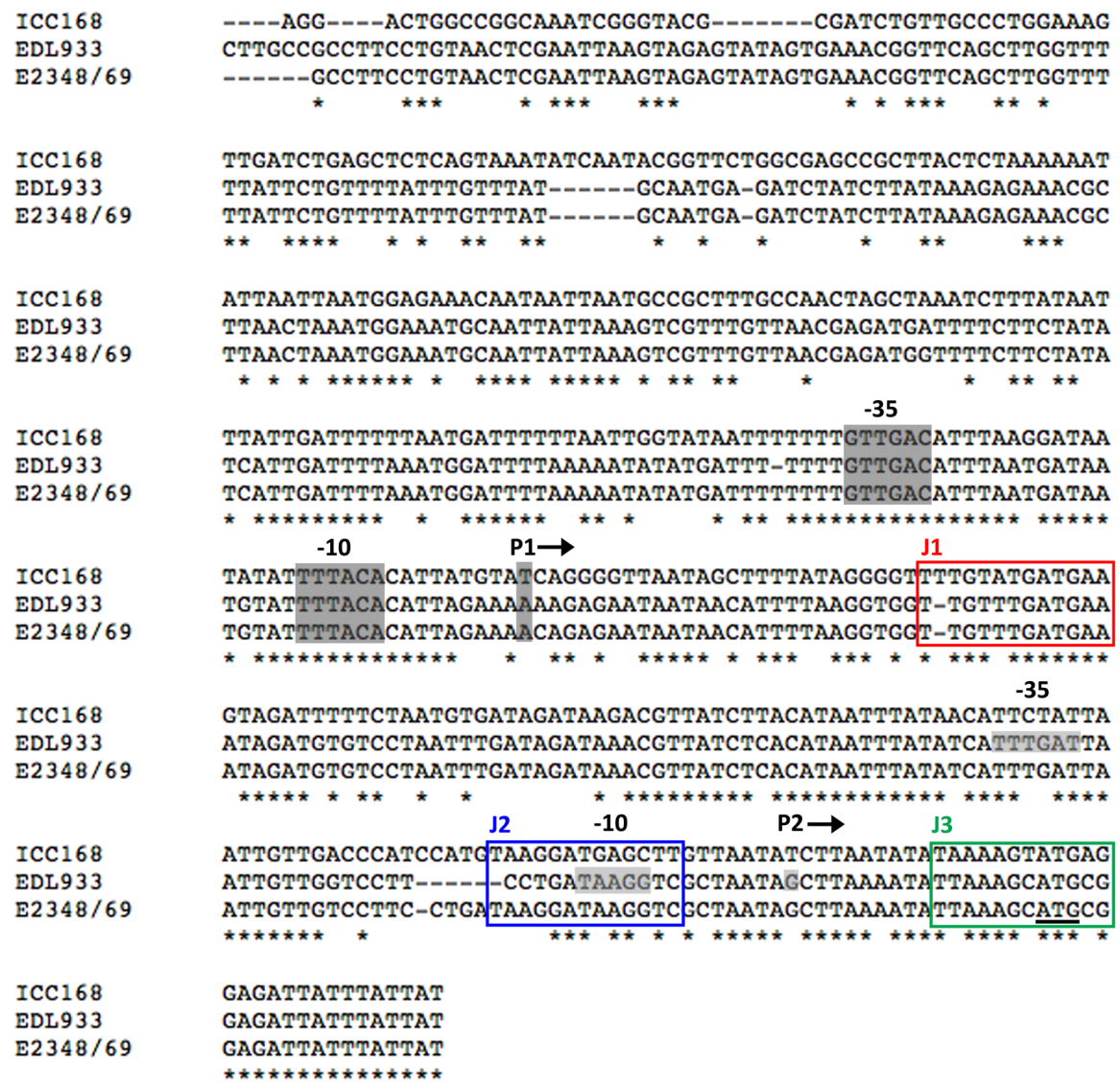

Figure 8-1 Alignment of the *ler* promoter region from EHEC, EPEC and *C. rodentium*. DNA sequence alignments of 400 bp upstream of the *ler* translational start codon (ATG underlined in black) from *C. rodentium* (ICC168), EHEC (EDL933) and EPEC (E2348/69). The LEE1 P1 distal promoter -35/-10 elements and the transcriptional start site (P1 arrow) arrow highlighted by dark grey boxes. The EHEC specific P2 transcriptional start site and associated -10/-35 elements are highlighted by light grey boxes. The putative YhaJ binding boxes are indicated in red (box 1), blue (box 2) and green (box 3). An asterisk (*) indicates a conserved base between all three sequences. A dash (-) indicates a gap in the sequence. The sequences were downloaded from NCBI and aligned using Clustal Omega. Figure adapted from Sharp & Sperandio 2007.

Table 8-2 Additional genes differentially expressed (P < 0.05) in response to D-serine

Name	Log^2 FC	P-value	Function
garR	6.291	9.9E-74	Tartronate semialdehyde reductase
Z4105	5.965	4.0E-30	Putative transport protein
yhaF	5.534	4.3E-41	Carbon utilisation
yhaU	5.241	1.5E-24	Putative D-galactarate transporter
yhaG	4.743	3.0E-21	D-galactarate dehydratase
yhaD	4.358	1.2E-40	Glycerate kinase
pspB	-4.340	2.1E-38	Phage shock protein
pspA	-4.255	3.9E-42	Phage shock protein
pspC	-4.248	1.6E-38	Phage shock protein
pspD	-3.934	4.1E-31	Phage shock protein
Z0314	3.724	1.7E-11	Hypothetical
yjbE	-3.614	5.1E-22	Extracellular polysaccharide (biofilm) production
ydaJ	-3.578	1.0E-20	Required for p-aminobenzoyl-glutamate utilization
cspA	3.506	5.9E-27	Major cold shock protein
intH	3.480	1.1E-36	Putative integrase for prophage CP-933H
ygcY	3.445	2.2E-13	Putative glucarate dehydratase
yahO	-3.426	1.9E-09	Predicted periplasmic protein, YhcN family
ygcX	3.416	5.1E-19	D-glucarate dehydratase; L-idarate dehydratase
tnaA	3.337	4.6E-20	Tryptophan indole-lyase
phoE	3.274	1.7E-03	Outer membrane phosphoporin protein E
yehV	-3.212	7.6E-05	Putative transcriptional regulator
csgB	-3.202	7.3E-24	Curlin nucleator protein
csgA	-3.184	1.8E-25	Curlin subunit
Z3921	-3.171	9.6E-05	NleG6-3
Z0312	3.095	2.0E-07	Partial O replication protein for prophage CP-933H
Z0461	3.017	5.0E-12	Putative permease; hexosephosphate transport
Z2427	-2.959	8.8E-16	Hypothetical
yigN	2.950	2.1E-16	DNA recombination protein
Z2225	-2.914	5.5E-05	ATP-dependent peptide transporter membrane subunit
pspE	-2.902	1.4E-20	Phage shock protein
Z2974	2.893	4.6E-19	Hypothetical
ilvN	2.850	6.3E-07	Acetolactate synthase
Z4104	2.838	7.1E-07	Glucokinase
Z2978	2.813	7.6E-13	Replication protein for prophage CP-933T
Z2223	-2.804	4.5E-14	ABC-type dipeptide transport system
Z2563	2.708	1.3E-05	Putative transposase
Z2979	2.686	1.6E-15	Stability/partitioning protein within prophage CP-933T
arsB	2.662	2.8E-03	Arsenite pump, inner membrane
sapB	-2.625	6.6E-06	Component of peptide uptake ABC transporter
Z0315	2.595	7.4E-04	Hypothetical
yehD	-2.567	6.9E-04	Predicted fimbrial-like adhesin protein
yddB	-2.562	4.2E-12	Putative TonB-dependent outer membrane receptor
Z0317	2.561	1.6E-05	Putative tail fiber protein from prophage CP-933H
Z0923	-2.533	4.7E-11	Putative homeobox protein
wcaH	-2.483	1.2E-03	GDP-mannose mannosyl hydrolase
wcaG	-2.369	1.4E-07	Putative nucleotide di-P-sugar epimerase or dehydratase

asnA	2.350	4.1E-17	Asparagine synthase A
lacY	2.327	3.5E-03	Lactose permease
Z2428	-2.292	8.7E-08	Pump protein
Z1510	-2.266	2.3E-09	Putative synthetase
pqqL	-2.265	7.0E-12	Putative zinc peptidase
Z2222	-2.248	2.0E-04	D-Ala-D-Ala dipeptidase
afuA	2.225	1.5E-07	ABC transporter protein
ybhO	-2.210	2.9E-04	Cardiolipin synthase 2
yihM	2.200	7.9E-05	Hypothetical
ybjM	2.194	3.8E-07	Predicted inner membrane protein
rcsA	-2.184	8.7E-09	Positive regulator for ctr capsule biosynthesis
Z0463	2.181	1.2E-05	Putative response regulator
ftn	-2.159	9.2E-07	Ferritin; negatively regulated by ryhB RNA as part of indirect positive Transcriptional regulation by Fur
Z2250	-2.155	4.5E-04	N-hydroxyarylamine O-acetyltransferase
Z1356	2.153	3.3E-03	Hypothetical
yohC	-2.142	2.1E-06	Hypothetical
yihL	2.136	1.5E-03	Putative transcriptional regulator
katE	-2.121	6.4E-09	Catalase hydroperoxidase II; response to oxidative stress
ycaC	-2.098	2.2E-06	Putative cysteine hydrolase
osmY	-2.084	8.8E-12	Osmotically induced periplasmic protein
Z1511	-2.084	6.3E-10	Pyrimidine utilization protein A
gntU	2.081	1.2E-05	Low affinity gluconate transporter
wcaM	-2.080	1.9E-04	Putative colanic acid biosynthesis protein
poxB	-2.080	1.1E-10	Pyruvate oxidase
nadB	2.073	4.3E-08	Quinolinate synthase, B protein
nadA	2.070	1.9E-08	Quinolinate synthase
ybaY	-2.066	2.2E-10	Predicted outer membrane lipoprotein
tktB	-2.066	2.0E-12	Transketolase B; binds Zn(II)
dadX	-2.023	2.9E-10	Alanine racemase
Z2224	-2.019	3.1E-04	Putative transport system permease protein
yrbA	2.018	1.2E-04	Hypothetical
wcaE	-2.014	2.6E-03	Putative glycosyl transferase
ftrA	-2.003	2.0E-04	Transcriptional regulation of genes involved in carbon and Amino Acid Transport and Metabolism
Z0316	1.996	5.7E-06	Hypothetical
Z4354	1.995	3.7E-07	Aldo-keto reductase
lldP	-1.969	4.4E-09	Dual role activator/repressor for lldPRD operon
yjfR	1.967	1.1E-03	Putative L-ascorbate 6-phosphate lactonase
yidF	1.962	2.0E-05	Putative DNA-binding transcriptional regulator
Z1617	-1.912	4.8E-12	Bifunctional enterobactin receptor/adhesin
adhP	-1.907	7.4E-09	Ethanol-active dehydrogenase/acetaldehyde-active reductase
gcvH	1.905	2.0E-08	Glycine cleavage
Z4018	-1.893	2.8E-03	Anaerobic nitric oxide reductase
Z2186	-1.890	5.3E-04	Trans-aconitate 2-methyltransferase
Z3249	-1.883	6.3E-09	Hypothetical
ddg	1.874	8.4E-07	Lipid A biosynthesis
lldR	-1.870	1.0E-06	Dual role activator/repressor for lldPRD operon

emrD	1.870	7.7E-04	Multidrug efflux protein
ecnB	-1.855	7.1E-10	Bacteriolytic lipoprotein entericidin B toxin
Z5547	1.847	2.6E-03	Hypothetical
Z2299	-1.838	3.3E-05	Putative transcriptional regulator LYSR-type
ycgB	-1.831	1.4E-07	SpoVR family protein
Z3260	-1.824	1.8E-09	Fructose-bisphosphate aldolase
gcvP	1.823	2.6E-11	Glycine dehydrogenase
Z2665	-1.822	1.5E-03	Hypothetical
ygjG	-1.822	7.9E-10	Putrescine:2-oxoglutaric acid aminotransferase; sigma 54-dependent
Z0753	1.821	3.0E-04	Hypothetical
Z3043	-1.815	1.8E-08	Hypothetical
hypA	1.810	5.6E-03	Hydrogenase 3 accessory protein
gmd	-1.808	8.7E-06	NAD dependent epimerase/dehydratase
wcaA	-1.801	2.8E-03	Putative glycosyl transferase
potH	-1.794	7.4E-05	Putrescine transporter subunit
Z4874	1.781	7.7E-05	Putative regulator
talA	-1.780	4.5E-09	Transaldolase A; creBC regulon
Z4353	-1.777	2.7E-04	Putative esterase/lipase
cpsB	-1.775	2.5E-04	Mannose-1-phosphate guanyltransferase
oriC	1.774	5.8E-06	Nucleoid-associated protein
malT	1.767	1.8E-11	Transcriptional activator for the mal regulon
Z1629	1.759	1.8E-03	Hypothetical
rihA	1.759	7.0E-05	Ribonucleoside hydrolase
aspA	1.753	2.4E-11	L-Aspartate ammonia-lyase
elaB	-1.748	1.1E-08	Hypothetical
Z3388	-1.741	5.5E-06	Hypothetical
Z0955	-1.728	5.2E-08	Hypothetical
yaeG	1.720	9.1E-06	Carbohydrate diacid transcriptional activator
bssR	-1.718	3.9E-07	Repressor of biofilm formation
potG	-1.715	1.0E-05	Putrescine ABC transporter
ybdL	-1.708	3.4E-04	Putative aminotransferase
yicP	1.698	4.1E-03	Adenine deaminase
rnk	1.695	1.1E-04	Nucleoside diphosphate kinase regulator
ybdK	-1.692	2.6E-03	Cysteine ligase
Z1924	-1.688	1.8E-04	Hypothetical
yeaX	-1.681	1.2E-03	Putative diogenase beta subunit
dps	-1.676	7.1E-10	DNA starvation/stationary phase protection protein
Z4398	1.672	5.0E-03	Hypothetical
ygeV	1.668	1.5E-08	Putative RpoN-dependent transcriptional activator
ilvB	1.664	2.1E-06	Acetohydroxy acid synthase I (AHAS-I)
yddA	-1.657	7.4E-04	Putative ATP-binding component of a transport system
Z2695	-1.654	2.2E-03	Hypothetical
Z2387	-1.651	4.2E-05	Hypothetical
yedL	-1.643	2.4E-06	N-acetyltransferases
Z3140	1.640	1.6E-04	Hypothetical
Z2777	-1.638	1.9E-08	Succinylarginine dihydrolase
astD	-1.636	1.7E-08	Succinylglutamic semialdehyde dehydrogenase
yciE	-1.631	1.2E-04	Hypothetical
sapF	-1.627	1.6E-04	Putative ATP-binding protein of peptide transport

			system
yeaQ	-1.624	8.4E-06	Hypothetical
ligB	1.622	3.2E-03	NAD-dependent DNA ligase
yeiA	1.619	4.8E-03	Dihydropyrimidine dehydrogenase
gcvT	1.619	5.2E-08	Aminomethyl transferase
metF	-1.614	3.0E-07	5,10-Methylenetetrahydrofolate reductase
ylaD	1.596	3.1E-04	Maltose O-acetyltransferase
dadA	-1.590	4.1E-08	D-amino acid dehydrogenase
Z2779	-1.589	3.5E-07	Arginine succinyltransferase
tatC	1.575	1.9E-06	Protein translocase, Sec-independent
yhdY	-1.573	1.1E-03	Putative transport system permease protein
fruA	1.572	1.4E-09	Fructose-specific PTS system IIBC component
ybaA	-1.562	3.1E-03	Hypothetical
Z1509	-1.558	1.8E-04	Hypothetical
otsB	-1.553	5.1E-04	Trehalose-6-phosphate phosphatase
argT	-1.552	5.1E-08	Lysine-, arginine-, ornithine-binding periplasmic protein
Z1507	-1.551	2.2E-04	Hypothetical
fruB	1.549	2.1E-09	Fructose-specific PTS IIA/HPr protein
nupG	1.543	1.4E-06	Nucleoside transporter
glpD	-1.539	1.8E-04	Glycerol-3-phosphate dehydrogenase
ychH	-1.538	1.4E-07	Predicted inner membrane protein
gltP	1.534	4.3E-07	Glutamate/aspartate proton symporter
Z2301	-1.526	6.5E-06	Hypothetical
dbpA	-1.524	1.3E-03	ATP-dependent RNA helicase
Z5978	-1.521	6.1E-06	Hypothetical
yeaT	1.521	2.7E-03	Putative transcriptional regulator LYSR-type
Z1508	-1.511	5.8E-05	Putative acetyltransferase
ribB	1.508	1.1E-07	DHBP synthase; stress induced
potI	-1.506	1.2E-04	Putrescine transporter subunit
lysC	1.506	2.3E-07	Aspartokinase III
fruK	1.496	1.0E-08	1-phosphofructokinase
cpsG	-1.484	2.3E-03	Phosphomannomutase
melR	1.475	1.2E-04	DNA-binding transcriptional regulator MelR
yebF	1.465	3.2E-07	Exported protein, function unknown
sdhB	-1.456	1.9E-06	Succinate dehydrogenase iron-sulfur subunit
Z2565	1.454	1.6E-07	Chaperone
Z2040	-1.450	4.0E-04	Hypothetical
tsx	1.448	3.7E-07	Outer membrane channel for nucleosides
acnA	-1.446	2.5E-06	Aconitate hydratase
ycdG	-1.434	1.2E-05	Putative pyrimidine permease, pyrimidine nitrogen catabolism
yjbH	-1.433	5.6E-03	Predicted porin
rbsA	1.431	5.2E-03	D-ribose high-affinity transport system
Z1930	-1.427	4.1E-06	Putative protease encoded within prophage CP-933X
ackA	1.427	3.4E-07	Acetate kinase
grxB	-1.423	5.4E-06	Glutaredoxin 2; regulated by RpoS and ppGpp
sugE	1.420	5.6E-03	Multidrug efflux pump
Z1120	1.416	1.7E-03	Bacteriophage P4 integrase
yacC	1.416	4.6E-03	Hypothetical
proP	-1.406	1.2E-06	Proline/betaine permease

yadI	1.405	3.8E-04	Hypothetical
yggB	-1.397	8.8E-07	Mechanosensitive channel protein
yidZ	1.396	4.3E-03	Putative transcriptional regulator LysR-type
etp	-1.391	1.8E-04	General secretory pathway protein E
metA	-1.390	5.5E-05	Homoserine O-succinyltransferase
Z3513	-1.380	1.6E-03	Formyl transferase
Z3931	-1.373	1.5E-03	Hypothetical
yiaG	-1.366	2.3E-03	Putative transcriptional regulator
malK	1.365	1.3E-04	Maltose/maltodextrin transporter ATP-binding protein
ygaP	1.363	4.9E-04	Hypothetical
ydjS	-1.361	4.6E-06	Probable succinylglutamate desuccinylase, arginine catabolism
Z1931	-1.361	4.7E-07	Outer membrane protease
yedU	-1.361	1.3E-06	Chaperone protein
pnuC	1.359	1.9E-04	Nicotinamide mononucleotide transporter
deaD	1.355	1.6E-07	Post-transcriptional Regulation factor W2, putative RNA helix-destabilizer
ydcP	-1.350	2.5E-03	Putative collagenase
feoA	1.342	1.9E-04	Ferrous iron uptake, required for full FeoB activity
phrB	-1.337	2.5E-03	Deoxyribodipyrimidine photolyase
potF	-1.332	9.9E-05	Spermidine/putrescine-binding periplasmic protein
wrbA	-1.327	7.9E-05	Stationary phase protein that binds TrpR repressor
yqiA	1.324	1.4E-04	Palmitoyl-CoA and pNP-butyrate esterase
yoaE	-1.319	9.2E-06	Predicted membrane protein
ynaJ	-1.318	1.3E-04	Hypothetical
rbsK	1.314	5.1E-04	Ribokinase
nikA	1.313	2.6E-03	Heme-binding periplasmic protein
yccZ	-1.302	1.6E-04	Polysaccharide biosynthesis/export protein
Z2971	1.300	1.8E-06	Hypothetical
Z4776	1.298	6.8E-04	Hypothetical
yjbA	-1.297	5.3E-04	Phosphate-starvation-inducible protein
pspG	-1.296	8.6E-06	Phage shock protein
greA	1.291	2.4E-04	Transcription elongation factor
gidA	1.285	7.1E-06	tRNA uridine 5-carboxymethylaminomethyl modification enzyme
yegH	-1.268	3.4E-04	Putative transport protein
yjbC	1.268	2.3E-03	23S rRNA pseudouridine synthase
gntT	1.264	2.5E-04	High-affinity transport of gluconate / gluconate permease
argD	-1.257	5.9E-05	Acetylornithine aminotransferase
malF	1.254	2.1E-04	Maltose transport complex, inner membrane-spanning subunit
Z1186	1.252	3.7E-03	Hypothetical
ygiW	-1.250	7.1E-04	Hypothetical
frdA	1.246	9.8E-05	Fumarate reductase flavoprotein
codB	-1.243	3.3E-03	Cytosine transport
yigM	1.226	2.8E-04	Hypothetical
Z2973	1.223	2.1E-05	Hypothetical
Z5881	1.223	2.2E-03	Hypothetical
Z0370	-1.213	3.2E-03	Hypothetical

gntK	1.213	2.3E-03	Gluconate kinase
fimB	1.206	4.7E-03	Regulator for fimA
aceA	-1.204	2.5E-04	Isocitrate lyase
mltD	1.203	1.4E-05	Membrane-bound lytic murein transglycosylase
fdoH	-1.201	2.5E-03	Formate dehydrogenase-O
yqhA	-1.197	2.4E-03	Conserved inner membrane protein
ilvL	-1.189	4.6E-04	*ilvG* operon leader peptide
ygjR	1.187	3.2E-04	Hypothetical
Z1038	-1.183	1.8E-04	Hypothetical
otsA	-1.181	3.0E-04	Trehalose-6-phosphate synthase
tyrB	1.180	1.4E-04	Aspartate aminotransferase
pckA	1.168	6.8E-06	Phosphoenolpyruvate carboxykinase
ydcF	-1.167	3.9E-03	Hypothetical
Z2713	-1.163	5.5E-03	Hypothetical
yfiQ	1.158	5.7E-05	Fused predicted acyl-CoA synthetase
slmA	1.151	2.5E-03	Nucleoid occlusion protein
nrdD	-1.145	1.3E-03	Ribonucleoside triphosphate reductase
pyrB	-1.141	5.2E-04	Aspartate carbamoyltransferase catalytic subunit
ygaM	-1.139	1.9E-03	Hypothetical
gmk	1.135	1.2E-04	Guanylate kinase
gapC	-1.134	3.1E-03	Glyceraldehyde 3-phosphate dehydrogenase
yhbC	1.130	1.7E-04	Hypothetical
frdB	1.124	4.5E-03	Fumarate reductase
Z4385	1.123	1.1E-04	ABC transporter ATP-binding protein
feoB	1.121	2.3E-05	Ferrous iron uptake GTP-binding membrane protein
ygfA	1.117	4.6E-03	5-formyltetrahydrofolate cyclo-ligase
yabB	1.116	1.1E-04	Cell division protein
radC	1.115	7.6E-04	DNA repair protein
Z2771	1.112	4.9E-03	Nucleotide excision repair endonuclease
Z1026	-1.110	2.5E-03	Catecholate siderophore receptor Fiu
Z2714	-1.109	6.2E-04	Hypothetical
Z2287	-1.108	5.5E-03	Hypothetical
Z1121	1.106	2.9E-03	Hypothetical
Z3731	-1.103	1.8E-03	Hypothetical
Z2972	1.100	6.8E-05	Hypothetical
metN	-1.094	6.2E-04	DL-methionine transporter ATP-binding subunit
gabD	-1.090	1.9E-03	Succinate-semialdehyde dehydrogenase I
Z1506	-1.085	5.2E-03	Putative 4-hydroxyphenylacetate 3-monooxygenase
acnB	-1.085	5.5E-05	2-methylaconitate hydratase
mioC	1.079	3.8E-04	NADPH-dependent FMN reductase
metK	-1.072	1.1E-04	S-adenosylmethionine synthase
cyoC	-1.068	1.9E-04	Cytochrome o ubiquinol oxidase subunit III
kgtP	-1.068	2.5E-04	Alpha-ketoglutarate transporter
yjdA	1.064	3.0E-03	Hypothetical
dnaA	1.058	1.2E-04	Chromosomal replication initiation protein
hisG	1.055	1.6E-03	ATP phosphoribosyltransferase
Z5898	-1.036	1.5E-04	Hypothetical
cyoD	-1.036	1.9E-04	Cytochrome o ubiquinol oxidase subunit IV
yigA	1.031	9.8E-04	Hypothetical
yjjY	-1.027	5.2E-04	Hypothetical

yeeA	1.024	3.7E-03	Hypothetical
ycfX	-1.021	1.0E-03	N-acetyl-D-glucosamine kinase
Z2888	1.019	3.6E-03	Putative copper resistance protein
purK	-1.017	2.4E-03	N5-carboxyaminoimidazole ribonucleotide synthase
leuA	1.017	1.3E-03	Leucine biosynthesis
hdeB	1.012	8.7E-04	Acid-resistance protein
mdoH	-1.011	4.0E-04	Glucosyltransferase
yeaD	-1.011	5.4E-04	Hypothetical
dcp	-1.010	5.7E-04	Dipeptidyl carboxypeptidase II
sdhA	-1.009	3.4E-04	Succinate dehydrogenase
ilvC	1.007	1.1E-04	Ketol-acid reductoisomerase
xerC	1.003	1.7E-03	Tyrosine recombinase
Z5098	1.000	3.2E-03	Transposase encoded within CP-933L
Z3117	-0.999	2.3E-04	Hypothetical
cydA	0.994	1.6E-04	Cytochrome d (bd-I) terminal oxidase subunit I
hdhA	-0.989	9.8E-04	7-alpha-hydroxysteroid dehydrogenase
Z2386	-0.977	1.7E-03	Hypothetical
trmE	0.976	2.2E-03	tRNA modification GTPase; acids stress induced
atpI	0.970	1.4E-03	Produces ATP from ADP in the presence of a proton gradient across the membrane
Z2267	-0.967	2.5E-04	Putative receptor
gltA	-0.967	3.0E-04	Citrate synthase; hexameric
Z2592	-0.961	5.5E-03	Hypothetical
wzxE	0.957	5.3E-03	Putative cytochrome
Z4386	0.955	7.3E-04	Putative iron compound receptor
Z4237	0.952	1.0E-03	Putative oxidoreductase
yjiY	0.946	1.5E-03	Putative carbon starvation protein
Z6021	-0.944	3.4E-03	Hypothetical
rhoL	0.941	3.8E-03	Putative rho operon leader peptide
yieM	0.939	2.3E-03	Hypothetical
dapF	0.936	1.5E-03	Diaminopimelate epimerase
birA	0.924	2.7E-03	Biotin protein ligase
phoH	-0.924	9.7E-04	Phosphate starvation-inducible protein
yccJ	-0.923	5.4E-03	Hypothetical
Z1592	-0.922	6.1E-04	Hypothetical
tatB	0.915	1.1E-03	Sec-independent translocase
yjeQ	0.912	4.8E-03	Ribosome-associated GTPase
ydgA	-0.911	1.2E-03	Hypothetical
lldD	-0.911	3.2E-03	L-lactate dehydrogenase
livJ	0.905	2.9E-03	High-affinity amino acid transport system
mdoG	-0.904	2.3E-03	Periplasmic glucan biosynthesis protein
ytfM	0.900	4.4E-03	Outer membrane protein
cydB	0.899	7.0E-04	Cytochrome bd-type quinol oxidase
osmE	-0.899	3.2E-03	Osmotically-inducible lipoprotein E
rho	0.898	8.6E-04	Rho-binding antiterminator
yjbJ	-0.896	3.5E-03	Putative stress related protein
pheT	-0.892	7.9E-04	Phenylalanyl-tRNA synthetase
xasA	-0.891	8.7E-04	Putative transporter
cyoB	-0.890	1.3E-03	Cytochrome o ubiquinol oxidase subunit I
NsrR	0.879	2.9E-03	Transcriptional repressor

ucpA	0.868	4.9E-03	Short chain dehydrogenase
ispB	0.866	4.9E-03	Octaprenyl diphosphate synthase
hns	-0.865	1.4E-03	Global DNA-binding transcriptional dual regulator
secE	0.863	2.2E-03	SecE/Sec61-gamma subunits of protein translocation complex
hemE	0.859	4.5E-03	Uroporphyrinogen decarboxylase
nac	-0.856	1.7E-03	Nitrogen assimilation transcriptional regulator
mraW	0.855	1.8E-03	S-adenosyl-methyltransferase
rfaF	0.854	2.7E-03	ADP-heptose:LPS heptosyltransferase II
mdh	-0.851	1.4E-03	Malate dehydrogenase
nusA	0.850	1.2E-03	Transcription elongation factor
cbl	-0.847	2.9E-03	LysR-type transcriptional regulator
Z1153	-0.841	1.7E-03	Hypothetical
purL	-0.840	1.9E-03	Phosphoribosylformylglycinamide synthase
pykF	-0.835	2.6E-03	Pyruvate kinase
fdoG	-0.835	4.9E-03	Formate dehydrogenase-O
tpx	-0.831	1.8E-03	Thiol peroxidase
nusG	0.823	2.7E-03	Modulates Rho-dependent transcription termination
Z2054	-0.822	1.9E-03	Putative killer protein encoded by prophage CP-933O
glk	0.814	4.1E-03	Hypothetical
slyB	-0.810	3.2E-03	Putative outer membrane protein
icc	0.797	4.5E-03	Cyclic 3',5'-adenosine monophosphate phosphodiesterase
csrA	-0.797	3.7E-03	Carbon storage regulator
potD	-0.791	4.1E-03	Spermidine/putrescine-binding periplasmic protein
glnK	-0.790	4.1E-03	Nitrogen regulatory protein
rmf	-0.766	3.5E-03	Ribosome modulation factor
rnpB	-0.745	4.3E-03	RNase P
Z2099	-0.732	5.6E-03	Hypothetical

Table 8-3 Summarised genome information of LEE+/*dsdCXA*+ isolates

Isolate[*1]	**Phylogroup**	**Sample source**[*2]
E. coli TW07793_AFAG	B2	Water
E. coli C295_10_AIAR	B2	Stool
E. coli KTE84_ANWB	B2	NA
E. coli 07798_AMUP	B2	NA
E. coli C93_11_AICD	B2	Stool
E. coli 3431_ADUM	D	NA
E. coli C353_09_AIAV	D	Stool
E. coli 5_0959_AEZX	B1	NA
E. coli O121_H19_str_MT_2_AGTJ	B1	NA
E. coli C238_91_AIAN	B1	Stool
E. coli O103_H25_str_CVM9340_AJVQ	B1	NA
E. coli O103_H25_str_NIPH_11060424_AGSG	B1	Stool (HUS patient)
E. coli M919_AEVZ	A	NA
E. coli H383	A	Stool
E. coli C260_92_AIAO	A	Stool
E. coli C213_10_AIAL	A	Stool

E. coli UMEA_3317_1_AWDF	A	Stool
E. coli 3_2303_AFAE	A	Water
E. coli C496_10_AIBB	A	Stool
E. coli TW00353_AMUM	A	NA
E. coli TW15901_AMUK	A	Beef
E. coli C732_98_AIBM	A	Stool
E. coli C751_03_AIBO	A	Stool
E. coli ATCC_35150_AWXM	E	NA
E. coli C161_11_AIAI	E	Stool
E. coli O145_H28_str_4865_96_AGTL	E	Stool (HUS patient)

*1 The isolate names refer to the genome annotation in NCBI. The four letter abbreviations following the isolate name refer to the NCBI accession number for that particular whole genome shotgun sequence and are not yet completed genomes.
*2 NA indicates unavailable information on the source of this isolate according to NCBI.

Table 8-4 Summary of isolates with an incomplete *yhaOKJ*[*1]

Isolate	*yhaJ*	*yhaK*	*yhaO*	*dsdCXA*[*2]	LEE[*3]
E. coli H263_AEFJ	+	-	+	+	-
E. coli TOP382_3_AOQE	+	+	-	+	-
E. coli TOP293_3_AOQI	-	+	+	+	-
E. coli 4_1_47FAA_ACTQ	-	+	+	-	-
E. coli 907715_AXTI	+	+	-	-	-
E. coli HVH_183_4_3205932_AVYJ	-	-	-	-	-
E. coli 908521_AXTM	-	-	-	-	-
S. boydii 965_58_AKNA	-	-	-	-	-
S. dysenteriae 1012_AAMJ	+	+	-	+	-
S. dysenteriae 155_74_AFFZ	-	-	-	+	-
S. flexneri K_227_AFGY	+	+	-	-	-
S. flexneri K_272_AFGX	+	+	-	-	-
E. coli SCD1_ATJZ	-	-	+	+	-
E. coli 2845350_AQDS	+	+	-	+	-
E. coli OP50_ADBT	-	+	-	+	-
E. coli ECA_727_AHHL	+	-	+	-	-
E. coli TOP2515_AOQT	-	+	+	-	-
E. coli MS_175_1_ADUB	-	-	-	-	-
E. coli MS_116_1_ADTZ	-	-	-	-	-
E. coli XH140A_AFVX	-	-	+	+	-
E. coli O157_H7_str_H093800014_CAOT	-	+	+	-	+
E. coli M21_AWQG	-	-	-	-	+
E. coli K1_AWQJ	-	-	-	-	+

*1 Presence or absence of either *yhaJ, yhaK* or *yhaO* is indicated by a + or – respectively
*2 The *dsdCXA* locus is considered present if intact according to the criteria in Figure 5-1
*3 The LEE is considered present according to the criteria in Figure 5-1

9 Publications

OPEN

The ISME Journal (2015) 9, 1039–1051

www.nature.com/ismej

ORIGINAL ARTICLE

The host metabolite D-serine contributes to bacterial niche specificity through gene selection

This article has been corrected since Advance Online Publication and an erratum is also printed in this issue

James PR Connolly[1,2], Robert J Goldstone[1,2], Karl Burgess[1,2], Richard J Cogdell[3], Scott A Beatson[4], Waldemar Vollmer[5], David GE Smith[1,2,6] and Andrew J Roe[1,2]

[1]*Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK;* [2]*School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK;* [3]*Institute of Molecular Cell and Systems Biology, University of Glasgow, Glasgow, UK;* [4]*School of Chemistry and Molecular Biosciences and Australian Infectious Diseases Research Centre, University of Queensland, St Lucia, Queensland, Australia;* [5]*Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK and* [6]*Moredun Research Institute, Pentlands Science Park, Bush Loan, Edinburgh, Midlothian, UK*

Escherichia coli **comprise a diverse array of both commensals and niche-specific pathotypes. The ability to cause disease results from both carriage of specific virulence factors and regulatory control of these *via* environmental stimuli. Moreover, host metabolites further refine the response of bacteria to their environment and can dramatically affect the outcome of the host–pathogen interaction. Here, we demonstrate that the host metabolite, D-serine, selectively affects gene expression in *E. coli* O157:H7. Transcriptomic profiling showed exposure to D-serine results in activation of the SOS response and suppresses expression of the Type 3 Secretion System (T3SS) used to attach to host cells. We also show that concurrent carriage of both the D-serine tolerance locus (*dsdCXA*) and the locus of enterocyte effacement pathogenicity island encoding a T3SS is extremely rare, a genotype that we attribute to an 'evolutionary incompatibility' between the two loci. This study demonstrates the importance of co-operation between both core and pathogenic genetic elements in defining niche specificity.**
The ISME Journal (2015) **9,** 1039–1051; doi:10.1038/ismej.2014.242; published online 19 December 2014

Introduction

Escherichia coli is a diverse Gram-negative bacterium that comprises several subgroups (A, B1, B2, C, D, E and F) based on genetic phylogeny (Wirth *et al.*, 2006; Sims and Kim, 2011). The genetic variation between these groups is vast with the 'core' genome consisting of a mere ~2000 genes in contrast to the *E. coli* pan genome of ~18 000 genes (Van, Elsas *et al.*, 2010). Within this variation, the acquisition and loss of genomic islands between strains, both in the context of virulence and environmental adaptation, plays a key role in defining niche specificity (Dobrindt *et al.*, 2004). Furthermore, the function of the core genome in defining a strain's metabolic capacity contributes to the ability of *E. coli* strains to occupy distinct ecological niches (Touchon *et al.*, 2009; Van, Elsas *et al.*, 2010; De Muinck *et al.*, 2013). Despite the vast array of research that has been carried out to understand the physiology, biochemistry and genetics of both pathogenic and commensal *E. coli* strains, identifying the specific attributes that define a strain's favored niche remains challenging.

E. coli within phylogroup B2 provide intriguing examples of niche diversification from intestinal commensalism and pathogenesis to strains highly virulent in the urinary tract (uropathogenic *E. coli* or UPEC) and other extra-intestinal sites (Kaper *et al.*, 2004; Brzuszkiewicz *et al.*, 2006; Wiles *et al.*, 2009). The latter are termed extraintestinal pathogenic *E. coli* because of their ability to cause disease beyond the gastrointestinal tract. Through the acquisition of combinations of virulence factors, such as fimbrial adhesins, capsule and iron acquisition systems, UPEC strains compete and thrive in the bladder, an organ very different to the intestine. Despite these defining features, UPEC strains continue to passage through the digestive system very successfully, without any apparent loss in intestinal fitness (Chen *et al.*, 2013). In contrast, intestinal pathogenic *E. coli* such as enterohaemorrhagic *E. coli* (EHEC) O157:H7, are highly niche-specific and are very rarely associated with the colonization of distal sites

Correspondence: AJ Roe, Institute of Infection, Immunity and Inflammation, University of Glasgow, 120 University Avenue, Glasgow G12 8TA, USA.
E-mail: andrew.roe@glasgow.ac.uk
Received 30 June 2014; revised 7 October 2014; accepted 13 November 2014; published online 19 December 2014

(Kaper *et al*., 2004; Wong *et al*., 2011). This is somewhat surprising when one considers that, at least *in vitro*, EHEC have a capacity to bind and attach to a wide variety of receptors present on a range of cell types including red blood cells, lung, cervical, as well as intestinal (Shaw *et al*., 2002; Torres and Kaper, 2003; Holmes *et al*., 2012). It is therefore imperative to understand why EHEC are limited to the gastrointestinal tract and what factors might be acquired to allow dissemination to new niches.

A key trait of a successful pathogen is the ability to compete with the established microbial residents to facilitate an infection (De Muinck *et al*., 2013). Recent work has highlighted the importance of metabolism in this process. Different *E. coli* pathotypes often utilize discrete sets of sugars ensuring they occupy distinct sites within the gastro-intestinal tract (Meador *et al*., 2014). In addition to sugar metabolism, one interesting difference between UPEC and EHEC strains is carriage of the locus for D-serine metabolism, *dsdCXA* (Cosloy and McFall, 1973; Nørregaard-madsen and Fall, 1995). This locus codes for a D-serine deaminase (DsdA), a D-serine inner membrane transporter (DsdX) and an essential LysR-type regulator of the system (DsdC). The role of *dsdCXA* has been defined to detoxify UPEC strains from inhibitory concentrations of the host metabolite D-serine encountered primarily in the urinary tract during infection (reported concentration range of ~28 to ~1 mM in urine) (Nørregaard-madsen and Fall, 1995; Anfora *et al*., 2007). Indeed, UPEC strains can metabolize D-serine as a sole carbon source (Anfora and Welch, 2006). Conversely, studies have shown that 95% of diarrheagenic *E. coli* strains tested fail to grow on D-serine as a sole carbon source. Prototypic *E. coli* O157:H7 isolates such as EDL933 contain a truncated *dsdCXA* locus, resulting in *dsdC* and a portion of *dsdX* being substituted with the *csrRAKB* sucrose utilization locus occurring at a chromosomal recombination hotspot (Jahreis *et al*., 2002; Moritz and Welch, 2006). The acquisition of the ability to utilize sucrose provides an obvious and tangible advantage to the pathogen, but the loss of *dsdCX* is obscure. There is a clear potential for EHEC strains to encounter D-serine in the environment, for example, through dietary intake (Friedman, 1999; Csapo, 2009) or even contact with urine post-fecal shedding from the host. Also, recent studies have demonstrated the ability of certain bacteria to naturally produce non-canonical D-amino acids (D-AAs) (Lam *et al*., 2009; Cava, de Pedro, *et al*., 2011).

The regulatory roles of D-AAs in bacteria have also been explored with implicated functions in spore germination in *Bacillus* species and recently, key roles in cell wall reorganization in both *Vibrio cholerae* and *Bacillus subtilis* (Lam *et al*., 2009; Cava, de Pedro, *et al*., 2011; Cava, Lam, *et al*., 2011). D-serine has been implicated to harbor a regulatory role in UPEC virulence during urinary tract infection (Roesch *et al*., 2003; Anfora *et al*., 2007).

On the basis of the aforementioned data, involving the role of D-AAs in signaling and showing the truncation of the *dsdCXA* locus in EHEC, we hypothesized that D-serine could influence niche selectivity of different *E. coli* pathotypes. We found that D-serine alone caused repression of the Type 3 Secretion System (T3SS) and induced the SOS response (Kenyon and Walker, 1980; Michel, 2005). These phenotypes were entirely independent of one another, highlighting the vast array of regulatory roles D-AAs can play in bacteria. We also show that carriage of both *dsdCXA* and the locus of enterocyte effacement (LEE) genes encoding a T3SS is extremely rare, a genotype that we attribute to an 'evolutionary incompatibility' between the two loci. This study demonstrates the importance of evolutionary cooperation between both core and pathogenic genetic elements in defining niche specificity. Overall, we show that D-serine influences both gene content and regulation of critical virulence factors in pathogenic *E. coli.*

Materials and methods

Bacterial strains, plasmids and growth conditions

All wild-type and mutant bacterial strains used in this study are listed in Supplementary Table S2 and plasmids used are listed in Supplementary Table S3. Mutant strains TUV93-0 Δ*yhiF*, CFT073 Δ*dsdA* and CFT073 Δ*dsdX*Δ*cycA* as well as complementation constructs (p*yhiF*, p*dsdA* and p*dsdX*) were generated as described elsewhere and generously supplied as gifts for use in this study (Anfora and Welch, 2006; Tree *et al*., 2011). Single colonies of bacteria were inoculated into 5 ml LB broth containing the appropriate antibiotics where specified (ampicillin 100 μg ml^{-1}; kanamycin 50 μg ml^{-1}; chloramphenicol 25 μg ml^{-1}) and cultured overnight at 37 °C, 200 r.p.m. Overnight cultures were used to inoculate pre-warmed MEM-HEPES (Sigma-Aldrich, St Louis, MO, USA; cat # m7278) at an OD^{600} of 0.05 and samples were cultured subsequently at 37 °C, 200 r.p.m. D-AAs for screening were purchased from Sigma-Aldrich and supplemented into MEM-HEPES at a concentration of 1 mM unless otherwise stated. Mitomycin C was used as a positive control for inducing the SOS response at a concentration of 5 μM.

In vitro GFP reporter-fusion assay of LEE promoter activity

Transcriptional GFP reporter-fusions for the LEE1–3 and *tir* promoters (PLEE1:GFP, PLEE2:GFP, PLEE3:GFP and P*tir*:GFP) were used to measure the effects of different D-AAs expression on the T3SS. A P*rpsM*::GFP reporter was used to control for any variation in expression of housekeeping genes. Reporter plasmids used in this study were generated as previously described (Roe *et al*., 2003) and are listed in Supplementary Table S3. Aliquots of

bacterial culture were taken at regular intervals and transferred to a black 96-well plate for GFP fluorescence measurement on a FLUOstar Optima Fluorescence Plate Reader (BMG Labtech, Jena, Germany). GraphPad Prism software version 5.0c (GraphPad Software, San Diego, CA, USA) was used to generate a standard curve of OD^{600} versus fluorescence (relative fluorescence unit) and obtain values at OD^{600} 0.7 for comparison between samples.

SDS-PAGE and immunoblot analysis of type III secreted proteins

Analysis of T3SS proteins was carried out as described previously (Tree *et al*., 2011). Cultures of bacteria in MEM-HEPES (as above) were grown to an OD^{600} of 0.7 and supernatants were obtained by centrifugation at 4000 r.p.m. for 15 min. Cell pellets were retained for analysis of whole cell proteins and were lysed using BugBuster Protein Extraction buffer (Merck, New Jersey, USA). Supernatants were syringe-filtered (0.45 μm) and secreted proteins (Sec) were precipitated overnight using 10% v/v TCA (Sigma-Aldrich) at 4 °C. Secreted proteins were harvested by centrifugation at 4000 r.p.m. (4 °C) for 1 h. Sec pellets were resuspended in Tris-HCl (pH 8.0) and 5 μl aliquots were analysed by SDS-PAGE using the Novex system (Invitrogen, Carlsbad, CA, USA). Primary antibodies used for immunoblotting were for EspD (1/6000), Tir (1/2000), RecA (1/4000) and GroEL (1/20000).

Total RNA extraction and mRNA enrichment

Bacterial cultures were grown as above and mixed with two volumes of RNAprotect reagent (Qiagen, Valencia, CA, USA), incubating for 5 min at room temperature. Cell pellets were harvested by centrifugation and total RNA was extracted using an RNeasy kit (Qiagen). Genomic DNA was removed post extraction using a TURBO DNase kit (Ambion, Carlsbad, CA, USA) and samples were enriched for mRNA using a MICROBexpress mRNA enrichment kit (Ambion). Samples for RNA-seq analysis were QC tested for integrity and rRNA depletion using an Agilent Bioanalyzer 2100 (University of Glasgow, Polyomics Facility).

RNA-seq transcriptome generation and data analysis

cDNA synthesis and sequencing was performed at the University of Glasgow Polyomics Facility on an Illumina Genome Analyser IIx (Illumina, San Diego, CA, USA), using 70 bp single-end reads and six samples per lane. Raw reads were QC checked using FastQC (Babraham Bioinformatics, Cambridge, UK) and trimmed accordingly using CLC Genomics Workbench (CLC Bio, Aarhus, Denmark). Data were normalized and analysed for differentially expressed genes using the Bioconductor packages DEseq and EdgeR (Anders and Huber, 2010; Robinson *et al*., 2010) with raw reads mapped to the EDL933 reference genome (NCBI accession number: NC_002655.2). The sequence reads reported in this paper have been deposited in the European Nucleotide Archive under study PRJEB7974 (samples ERS627411 and ERS627412). Differentially expressed genes were identified using a significance cutoff of $\leqslant 0.05$ and changes of interest were validated using qRT-PCR. An experiment investigating the global effects of D-serine on *E. coli* O157:H7 consisted of three biological replicates of WT TUV93-0 cultured in MEM-HEPES and two biological replicates of TUV93-0 cultured in MEM-HEPES supplemented with 1 mM D-serine.

Quantitative real time-PCR (qRT-PCR)

Validation of differentially expressed genes identified *via* RNA-seq was carried out by qRT-PCR using KAPA SYBR FAST Universal qRT-PCR master mix (KAPA Biosystems, Woburn, MA, USA) and M-MLV Reverse Transcriptase (Promega, Madison, WI, USA). Total RNA was extracted as above and was quantified using a Nanodrop 2000 (Thermo Scientific, Waltham, MA, USA). Samples for comparison were normalized to a concentration of $50\,ng\,\mu l^{-1}$ using RNAse free TE buffer (Ambion). qRT-PCR analysis was performed using a two-step reaction; cDNA synthesis (37 °C for 15 min and 95 °C for 10 min) followed by qRT-PCR (40 cycles of 95 °C for 10 s, 55 °C for 30 s). Individual reactions were performed in triplicate to eliminate technical variance and reaction for each gene to be analysed was performed in biological triplicate. qRT-PCR reactions were carried out using the ECO Real-Time PCR System (Illumina) according to the manufacturers specifications and the data were analysed according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) using the housekeeping gene *gapA* as an internal control. All primers used are listed in Supplementary Table S4.

HeLa cell adhesion assay and microscopy

Coverslips were seeded with 4×10^4 HeLa cells in a multi-well plate and incubated overnight in MEM-HEPES at 37 °C with 5% CO2. TUV93-0 transformed with a pRFP plasmid were used for adhesion assays. Bacterial cultures for infection were grown in MEM-HEPES at 37 °C until at an OD^{600} of 0.6. Seeded cells were washed with fresh media and infected with 100 μl bacterial culture (adjusted to OD^{600} of 0.1) in 500 μl fresh MEM-HEPES with or without 1 mM D-serine. Plates were centrifuged at 400 r.p.m. for 3 min and incubated at 37 °C with 5% CO2 for 2 h. Wells were washed with fresh media to remove unbound bacteria and incubated as above for a further 3 h. Wells were washed three times with sterile PBS before fixing for 20 min with 250 μl PFA (2%). Wells were washed again with PBS. Wells were incubated for 5 min with 250 μl of Triton X-100 (0.5%) and washed. Host cell actin was stained with Phalloidin-488 (Invitrogen) for 1 h. Wells were washed a final time before mounting on microscope

slides with fluorescent mounting medium (Dako, Cambridge, UK). Slides were imaged using a Zeiss M1 Axioimager microscope and data were acquired and deconvoluted using the Zen Pro software (Zeiss, Jena, Germany). Adhesion assays were performed in triplicate.

Growth on D-serine as a sole carbon source
Assessment of a strain's ability to grow using D-serine as a sole carbon source was carried out using MOPS minimal media agar plates supplemented with D-serine, as described previously (Anfora and Welch, 2006). MOPS minimal media consisted of 1.32 mM K_2HPO_4, 9.52 mM NH_4Cl, 0.523 mM $MgCl_2$, 0.276 mM K_2SO_4, 10 μM $FeSO_4$, 0.5 μM $CaCl_2$, 50 mM NaCl, 40 mM MOPS, 4 mM Tricine, 3 nM $(NH_4)_6(MO_7)_{24}$, 0.4 μM H_2BO_3, 30 nM $CoCl_2$, 10 nM $CuSO_4$, 80 nM $MnCl_2$ and 10 nM $ZnSO_4$. Plates were supplemented with either 43.4 mM glycerol (as a positive control) or 4.76 mM D-serine (as a sole carbon source) as previously described elsewhere (Anfora and Welch, 2006). Strains of interest were streaked and grown at 37 °C on MOPS plates.

Extraction of whole-cell metabolites and HPLC analysis
Preparation of metabolites was adapted from Creek *et al.* (2011). Bacterial cultures were grown to an OD^{600} of 0.6 in MEM-HEPES, supplemented with and without 1 mM D-serine, and were subsequently diluted to an OD^{600} of 0.1 in fresh, warm MEM-HEPES to a final volume of 40 ml. Cultures were quenched by rapid submersion in ethanol and dry ice to cool samples to 4 °C. Cells were then centrifuged at 3500 r.p.m. (4 °C) for 10 min. Bacterial cell pellets were resuspended in 1 ml of supernatant and transferred to an ice-cold 1.5 ml microcentrifuge tube before obtaining the final cell pellet at 3500 r.p.m. (4 °C) for 5 min. The supernatant was completely removed at this stage and metabolites were extracted from the cell pellet by addition of 200 μl of chloroform/methanol/water (1:3:1) at 4 °C and tubes were vortexed vigorously for 1 h at 4 °C. Cellular debris and precipitate was removed by a final centrifugation at 13 000 r.p.m. (4 °C) for 5 min. Supernatant was transferred to a fresh tube and stored at − 80 °C until LC-MS analysis.

Samples were separated on a Chirobiotic T2 column (2 mm × 25 cm; Sigma-Aldrich) using an isocratic flow at 85% ethanol/15% water at 200 μl min^{-1}. Amino acids were detected using atmospheric pressure chemical ionization in an Orbitrap Elite (Thermo Scientific) at 70 000 resolution. Source settings were configured as described previously (Desai and Armstrong, 2004). Analysis was performed using the ExactFinder software (Thermo Scientific).

Bioinformatic analysis of dsdCXA and the LEE carriage in E. coli
The nucleotide sequences for 159 core genes in *E. coli* were elaborated as described in a recent publication (R. Goldstone, in preparation). Briefly, at each iteration, the core gene set, initialized as the nucleotide sequence for genes present in MG1655, was aligned to the next *E. coli* genome sequence using blastn (Altschul *et al.*, 1990). Genes aligning at >70% identity and >80% of the length of the coding sequence were retained in the core gene set for use in the next iteration. This analysis resulted in the identification of 159 genes. The nucleotide sequences of each core gene were extracted from the *E. coli* genomes, aligned by Muscle (Edgar, 2004), concatenated, and a maximum likelihood tree calculated using PhyML (Guindon and Gascuel, 2003) under the general time reversible model of nucleotide substitution. Dendrograms were visualised using the APE package (Paradis *et al.*, 2004) implemented in R (R Development Core Team, 2012).

The amino acid sequences for genes present within O-island 148, encoding the LEE T3SS system were collected from the sequence for EDL933 (NCBI accession number: NC_002655.2), and amino acid sequences for DsdX, DsdC and DsdA were collected from the sequence for CFT073 (NCBI accession number: NC_004431.1). These sequences were iteratively aligned against the *E. coli* genomes using tblastn and called as present in a genome if the protein sequence aligned at greater than 70% identity over greater than 80% to the translated nucleotide sequence of the genome in question. The distribution of the genes across the *E. coli* core-genome dendrogram was visualized using the Diversitree package (Fitzjohn, 2012) implemented in R (R Development Core Team, 2012).

For statistical analysis to take account of genes incorrectly labelled as 'absent' because of the splitting of coding sequences across contigs, the LEE was called as 'present' if an isolate possessed 21 or more of the following: *espF, Z5102, escF, Z5014, espB, espD, espA, sepL, escD, eae, Z5111, tir, Z5113, Z5114, Z5115, sepQ, Z5117, Z5118, escN, escN, Z5121, sepZ, Z5123, escJ, Z5125, escC, cesD, Z5128, Z5129, Z5131, escU, escT, escS, escR, Z5136, Z5137, Z5138, Z5139, Z5140, Z5142* and *Z5143*. *dsdCXA* was called as present only if all three of *dsdX, dsdC and dsdA* were identified. Statistical significance for the association between *dsdCXA* and the LEE was evaluated by Fisher's exact test.

Results

Global analysis of the effects of D-serine on EHEC
Previous work has demonstrated D-AAs can play an important role in the regulation of cellular processes and that D-serine is particularly important for UPEC pathogenesis (Anfora *et al.*, 2007; Cava, Lam, *et al.*, 2011). Our hypothesis was that D-serine might also play an important role in regulating gene expression in *E. coli* O157:H7. To investigate this, we used RNA-sequence analysis (RNA-seq), which provided us with a global view into the effects of D-serine on

gene expression (Wang *et al.*, 2009). A concentration of 1 mM D-serine was used as this represents a physiologically relevant level as previously reported (Anfora *et al.*, 2007). We also found concentrations upwards of 2 mM D-serine had no significant effects on growth or viability in the media tested (data not shown). Significantly ($P \leqslant 0.05$) up- and downregulated genes were identified using the DESeq and EdgeR packages of Bioconductor (Anders and Huber, 2010; Robinson *et al.*, 2010). The most strongly upregulated genes included members of the SOS-regulon (Fernández De Henestrosa *et al.*, 2000; Salgado *et al.*, 2013). This comprised 21 genes including the well-characterized SOS components *lexA*, *recA* and *sulA* (Figure 1a; Supplementary Table S1). These genes are normally associated with a global response to DNA-damage (Kenyon and Walker, 1980; Michel, 2005) and the addition of a single D-AA seemed inconsistent with such detrimental effects. These results were confirmed by immunoblotting for the SOS anti-repressor RecA, which was markedly upregulated compared with the untreated control (Figure 1b) and RNA-seq data were also validated by qRT-PCR (Supplementary Figure 1).

Addition of D-serine significantly reduced the expression of 26 out of the 42 genes that comprise the LEE, a horizontally acquired pathogeni-

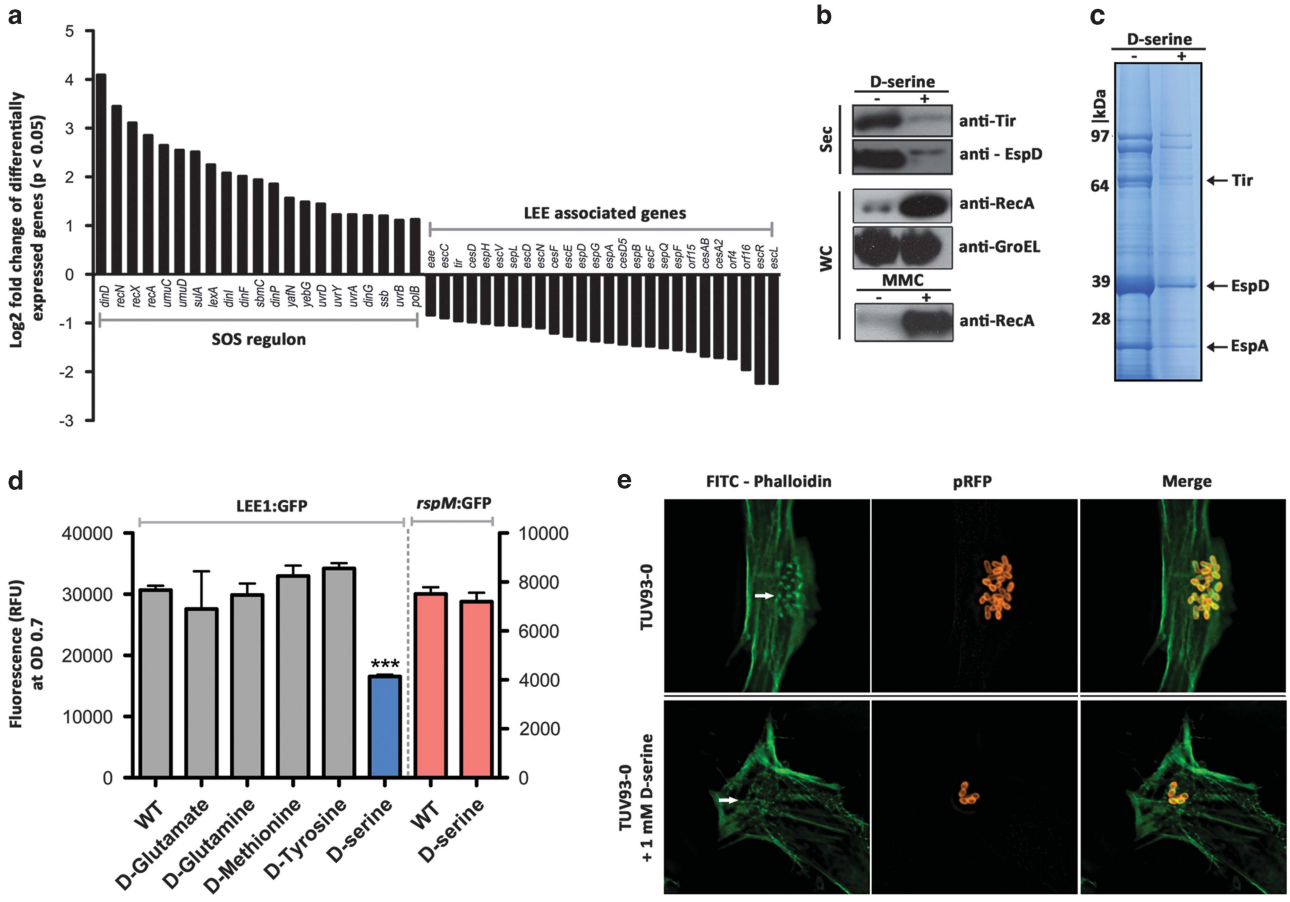

Figure 1 D-serine induces the SOS response and represses the T3SS in *E. coli* O157:H7. (**a**) Co-ordinated induction of the SOS response and repression of the LEE in TUV93-0 cultured in MEM-HEPES supplemented with D-serine as identified by RNA-seq analysis. Histogram bars indicate the Log^2 fold change of differentially expressed genes wit a *P*-value of $\leqslant 0.05$. (**b**) Immunoblot analysis of secreted proteins (Sec) and whole-cell lysate (WC) from TUV93-0 cultured in MEM-HEPES alone (−) and supplemented with 1 mM D-serine (+). Tir, EspD and RecA were used as markers for type 3 secretion and the SOS response. Equal amount of Sec fractions were analysed in − and + lanes. GroEL was used as a loading control for the WC fractions and these corresponded to the Sec fractions. Cultures supplemented with 5 μM mitomycin C to induce the SOS response were used as a positive control. (**c**) SDS-PAGE analysis of the type 3 secretion profile from TUV93-0 cultured in MEM-HEPES alone (−) and supplemented with 1 mM D-serine (+). Protein bands corresponding to Tir, EspD and EspA were identified by MS-MS analysis and are indicated with black arrows. (**d**) Screening the effects of multiple D-AAs on LEE1 expression using a GFP tagged LEE1 promoter fusion reporter (PLEE1:GFP). A subset of all D-AAs tested (D-glutamate, D-glutamine, D-methionine, D-tyrosine and D-serine) is shown with D-serine alone showing a significant decrease in LEE1 expression (blue bar). *** denotes $P \leqslant 0.001$ calculated from three biological replicates. An *rpsM* GFP reporter was used as a housekeeping control (P*rpsM*:GFP) and is indicated by red bars on the right *y* axis. Relative fluorescence units (RFU) were derived from a standard curve of optical density at 600 nm (OD^{600}) measured over time. Bacteria were cultured in MEM-HEPES to promote expression of the T3SS. (**e**) Wide-field fluorescence microcopy images of HeLa cells infected with TUV93-0 (with and without the addition of 1 mM D-serine to the growth medium). Host HeLa cells were stained with FITC-Phalloidin (green) and bacteria with carrying a constitutively expressed RFP construct (red). Actin condensation, corresponding to pedestal formation, is indicated by a white arrow.

city island which encodes a major virulence factor, the T3SS, across five operons, LEE1 to LEE5 (Supplementary Table S1) (McDaniel *et al.*, 1995; Croxen and Finlay, 2010; Büttner, 2012). The LEE-encoded genes downregulated included those encoding for membrane structural components, the needle proteins and secreted effectors (Figure 1a). The normal function of the LEE is to translocate bacterial effector proteins into host cells and subvert normal cell function (Büttner, 2012). Secretion of several of these effector proteins can be readily assayed by growing EHEC in media that induces secretion in the absence of host cells (Tree *et al.*, 2011). Separation by SDS-PAGE and visualization with Coomassie blue allows several well-characterized proteins to be detected including Tir, EspA and EspD. Addition of 1 mM D-serine caused a marked reduction in the production of effector proteins (Figure 1c), a phenotype that was consistent with downregulation of the LEE. The results were confirmed by immunoblotting for both Tir and EspD, which showed a greater than sevenfold reduction in effector protein secretion after addition of D-serine (Figure 1b). As a control, the corresponding whole-cell fractions were probed for GroEL, which showed no differences (Figure 1b).

Downregulation of the T3SS is D-serine-specific and affects host cell binding

To test whether the phenotype was specific to D-serine, we analysed the effects of numerous D-AAs on expression of the LEE. Addition of 1 mM of D-glutamate, D-glutamine, D-methionine, D-tyrosine, D-valine, D-tryptophan, D-phenylalanine and D-leucine had no effect on LEE expression, as determined by measuring transcription of the LEE master regulator Ler (Figure 1d) (Elliott *et al.*, 2000). In contrast, addition of 1 mM D serine reduced *ler* expression by 54% ($P \leqslant 0.0001$). As a control, we tested transcription from the promoter for the ribosomal protein, RpsM, which was unaffected (Figure 1d). Moreover, L-serine had no effect, highlighting the specificity of the phenotype solely to D-serine. Finally, we found D-serine to be significantly detrimental to LEE expression at concentrations of $\geqslant 100\,\mu M$ in MEM-HEPES (Supplementary Figure 2).

Given that D-serine selectively downregulated the T3SS, we then tested the ability of EHEC to bind and intimately attach to host cells. WT bacteria use their T3SS to translocate effector proteins into the host cell resulting in distinctive areas of condensed host-cell actin, called pedestals (Figure 1e) (Croxen and Finlay, 2010). Addition of 1 mM D-serine during infection does not affect bacterial growth rate in the media used but resulted in 77% fewer infected host cells relative to the untreated control ($P < 0.001$). Moreover, D-serine also reduced the bacterial numbers on successfully infected host cells by 62% ($P < 0.01$) and fewer of the attached bacteria formed pedestals (Figure 1e and Supplementary Figure 3A). As a control, an isogenic TUV93-0 Δ*tir* strain was used. As expected, this failed to translocate Tir and form pedestals (Supplementary Figure 3B). The reduction in LEE expression at D-serine concentrations of $\geqslant 100\,\mu M$ implied that the levels in the gastro-intestinal tract would be significantly lower to permit LEE functionality. Measurement of D-serine in the colon of five BALB/C mice revealed levels of $\sim 1\,\mu M$ relative to a D-serine standard (data not shown) some 100 times lower than the concentration required for inhibition of LEE expression.

T3SS repression by D-serine is independent of the SOS response

Previous work has demonstrated that the majority of EHEC cannot grow on D-serine as a sole carbon source, the assumption being that these strains do not express a deaminase capable of metabolizing this amino acid (Moritz and Welch, 2006). Correspondingly, the EHEC strain used in this study (TUV93-0) failed to grow on MOPS minimal agar plates containing D-serine as a sole carbon source (Figure 2a). We postulated that the accumulation of intracellular D-serine resulted in expression of the SOS response. This hypothesis was tested in two ways. Firstly, a plasmid-borne copy of D-serine deaminase (p*dsdA*) was introduced into EHEC and secondly, the UPEC strain CFT073 was deleted for *dsdA*. In both cases, the SOS response and the level of intracellular D-serine were measured. Transformation of EHEC with p*dsdA*, enabled growth on D-serine plates (Figure 2a) showing clearly that the strain was now capable of actively metabolizing D-serine. Intracellular D-serine levels of EHEC p*dsdA* were reduced > sevenfold compared with the WT, consistent with D-serine breakdown (Supplementary Figure 4A and B). Interestingly, the expression of RecA levels in EHEC p*dsdA* showed that they were restored to the same levels as seen in the absence of D-serine (Figure 2b). Furthermore, intracellular accumulation of D-serine was evident in a UPEC Δ*dsdA* mutant and this resulted in strong RecA expression, indicating activation of the SOS response (Figure 2c). In contrast, analysis of the secreted protein profile and immunoblotting for EspD showed that even when EHEC can metabolize D-serine, the T3SS remains inhibited and production of effector proteins (Figure 2b) and the ability to attach to host cells (Supplementary Figure 3C) is reduced. These data show that the SOS response is dependent on intracellular accumulation of D-serine and that repression of the T3SS by D-serine occurs even in the absence of the SOS response.

D-serine regulates the T3SS through yhiF and ihfA

Examination of the RNA-seq data revealed differential expression of known regulators that might explain the repression of the LEE. We identified that D-serine differentially affected the expression of two

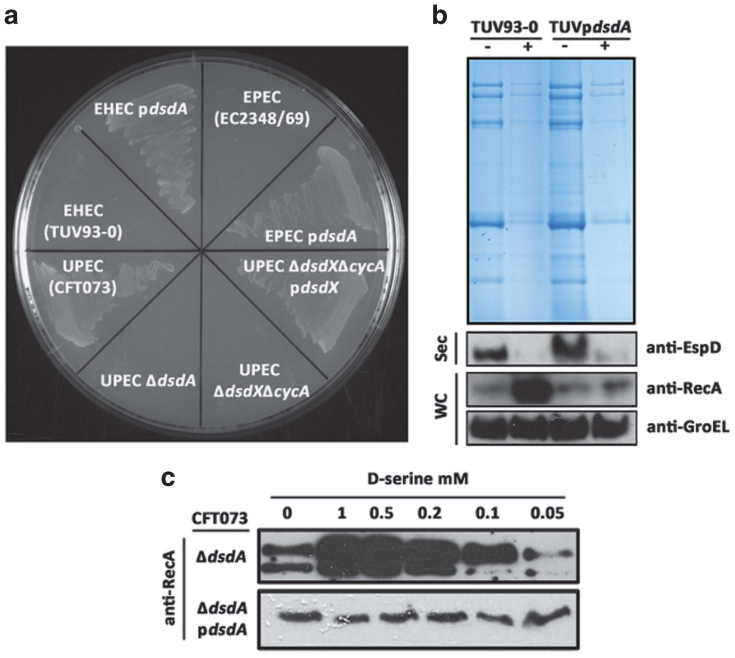

Figure 2 The effects of D-serine accumulation on growth, the SOS response and the T3SS. (**a**) Growth of various *E. coli* strains on MOPS minimal agar plates containing D-serine as a sole carbon source. WT UPEC (CFT073) or EHEC (TUV93-0) and EPEC (EC234/69) complemented with p*dsdA* were capable of growing on D-serine as a sole carbon source. WT EHEC and EPEC as well as a UPEC Δ*dsdA* mutant were unable to grow. A UPEC Δ*dsdX/cycA* D-serine transporter double mutant was also unable to grow. (**b**) SDS-PAGE and immunoblot analysis of secreted proteins from TUV93-0 and TUV93-0 complemented with *dsdA* from UPEC CFT073 in trans (TUV*pdsdA*). Strains were cultured in MEM-HEPES alone (−) and supplemented with 1 mM D-serine (+). Immunoblot analysis of EspD levels from the secreted fraction (Sec) and RecA and GroEL from the whole-cell lysate (WC) are displayed. Equal amounts of Sec fraction were loaded for each lane and WC fractions from the corresponding Sec fractions were used to indicate identical total protein levels between samples. GroEL was used as a loading control. (**c**) Immunoblot analysis of RecA expression in UPEC Δ*dsdA* over five increasing D-serine concentrations. Complementation with p*dsdA* resulted in no increased RecA expression.

DNA-binding transcriptional activators, IhfA (Integration Host Factor alpha-subunit) and the GAD acid stress response regulator YhiF. Previous work has shown Integration Host Factor (IHF) directly binds the LEE1 promoter as a positive regulator of the T3SS. YhiF is a member of the LuxR family of transcriptional regulators and has been shown to negatively regulate the LEE2 and LEE5 promoters (Friedberg *et al*., 1999; Tatsuno *et al*., 2003; Tree *et al*., 2011). The upregulation of *yhiF* and downregulation of *ihfA* were verified using qRT-PCR analyses (Figure 3a). As both IhfA and YhiF have been linked to regulation of the LEE, we evaluated their contribution to the D-serine phenotype. Given that *yhiF* was upregulated in the presence of D-serine, we hypothesized that a Δ*yhiF* deletion strain might be less sensitive to this amino acid. Similarly, as *ihfA* was downregulated in response to D-serine, the prediction is that constitutive complementation of this activator might similarly reduce the repressive effects. As demonstrated in Figure 3b, deletion of *yhiF* provided protection from the repressive effects of D-serine. This phenotype could be complemented by transformation of the mutant with a plasmid-borne copy of *yhiF*. Similarly, EHEC transformed with a plasmid constitutively expressing *ihfA* (p*ihfA*) were shown to be insensitive to D-serine addition. Furthermore, to investigate whether YhiF and IhfA were cross-regulated in response to D-serine, qRT-PCR analysis of each transcript in the corresponding deletion or complementation background was carried out (Figure 3a). These results revealed that differential expression of these regulators in response to D-serine was dependent on each other. In the Δ*yhiF* strain, *ihfA* was no longer downregulated in response to D-serine and similarly in the p*ihfA* strain, *yhiF* levels were comparable with that of the wild type (Figure 3a). These data provide convincing evidence that *yhiF* and *ihfA* co-operate to regulate LEE expression following addition of D-serine. A simple model summarizing these data is presented in Figure 3c.

Carriage of dsdCXA and the LEE is rare

Our data showed that D-serine had potent effects on the expression and function of the T3SS. This sensitivity to D-serine suggested that strains carrying the LEE would be limited to colonization of environments with low concentrations of this amino acid, hence there would be no requirement to metabolize D-serine and low carriage of the full *dsdCXA* locus. We assessed the frequency of carriage of the LEE and the *dsdCXA* locus in 1591

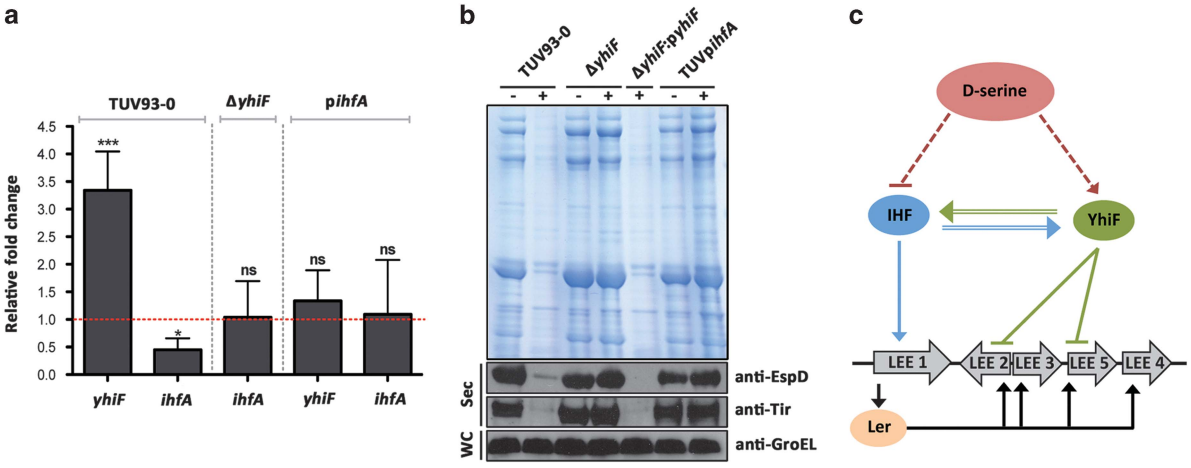

Figure 3 The mechanism of D-serine repression of the T3SS in *E. coli* O157:H7. (**a**) qRT-PCR validation of differentially expressed LEE regulators (*yhiF* and *ihfA*) in response to D-serine identified by RNA-seq. The expression of *yhiF* and *ihfA* in response to D-serine was investigated by analysing relative mRNA transcript levels of each regulator in the corresponding complementation or knockout background. The wild-type TUV93-0, Δ*yhiF* and p*ihfA* backgrounds are indicated above the bars and are separated by gray dashed lines. The red dashed line indicates relative baseline expression, with genes expressed above this being upregulated and genes expressed below this being downregulated. ns, * and *** denote no significance, $P \leqslant 0.05$ and $P \leqslant 0.001$, respectively, calculated from three biological replicates. (**b**) SDS-PAGE and immunoblot analysis of secreted proteins (Sec) from TUV93-0, Δ*yhiF*, Δ*yhiF* complemented with *yhiF* in trans (Δ*yhiF*:p*yhiF*) and TUV93-0 complemented with *ihfA* in trans (TUV:p*ihfA*). Strains were cultured in MEM-HEPES alone (−) and supplemented with 1 mM D-serine (+). Equal amounts of the Sec fraction were loaded for each lane and whole-cell lysate (WC) fractions from the corresponding Sec fractions were used to indicate identical total protein levels between samples. GroEL was used as the loading control. (**c**) Regulatory model of D-serine repression on the LEE pathogenicity island *via* YhiF and IhfA. Arrows indicate positive regulation, whereas blunt end lines indicate negative regulation. Solid lines indicate a direct signal, double lines indicate cross regulation and dashed lines indicate an unknown signal. Expression of LEE1 to LEE5 operons is driven through Ler. IHF directly upregulates the LEE1 promoter, whereas YhiF represses the LEE2 and LEE5 promoters. Wild-type TUV93-0 supplemented with D-serine leads to downregulation of *ihfA* and upregulation of *yhiF*. *ihfA* and *yhiF* respond co-operatively in response to D-serine and differential expression of both regulators is required for D-serine-mediated repression of the LEE.

strains of sequenced *E. coli* isolates available in the NCBI database (as available on 3/6/2014) (Figure 4a). Among strains carrying either the LEE or the *dsdCXA* locus, there was a strong correspondence to absence of the reciprocal locus (odds ratios of 16.49 and 21.88, respectively). In contrast, among LEE negative strains, there was no correspondence with presence or absence of the *dsd* locus (odds ratio 0.99) with a corollary in strains lacking *dsdCX* also being LEE negative (odds ratio 1.37) (Figure 4b). These data demonstrate that for LEE negative strains, there was an equal probability of the *dsd* locus being carried or absent (Figure 4b). In contrast, for strains carrying either the LEE or the *dsdCXA* locus, there was a significant correspondence to absence of the reciprocal locus (odds ratios of 16.49 and 21.88, respectively; $P<0.0001$). These data demonstrate a clear distinction in the carriage of these two loci, which is consistent with them being functionally incompatible.

Analysis of *dsdA* shows that it is highly conserved across the *E. coli* lineage; indeed, over 98% of strains (1561 of 1591) investigated in this study carried *dsdA*. Mapping the nucleotide diversity of *dsdA* results in a phylogenetic tree very similar to one constructed using sequences from the core genome (Supplementary Figure 5). This suggests that *dsdCXA* are ancestral and that the ability to metabolize D-serine was a capacity of the progenitor *E. coli*. Moreover, it is also important to address whether loss of *dsdCX* preceded acquisition of the LEE. Our analysis shows that 1.6% of strains analysed carried both *dsdCXA* and the LEE indicating that the LEE can be acquired in a *dsdCXA* background but that subsequent selective pressures typically result in loss of *dsdCX* (Figure 4b).

To further test our hypothesis, we looked more closely within phylogroup B2 (Figure 4c), represented by 427 strains in our analysis. This phylogroup contains primarily extraintestinal pathogenic *E. coli* strains including the well-characterized UPEC strains 536, UTI89 and CFT073 as well as strain S88, a neonatal meningitis associated isolate. Notably, B2 also contains a number of intestinal pathogenic *E. coli* strains that provide an informative comparator. The vast majority of B2 strains (344/427 = 81%) carried *dsdCXA*. Strikingly, the enteropathogenic *E. coli* (EPEC) strains within this phylogroup, representing the majority of the intestinal pathogenic *E. coli* strains, all lack *dsdCX,* but carry the LEE (Figure 4c). This is interesting as these EPEC strains are more closely related to the extraintestinal pathogenic *E. coli* isolates phylogenetically, but have undergone two clear genetic changes that correspond with adaptation to a strictly gastrointestinal tract niche. This highlights the powerful selective pressure to lose *dsdCX* when the LEE is acquired.

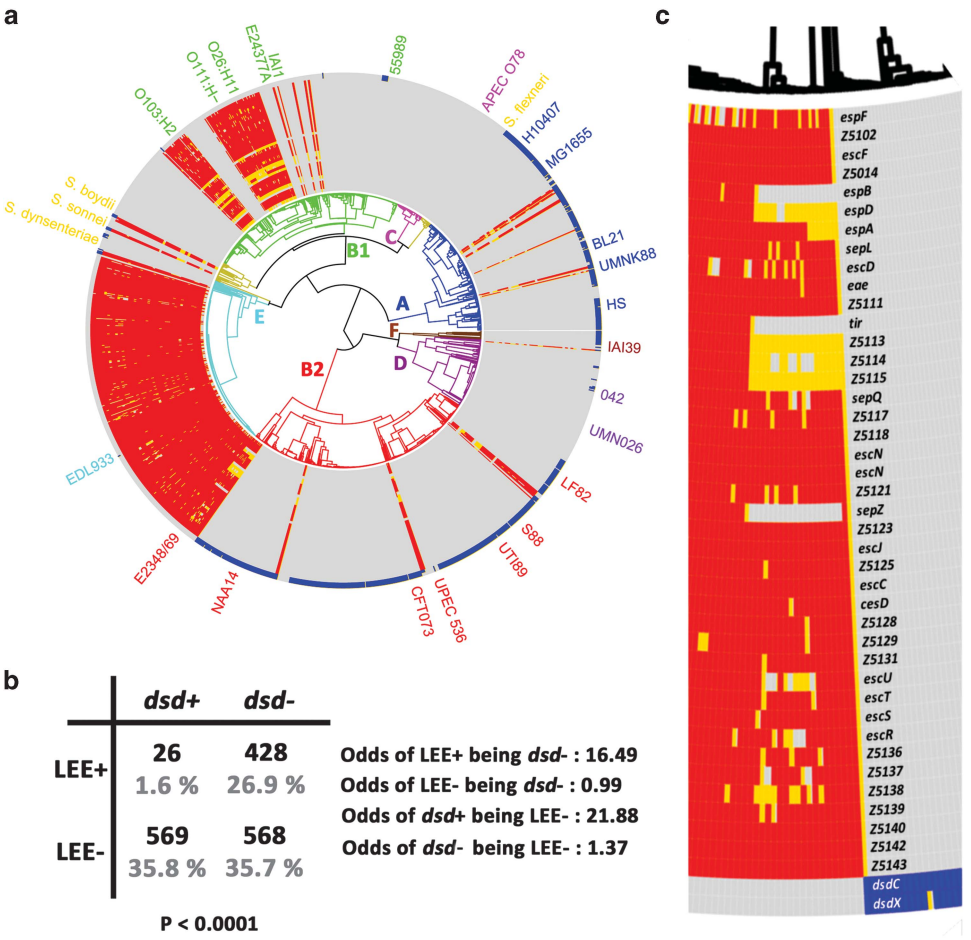

Figure 4 Carriage of both the LEE and *dsdCXA* is rare amongst *E. coli* strains. (**a**) Circularized phylogenetic tree of 1591 *E. coli* and *Shigella* isolates overlaid with gene carriage for the LEE island (>80% identity over >80% of the coding sequence highlighted in red; >50% identity over >50% of the coding sequence highlighted in yellow) and *dsdCX* locus (>80% identity over >80% of the coding sequence highlighted in blue; >50% identity over >50% of the coding sequence highlighted in yellow). Phylogenetic sub-grouping is indicated by branch colour coding as follows: Phylogroup A = Blue; Phylogroup B1 = Green; Phylogroup B2 = Red; Phylogroup C = Magenta; Phylogroup D = Purple; E = Cyan; Phylogroup F = Brown; *Shigella* = Gold. Key strains are labelled according to their location on the tree. (**b**) 2 × 2 contingency matrix for carriage of the LEE and *dsdCXA* loci. The table is based on the 1591 strains investigated for carryiage of the LEE, *dsdCXA,* both loci or neither loci (a '+' denotes presence; a '−' denotes absence). A strain is assumed *dsd* positive (+) if it carries coding sequence for *dsdC, dsdX* and *dsdA*. A strain is assumed *dsd* negative (−) if it carries the *dsdCX* truncation. The number of strains for each scenario and their percentage distribution amongst the 1591 strains investigated are given in black and grey, respectively. Statistical significance was calculated using a Fisher's exact test ($P<0.0001$). Odds ratio (OR) of LEE versus *dsdCXA* carriage are indicated also. (**c**) Expansion of LEE+ EPEC isolates from phylogroup B2 exemplifies the clear distinction between LEE and *dsdCXA* carriage.

Discussion

Understanding the basis by which a pathogen targets a particular niche is critically important when designing potential intervention strategies such as vaccines, probiotics and anti-virulence agents. Furthermore, even when we have a good understanding of the molecular basis to pathotype tropism, predicting how evolutionary pressures will shape emergent strains is challenging. For example, the 2011 outbreak of *E. coli* O104:H4 was caused by an outbreak strain carrying virulence genes for Shiga toxin production and aggregative adherence to intestinal epithelial cells, a previously rare combination (Bielaszewska *et al.*, 2011; Karch *et al.*, 2012). On this basis, we have considered why *E. coli* O157:H7 is generally limited to the gastrointestinal tract, and why it is rarely associated with the colonization of other niches. Comparison with UPEC isolates is particularly informative as these strains compete well both in the gastro-intestinal tract and are capable of extraintestinal disease (Chen *et al.*, 2013).

Previous work has demonstrated that the *dsdCXA* locus is carried extensively by UPEC isolates. Indeed, the ability to metabolize D-serine is entirely compatible with strains that infect the bladder as urine contains ~28 to 1 mM of D-serine (Anfora *et al.*, 2007). Thus, the catabolism of D-serine represents a positive *E. coli* fitness trait during

urinary tract infection. Our data have shown that the concentration of D-serine in the gastrointestinal tract of mice to be far lower than this, at ~1 μM. On the basis of this, our working model is that LEE-positive strains are largely restricted to the gastrointestinal tract as this environment has insufficient D-serine concentrations required to block expression of the LEE. In contrast, distal sites such as the bladder contain concentrations of D-serine that would block LEE mediated adhesion.

D-isomers of AAs provide not only potential carbon sources but have also been shown to be important regulatory signaling molecules (Cava, Lam, *et al.*, 2011). On the basis of this, we postulated that D-AAs might be important in the regulation of virulence genes in *E. coli* O157:H7. Testing a variety of D-AAs showed that D-serine selectively affected LEE gene expression. Whereas D-serine can be bacteriostatic in minimal medium at ~0.475 mM (Cosloy and McFall, 1973; Nørregaard-madsen and Fall, 1995), addition of up to 3 mM D-serine did not affect the growth rate of *E. coli* O157:H7 grown in minimal essential media (MEM-HEPES) (data not shown). This media is typically used for culture of O157:H7 as it normally induces expression of the T3SS (Tree *et al.*, 2011).

Remarkably, addition of D-serine reduced T3SS expression by modulating the expression of the majority of the LEE genes. The resulting phenotype was clear: strong reduction in the secretion of T3SS effector proteins and reduced attachment to host cells. Although expression of the LEE was reduced, D-serine also caused a selective upshift in gene expression. Strikingly, exposure to D-serine resulted in an upregulation of 21 genes of the SOS-regulon. This 'stress' response was confirmed by immunoblotting for RecA and qRT-PCR.

The induction of the SOS response was interesting as it is normally associated with a global response to DNA damage (Kenyon and Walker, 1980; Michel, 2005). However, induction of an SOS response can be achieved by alternative mechanisms, for example, by addition of β-lactams that inhibit the *ftsI* gene product, penicillin binding protein 3 (Maiques *et al.*, 2006; Miller *et al.*, 2013). Previous work has shown that *E. coli* can covalently link certain D-AAs to its peptidoglycan and that in the case of *V. cholerae*, cause a rod-to-sphere transition (Lam *et al.*, 2009). The exact mechanism by which D-serine induces the SOS response will be pursued as a separate study, but it is exciting to speculate that this may be linked to changes in peptidoglycan cross-linking as we also observed more spherical cells upon addition of D-serine to the growth media (unpublished data).

As D-serine supplemented in the growth environment resulted in induction of the SOS response and repression of the LEE, we tested whether these responses were co-dependent. Transformation of O157:H7 with a plasmid (p*dsdA*) expressing D-serine deaminase from UPEC, allowed the strain to metabolize D-serine and restored RecA expression to the same level as seen in the absence of D-serine (Figure 2b). In contrast, analysis of the secreted protein profile and immunoblotting for EspD showed that even when O157:H7 can metabolize D-serine, the T3SS is inhibited and expression of effector proteins is reduced (Figure 2b). These data show that D-serine causes repression of the T3SS even in the absence of the SOS response. Indeed, inspection of our transcriptomic analysis revealed that two known regulators of the LEE were affected by addition of D serine: YhiF and IhfA (Friedberg *et al.*, 1999; Tatsuno *et al.*, 2003; Tree *et al.*, 2011). IhfA is the alpha-subunit of the IHF and is a member of the nucleoid-associated proteins, which are involved in a number of cellular processes including transcriptional regulation (Goosen and van de Putte, 1995). The involvement of IHF in multiple regulons has been explored recently in *E. coli* K-12, revealing over 30% of all predicted operons containing IHF-binding regions and ~10% of these showing differential expression in an *ihfAB* double mutant (Prieto *et al.*, 2012). Specifically, IHF has also been shown to play a critical and direct role in positive regulation of the LEE1 promoter, *via* the Ler master regulator which subsequently activates operons LEE2 to LEE5 (Friedberg *et al.*, 1999). The second regulator identified, YhiF, is a member of the GAD acid stress response regulators which have been identified as negative regulators of the LEE2 and LEE5 promoters (Tatsuno *et al.*, 2003). In line with this knowledge, we identified that both regulators are differentially expressed to promote repression of the LEE. Interestingly, IhfA and YhiF were found to be co-operative in response to D-serine suggesting an interplay that leads to repression of the LEE *via* these two regulators (Figure 3a). Constitutive expression of *ihfA* removed the repression of the LEE and restored *yhiF* levels to that of the wild type suggesting that YhiF may repress *ihfA* at the transcriptional level. Similarly, deletion of *yhiF* restored wild-type *ihfA* expression levels and removed LEE repression activated by D-serine. We propose that IHF and YhiF form a regulatory feedback loop in which repression of *ihfA* down-regulates LEE1 expression and upregulation of *yhiF* represses the LEE further at the LEE2 and LEE5 operons (Figure 3c). Regulators of the GAD acid stress response have previously been shown to be overridden by alternate regulatory systems to modulate the LEE (Tree *et al.*, 2011; Branchu *et al.*, 2014), so it is conceivable to imagine a D-serine-specific response system that is mediated through key regulators of the LEE. However, despite this evidence that IhfA and YhiF regulate the LEE in response to D-serine, whether this amino acid acts directly on these regulators remains to be tested.

Exposure of *E. coli* O157:H7 to D-serine caused a distinct pattern of gene expression. We postulated that the importance of this might affect carriage of

the *dsdCXA* locus for two rather distinct reasons. Firstly, a functional *dsdA* gene did not overcome the repression of the LEE—such that the T3SS is not functional when D-serine is present in the environment. This suggested that carriage of *dsdCXA* would offer little advantage for pathogens with the LEE. Secondly, a functional *dsdA* gene prevented the SOS response thereby fundamentally changing the gene expression profile of the pathogen. To test this, we assessed the frequency of carriage of LEE and *dsdCXA* genes in 1591 strains of sequenced *E. coli* isolates and showed a strong distinction in the carriage of the LEE and *dsdCXA*, which is consistent with them being functionally incompatible. In place of *dsdCX,* most strains contain the locus for sucrose utilization (*cscBKAR*), an observation that has been made for some *E. coli* K12 strains in the 1970s (Alaeddinoglu and Charles, 1979). Insertion of the sucrose operon is facilitated by the presence of the *argW* gene, which codes for an arginine-specific tRNA and acts as a hotspot for recombination (Jahreis *et al.*, 2002). Clearly, the ability to metabolize sucrose provides a tangible selective advantage to strains carrying this operon but, in addition, the concurrent loss of *dsdCX* affects the gene expression response of *E. coli* O157:H7 to D-serine. Some rare isolates (26 from 1591 genomes analysed) appear to carry both the LEE and *dsdCXA*. These strains were all isolated from stool samples suggesting they would be classified as intestinal pathogenic *E. coli*. Our interpretation is that these strains have acquired the LEE without opportunity for subsequent selective pressures to result in replacement of *dsdCXA*.

As D-serine affects expression of the LEE, it seems likely that highlights that this amino acid provides an important signal as to where and when to attach to host tissue. Sites commonly infected by extra-intestinal pathogenic *E. coli* strains such as the bladder and meninges both contain high levels of D-serine (Wolosker *et al.*, 2008) and we have shown that the LEE is rarely carried in these pathotypes as it does not function when in physiologically relevant concentrations of this amino acid. It should be noted that some rare, but serious, cases of urinary tract infection have resulted from Shiga toxin positive strains, although genetic information on their gene context is very sparse (Tarr *et al.*, 1996). Interestingly, the archetype *E. coli* K1 neonatal meningitis strain, strain RS218, has two copies of the *dsdCXA* genes (Moritz and Welch, 2006), facilitating infections in this tissue. Fascinatingly, one NMEC strain, *E. coli* O7:K1 CE10, has been reported to carry a functional T3SS that is involved in the invasion and intracellular survival in human brain microvascular endothelial cells (Yao *et al.*, 2009). However, this strain does not carry the LEE but instead appears to have a functional 'second type' or ETT2 secretion system that delivers effector proteins into host cells. On the basis of our data, we would expect that the CE10 ETT2 system is functional in D-serine and regulated very differently compared with the LEE. In this way, the NMEC strain has adapted to tolerate and metabolize D-serine whilst still retaining a functional delivery system for effector proteins. Overall, our study provides novel insights into how a single amino acid affects gene regulation in *E. coli* O157:H7 and that this can have stark implications for the niche specificity of this pathotype.

Conflict of Interest

The authors have confirmed that they have no financial, personal or professional interests that could be construed to have influenced the paper.

Acknowledgements

We are very grateful to Professor Jose Penades for fruitful discussions and critical appraisal of the manuscript, Professor David Gally for generously supplying the Δ*yhiF* strain and the complementation plasmid, and Professor Rodney Welch for his generous gift of the CFT073 strain, corresponding mutants and complementation plasmids. JPRC was supported by an MRC studentship (DTP MR/J50032X/1). RJC and DGES acknowledge funding from the BBSRC ref (BB/1017283/1).

References

Alaeddinoglu NG, Charles HP. (1979). Transfer of a gene for sucrose utilization into Escherichia coli K12, and consequent failure of expression of genes for D-serine utilization. *J Gen Microbiol* **110**: 47–59.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Anders S, Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Anfora AT, Haugen BJ, Roesch P, Redford P, Welch Ra. (2007). Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by Escherichia coli CFT073. *Infect Immun* **75**: 5298–5304.

Anfora AT, Welch Ra. (2006). DsdX is the second D-serine transporter in uropathogenic Escherichia coli clinical isolate CFT073. *J Bacteriol* **188**: 6622–6628.

Bielaszewska M, Mellmann A, Zhang W, Köck R, Fruth A, Bauwens A *et al.* (2011). Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* **11**: 671–676.

Branchu P, Matrat S, Vareille M, Garrivier A, Durand A, Crépin S *et al.* (2014). NsrR, GadE, and GadX interplay in repressing expression of the Escherichia coli O157:H7 LEE pathogenicity island in response to nitric oxide Mulvey, MA (ed). *PLoS Pathog* **10**: e1003874.

Brzuszkiewicz E, Bru H, Liesegang H, Emmerth M, Tobias O, Albermann K *et al.* (2006). How to become a uropathogen: Comparative genomic analysis of

extraintestinal pathogenic Escherichia coli strains. *Proc Natl Acad Sci USA*. **103**: 12879–12884.
Büttner D. (2012). Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol Mol Biol Rev* **76**: 262–310.
Cava F, de Pedro Ma, Lam H, Davis BM, Waldor MK. (2011). Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *EMBO J* **30**: 3442–3453.
Cava F, Lam H, de Pedro Ma, Waldor MK. (2011). Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cell Mol Life Sci* **68**: 817–831.
Chen SL, Wu M, Henderson JP, Hooton TM. (2013). Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Sci Transl Med* **5**: 184ra60.
Cosloy SD, McFall E. (1973). Metabolism of D-serine in Escherichia coli K-12: mechanism of growth inhibition. *J Bacteriol* **114**: 685–694.
Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KEV. (2011). Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: Improved metabolite identification by retention time prediction. *Anal Chem* **83**: 8703–8710.
Croxen Ma, Finlay BB. (2010). Molecular mechanisms of Escherichia coli pathogenicity. *Nat Rev Microbiol* **8**: 26–38.
Csapo J. (2009). The D-amino acid content of foodstuffs (a review). *Acta Univ Sapientiae, Aliment* **1**: 5–30.
De Muinck EJ, Lagesen K, Afset JE, Didelot X, Rønningen KS, Rudi K *et al.* (2013). Comparisons of infant Escherichia coli isolates link genomic profiles with adaptation to the ecological niche. *BMC Genomics* **14**: 81.
Desai MJ, Armstrong DW. (2004). Analysis of native amino acid and peptide enantiomers by high-performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *J Mass Spectrom* **39**: 177–187.
Dobrindt U, Hochhut B, Hentschel U, Hacker J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.
Edgar RC. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
Elliott SJ, Sperandio V, Girón JA, Mellies JL, Wainwright L, Steven W *et al.* (2000). The locus of enterocyte effacement (LEE) -encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic Escherichia coli. *Infect Immun* **68**: 6115–6126.
Fernández De Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H *et al.* (2000). Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol* **35**: 1560–1572.
Fitzjohn RG. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol* **3**: 1084–1092.
Friedberg D, Umanski T, Fang Y, Rosenshine I. (1999). Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic Escherichia coli. *Mol Microbiol* **34**: 941–952.
Friedman M. (1999). Chemistry, nutrition, and microbiology of D-amino acids. *J Agric Food Chem* **47**: 3457–3479.
Goosen N, van de Putte P. (1995). The regulation of transcription initiation by integration host factor. *Mol Microbiol* **16**: 1–7.
Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
Holmes A, Lindestam Arlehamn CS, Wang D, Mitchell TJ, Evans TJ, Roe AJ. (2012). Expression and regulation of the Escherichia coli O157:H7 effector proteins NleH1 and NleH2. *PLoS One* **7**: e33408.
Jahreis K, Bentler L, Bockmann J, Meyer A, Siepelmeyer J, Joseph W *et al.* (2002). Adaptation of sucrose metabolism in the adaptation of sucrose metabolism in the Escherichia coli wild-type. *J Bacteriol* **184**: 5307–5316.
Kaper JB, Nataro JP, Mobley HL. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* **2**: 123–140.
Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Johannes L *et al.* (2012). The enemy within us: lessons from the 2011 European Escherichia coli O104:H4 outbreak. *EMBO Mol Med* **4**: 841–848.
Kenyon CJ, Walker GC. (1980). DNA-damaging agents stimulate gene expression at specific loci in Escherichia coli. *Proc Natl Acad Sci USA* **77**: 2819–2823.
Lam H, Oh D-C, Cava F, Takacs CN, Clardy J, de Pedro Ma *et al.* (2009). D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* **325**: 1552–1555.
Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402–408.
Maiques E, Úbeda C, Campoy S, Lasa Í, Novick RP, Barbé J *et al.* (2006). β-Lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in Staphylococcus aureus. *J Bacteriol* **188**: 2726–2729.
McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* **92**: 1664–1668.
Meador JP, Caldwell ME, Cohen PS, Conway T. (2014). Escherichia coli pathotypes occupy distinct niches in the mouse intestine. *Infect Immun* **82**: 1931–1938.
Michel B. (2005). After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol* **3**: e255.
Miller C, Thomsen LE, Gaggero C, Mosseri R, Ingmer H, Cohen SN. (2013). SOS response induction by β-lactams and bacterial defense against antibiotic lethality. *Science* **305**: 1629–1631.
Moritz RL, Welch RA. (2006). The Escherichia coli argW-dsdCXA genetic island is highly variable, and E. coli K1 strains commonly possess two copies of dsdCXA. *J Clin Microbiol* **44**: 4038–4048.
Nørregaard-madsen M, Fall EMC. (1995). Organization and transcriptional regulation of the Escherichia coli K-12 D-serine tolerance locus. These include: organization and transcriptional regulation of the Escherichia coli K-12 D -serine tolerance locus. *J Bacteriol* **177**: 6456–6461.
Paradis E, Claude J, Strimmer K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee ASN, Luscombe NM. (2012). Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and

HU in Escherichia coli K12. *Nucleic Acids Res* **40**: 3524–3537.

R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Roe AJ, Yull H, Naylor SW, Martin J, Smith DGE, Gally DL *et al.* (2003). Heterogeneous surface expression of EspA translocon filaments by Escherichia coli O157:H7 is controlled at the posttranscriptional level. *Infect Immun* **71**: 5900–5909.

Roesch PL, Redford P, Batchelet S, Moritz RL, Pellett S, Haugen BJ *et al.* (2003). Uropathogenic Escherichia coli use d-serine deaminase to modulate infection of the murine urinary tract. *Mol Microbiol* **49**: 55–67.

Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñiz-Rascado L, García-Sotelo JS *et al.* (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res* **41**: D203–D213.

Shaw RK, Daniell S, Frankel G, Knutton S. (2002). Enteropathogenic Escherichia coli translocate Tir and form an intimin-Tir intimate attachment to red blood cell membranes. *Microbiology* **148**: 1355–1365.

Sims GE, Kim S-H. (2011). Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). *Proc Natl Acad Sci USA* **108**: 8329–8334.

Tarr PI, Fouser LS, Stapleton AE, Wilson RA, Kim HH, Vary JC *et al.* (1996). Hemolytic-uremic syndrome in a six-year-old girl after a urinary tract infection with Shiga-toxin-producing Escherichia coli O103:H2. *N Engl J Med* **335**: 635–638.

Tatsuno I, Nagano K, Taguchi K, Rong L, Mori H, Sasakawa C. (2003). Increased adherence to Caco-2 cells caused by disruption of the yhiE and yhiF genes in enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* **71**: 2598–2606.

Torres AG, Kaper JB. (2003). Multiple elements controlling adherence of enterohemorrhagic Escherichia coli O157:H7 to HeLa cells multiple elements controlling adherence of enterohemorrhagic Escherichia coli O157:H7 to HeLa cells. *Infect Immun* **71**: 4985–4995.

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P *et al.* (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* **5**: e1000344.

Tree JJ, Roe AJ, Flockhart A, McAteer SP, Xu X, Shaw D *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. *Mol Microbiol* **80**: 1349–1365.

Van, Elsas JD, Semenov AV, Costa R, Trevors JT. (2010). Survival of Escherichia coli in the environment: fundamental and public health aspects. *ISME J* **5**: 173–183.

Wang Z, Gerstein M, Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57–63.

Wiles TJ, Kulesus RR, Mulvey MA. (2009). Origins and virulence mechanisms of uropathogenic Escherichia coli. *Exp Mol Pathol* **85**: 11–19.

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH *et al.* (2006). Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* **60**: 1136–1151.

Wolosker H, Dumin E, Balan L, Foltyn VN. (2008). D-amino acids in the brain: D-serine in neurotransmission and neurodegeneration. *FEBS J* **275**: 3514–3526.

Wong ARC, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF *et al.* (2011). Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. *Mol Microbiol* **80**: 1420–1438.

Yao Y, Xie Y, Perace D, Zhong Y, Lu J, Tao J *et al.* (2009). The type III secretion system is involved in the invasion and intracellular survival of Escherichia coli K1 in human brain microvascular endothelial cells. *FEMS Microbiol Lett* **300**: 18–24.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)