Investigating the role of sugar uptake and metabolism in the regulation of Enterohaemorrhagic *Escherichia coli* virulence

Curtis Cottam

A Thesis Submitted for the Degree of Doctor of Philosophy

2020-2023

Biosciences Institute

Faculty of Medical Sciences

Newcastle University

Abstract

Enterohaemorrhagic *Escherichia coli* (EHEC) is a major foodborne pathogen of the human colon, and cause for zoonotic disease. Transmitted via the faecal-oral route, EHEC exhibits a remarkably low infectious dose, resulting in outbreaks that manifest in bloody diarrhoea, and in extreme cases, acute renal failure. Treatments against EHEC are limited due to conventional antibiotics exacerbating infection. Understanding EHEC pathogenesis is therefore crucial to the development of novel treatment strategies.

EHEC have evolved to sense environmental nutrients as “signals” to fine-tune the expression of their primary virulence factor, the Type 3 Secretion System (T3SS), which is essential for host-cell colonisation. These signals include sugars and fats ingested as part of the diet, or by-products of metabolism by the gut microbiota. However, the mechanisms underlying how EHEC exploit these different nutrients are poorly understood.

Here, a novel ATP-binding cassette (ABC) transporter in the murine pathogen *Citrobacter rodentium*, commonly used as a surrogate model for EHEC, was characterised. This system, known to be upregulated during murine infection, is specific for D-ribulose and likely aids colonisation of the mouse gut. Searches in EHEC revealed a similar ABC transporter encoded on a horizontally-acquired genetic element to be significantly enriched amongst EHEC strains. Transcriptionally, the locus was regulated exclusively by L-arabinose, in an AraC-dependent manner. Furthermore, growth on L-arabinose significantly enhanced T3SS expression and the ability to attach to host cells. Deletion of the genes required for L-arabinose uptake, metabolism and associated regulation revealed this phenotype to rely on L-arabinose breakdown and not merely “sensing” its presence in the environment. Collectively, this work suggests L-arabinose metabolism to be important in EHEC pathogenesis through providing a source of nutrition and enhancing virulence gene regulation. It is proposed that these systems and their substrates allow EHEC to outcompete the native gut microbiota, with their downstream metabolism contributing to virulence regulation.

Acknowledgments

First and foremost, I would like to give a special thank-you to my supervisor, Dr. James Connolly, for providing me with this incredible opportunity. His unwavering support, encouragement, and boundless enthusiasm over the last three years have been instrumental in shaping me into the scientist I am today. I am also deeply grateful to my co-supervisory team, Dr. Elisabeth (Lis) Lowe and Dr. Christopher Stewart, whose expertise and guidance have allowed me to explore and conduct research in their respective fields. A further thank-you to Lis and her partner in crime, Dr. David Bolam, for their jokes, Alvino's antics, and occasional firing with Nerf darts.

Several thanks are also needed for all the technical support I have been given throughout the project including that from Helen Glenwright for her expertise and time in tissue/cell culture, Jon Chapman for his guidance during organoid experiments and Maisie Palmer for her patience with my (lack of) microscopy skills.

I cannot forget to acknowledge the fantastic group of fellow PhD students who have carried me through along the way. From the Khan lab, I'd like to extend my appreciation to Beth and Jack. From the Lowe/Bolam lab, my thanks go to Cossette, Diana, Joe, and Omar. And from the Palmer lab, a heartfelt shout-out to Andy, Ellie, José, Kieran, Stephen, and Yaping (YP). Your support, camaraderie, laughter, and countless post-work pints at the pub have not only made the tough times bearable and the good times even better but incredibly memorable. My thanks for those around me also extend to the ever "professional" group of post-docs: Amy, Emm, Merel and Nicky for their help (and the occasional much-needed dose of tequila).

I'd also like to give my appreciation to my friends back in the motherland for pretending to know what I've been doing over the last few years. But also reminding me that there is more than just work. Unfortunately, I cannot prescribe for, diagnose, or solve your medical problems. Similarly, my family, who have always

encouraged me to aspire high and given me unmeasurable support (and not just financially).

Finally, the biggest and most special thanks to Lauren with MMI5JU (and Beck/Laing/Yates clan) for being my rock, helping me keep my sanity and coming along on the journey with me. I genuinely could have not done this without you.

Table of Contents

List of Figures

List of Tables

Abbreviations

Δ	Deletion
A/E	Attaching and effacing
ABC	ATP-binding cassette
AHL	Acyl-homoserine lactone
AI	Autoinducer
AIEC	Adherent Invasive *E. coli*
Amp	Ampicillin
Ara	L-arabinose
ATP	Adenosine triphosphate
AUG	L-arabinose utilisation gene
BHI	Brain heart infusion
BMM	*Bacteroides* minimal media
bp	Base pairs
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CFU	Colony forming units
Cm	Chloramphenicol
COGs	Cluster of Orthologous Groups of proteins
CR	Colonisation resistance
CRP	cAMP receptor protein
DAEC	Diffusely Adherent *E. coli*
DEG	Differentially expressed gene
DNA	Deoxyribonucleic acid
EAEC	Enteroaggregative *E. coli*
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ETEC	Enterotoxigenic *E. coli*
ETT2	*E. coli* T3SS 2
ExPEC	Extraintestinal pathogenic *E. coli*
FA	Fatty acid
FBS	Foetal bovine serum
FDR	False discovery rate
Gb_3	Globotriaosylceramide 3
gDNA	Genomic DNA
Gent	Gentamicin
GH	Glycoside hydrolase
GI	Gastrointestinal
H-NS	Histone-like nucleoid structuring protein
HCP	Haemorrhagic coli pilus
HGT	Horizontal gene transfer
HK	Histidine kinase
HTH	Helix-turn-helix
HUS	Haemolytic uremic syndrome
IEC	Intestinal epithelial cell
IHF	Integration host factor

IM	Inner membrane
InPEC	Intestinal *E. coli*
Kan	Kanamycin
Kb	Kilobases
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Lysogeny broth
LCFA	Long chain FA
LDS	Lithium dodecyl sulphate
LEE	Locus of enterocyte effacement
Ler	LEE encoded regulator
LPS	Lipopolysaccharide
LUX	Luminescence
MLST	Multi locus sequence typing
MOI	Multiplicity of infection
mRNA	Messenger RNA
MW	Molecular weight
N-WASP	Neuronal Wiskott-Aldrich syndrome protein
Nal	Nalidixic acid
NAP	Nucleoid associated protein
NBD	Nucleotide binding domain
NCBI	National Centre for Biotechnology Information
NLE	Non-LEE encoded
NMEC	Neonatal meningitis-associated *E. coli*
NS	Non-significant
OD	Optical density
OI	O-island
OM	Outer membrane
ORF	Open reading frame
PAI	Pathogenicity island
PBP	Periplasmic binding protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PPP	Pentose phosphate pathway
QS	Quorum sensing
^{R}	Resistance
RLU	Relative luminescence units
RNA	Ribonucleic acid
RNAP	RNA polymerase
rpm	Revolutions per minute
RR	Response regulator
RT-qPCR	Reverse transcriptase quantitative PCR
SCFA	Short chain FA
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression

sRNA	Small RNA
STEC	Shiga toxigenic *E. coli*
Stx	Shiga toxin
T2SS	Type 2 secretion system
T3SS	Type 3 secretion system
TAE	Tris-acetate-EDTA
TCA	Trichloroacetic acid
TCS	Two-component system
TEER	Trans-epithelial electrical resistance
Tir	Translocated intimin receptor
TLC	Thin layer Chromatography
TMD	Transmembrane domain
tRNA	Transfer RNA
UPEC	Uropathogenic *E. coli*
V	Volts
VF	Virulence factor
WT	Wild type

1. Introduction

1.1. *Escherichia coli*

Escherichia coli is a Gram-negative rod-shaped bacterium belonging to the Enterobacteriaceae family and was first described by Theodor Escherich in 1885 (Riley, 2020). Its exceptional genetic tractability has facilitated the understanding of several fundamental concepts in modern genetics, spanning transcription, translation, and DNA replication (Blount, 2015). *E. coli* has therefore become widely recognised as the 'workhorse' of modern molecular biology. Additionally, due to its ability to grow in both the presence and absence of oxygen, as well as use miscellaneous nutrients, *E. coli* exhibits exceptional metabolic versatility (Geurtsen et al., 2022). This has enabled the species to survive in diverse environments such as water, soil, and the mammalian gut (Foster-Nyarko & Pallen, 2022).

1.2. Diversification of *E. coli*

1.2.1. Commensal *E. coli*

A pioneer species of the human gut, *E. coli* is ubiquitously found following colonisation shortly after birth and constitutes 0.1-5 % of the intestinal microbiome (Blount, 2015; Martinson & Walk, 2020). Thereafter, *E. coli* typically exist as a harmless commensal, defining the intestinal environment and providing benefit to the host via their ability to produce essential vitamins (K and B_{12}) (Blount, 2015), exclude pathogens through competition (Fabich et al., 2008) and drive mucosal integrity (Hering et al., 2014). For these reasons, protective *E. coli* strains such as *E. coli* Nissle 1917 have been used as probiotics in the treatment and prevention of several gastrointestinal (GI) disorders, primarily ulcerative colitis, and irritable bowel syndrome (Gronbach et al., 2010; Pradhan & Weiss, 2020). However, variants of *E. coli* deviate from being beneficial, including those traditionally used as probiotics, when it is environmentally favourable to do so (Bleich et al., 2008; Gronbach et al., 2010). Instead, these deviators negatively impact host health and can act as the progenitor to disease.

1.2.2. Commensal-pathogen switch

Studies have demonstrated that compositional changes to the gut milieu, owing to altered microbiota composition, a compromised immune system, and uncompetitive niches, can induce a phenotypic switch amongst some commensals (Bhat et al., 2019).

Whilst the environment is undoubtedly extremely important in driving this switch for commensal *E. coli* to become pathogenic, the emergence of virulence is in fact far more complex. Selective pressures conferred by the environment can often lead to genetic changes that result in a pathogenic phenotype (Touchon et al., 2020). Genetic changes that allow for commensal-pathogen transitioning can include horizontal gene transfer (HGT) via mobile elements such as plasmids and phage, antagonistic pleiotropy whereby anti-virulence genes are inactivated, or pathoadaptive mutations that result in a change of gene function (Denamur et al., 2020) (**Figure 1-1**). The apparent malleability of the genome is therefore likely to benefit the organism under hostile conditions where changes are required to survive, therefore driving the evolution, diversification, and adaptation of the species. Due to the association with pathogens, genetic features such as virulence plasmids and pathogenicity islands (PAIs) are frequently used in their identification and classification. PAIs are large (i.e., 10-200 Kb) integrative islands with their own genomic characteristics, and therefore useful pathogenic signatures (Desvaux et al., 2020). Furthermore, differences in the G+C content of PAIs are reflective of recent acquisition and highlight integrative events, increasing their identifiability (Desvaux et al., 2020).

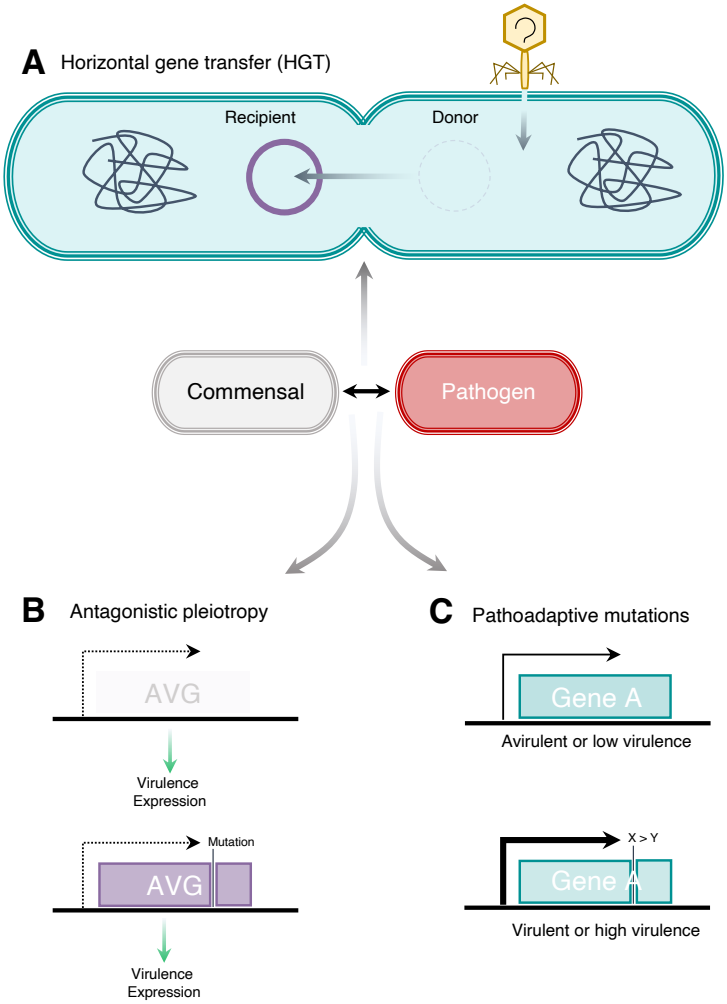


Figure 1-1. Overview of genetic changes that drive the commensal-pathogen switch in *E. coli*. Genetic changes that can influence commensal *E. coli* to become pathogenic include: (**A**) HGT of mobile elements that can confer virulence traits; (**B**) antagonistic pleiotropy whereby anti-virulence genes (AVG) are inactivated (i.e., by mutations, deletions); (**C**) pathoadaptive mutations such as point mutations which cause a change of function in the resultant protein, making them become virulent from non-virulent or more greatly virulent when already responsible for virulence.

The role of genetic changes has been substantiated by the ability to experimentally transform the commensal *E. coli* strain K-12 into a pathogen following minor changes to its genomic content. Examples include mutations in histone-like proteins (such as HUα) (Koli et al., 2011) and single transposon gene insertions that allow for increased intracellular macrophage survival (Proença et al., 2017). $HU\alpha^{E38K,V24L}$ mutants were found to undergo significant transcriptional reprogramming to favour the expression of pathogenicity genes such as haemolysin E and curli fibres (Koli et al., 2011). The *E. coli* genome is therefore extremely 'plastic' and an ability to easily acquire genetic changes has contributed to the phenotypic diversification of the species. However, the notion that singular genetic changes are sufficient to drive commensal-pathogen

transitioning, is an oversimplification, especially as non-pathogenic *E. coli* carrying genes for virulence traits have been isolated (Kaper et al., 2004). This is best exemplified by *E. coli* Nissle, which shares numerous virulence factors (VFs) (i.e., adhesins, toxins and effectors) with clinically relevant pathogenic *E. coli* strains such as the uropathogenic strain CFT073 (Van Der Hooft et al., 2019). It may be that in the context of the rest of the *E. coli* Nissle genome, the presence of these VFs simply do not confer pathogen-like phenotypes. Moreover, these factors when expressed may have different roles for the lifestyle of *E. coli* Nissle.

1.2.3. Evolution of *E. coli* pathotypes

A strong phylogenetic structure exists amongst *E. coli*, and strains can be sub-categorised into eight main phylogenetic groups (phylogroups): A, B1, B2, C, D, E, F and G (Desvaux et al., 2020). These lineages are based on PCR patterns and multi-locus sequence typing (MLST) of candidate house-keeping genes (Geurtsen et al., 2022). However, whilst only < 3 % nucleotide divergence is seen across conserved genes in *E. coli*, a much greater difference can be observed in the gene content of genome pairs, diverging by > 30 % (Touchon et al., 2009). Therefore, whilst extremely useful in rapid epidemiological surveillance, the use of MLST in inferring the true phylogeny of bacteria remains controversial (Tsang et al., 2017). In fact, direct comparison of phylogenetic reconstruction using MLST differs considerably to that based on genome sequencing and single nucleotide polymorphisms (SNPs), which correspond with each other, therefore failing to represent potential microevolution events (Tsang et al., 2017).

Phylogroup D is widely accepted as having diverged first from the common ancestor (Sims & Kim, 2011) and displays the greatest similarity with the *E. coli* origin (Gonzalez-Alba et al., 2019). Pronounced differences can therefore be seen at the phylogroup level. For example, the genome of strains belonging to A and B1 appear smaller than B2 or D strains (Touchon et al., 2020). However, while grouping may suggest similarity between strains of the same group, phenotypic and genotypic differences within individual phylogroups do exist, with no single group comprising only a single pathotype (Touchon et al., 2020). A pathotype is defined as a pathogenic variant of *E. coli* that is taxonomically related to those strains that asymptomatically colonise the host (Riley, 2020). These intra-phylogroup differences are the likely

outcome of the high genome plasticity *E. coli* possess as well as extremely high recombination event occurrence. In fact, the *E. coli* genome ranges in size from 4.2 to 6 Mb, with all strains only sharing around 2,000 genes, comprising the estimated 'core' genome (Touchon et al., 2020). As a 'core' set of genes, they encode mainly essential gene products necessary for normal cellular functioning (Gonzalez-Alba et al., 2019). The pangenome (defined as the total number of genes present in a species, across all isolates (Brockhurst et al., 2019)), on the other hand, greatly exceeds this value and differs dependent on the number of strains included in analyses (Geurtsen et al., 2022). As an instance, the pangenome of *E. coli* was previously reported to encompass around 18,000 genes (Touchon et al., 2009). However, with the increased sequencing of *E. coli* genomes over the past decade (Brockhurst et al., 2019), it is now believed to consist of roughly > 75,000 genes (Touchon et al., 2020).

Diversity is observed amongst the strategies pathogenic *E. coli* employ during pathogenesis and are often used to define the 'pathotype'. However, there appears only to be weak associations with pathotype class and phylogroup, reflecting some disorder of *E. coli* phylogenomics. Instead, more general assumptions can be made on the associated pathotypes with each phylogroup. For example, phylogroup A is described as largely encompassing commensal *E. coli* strains whereas B2 is generally accepted as being comprised predominantly of extraintestinal pathotypes (Desvaux et al., 2020). Pathotypes can be broadly categorised based on their site of colonisation (i.e., the intestine, urinary tract, etc.), with isolates associated with both intestinal and extraintestinal disease being extremely rare.

A total of nine *E. coli* pathotypes are widely recognised. Of these, the majority (7/9) are enteric pathogens (intestinal pathogenic *E. coli* or InPEC) responsible for causing diarrhoea and intestinal disorders. InPEC comprises of the pathotypes Shiga toxin-producing *E. coli* (STEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli*, enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC) and adherent-invasive *E.coli* (AIEC) (Kaper et al., 2004; Santos et al., 2020). Strains of each pathotype are grouped by similarity in virulence associated traits such as their virulence factor repertoire, tissue tropism, interaction with the host and clinical symptoms (Pokharel et al., 2023). Unlike InPEC, extraintestinal pathogenic *E. coli* (ExPEC) are more difficult to categorise into pathotypes due to there being no

single or set of VFs that can be used as markers in the identification and classification of these strains (Santos et al., 2020). Instead ExPEC are typically defined by their site of isolation during an infection (Santos et al., 2020). The most common ExPEC pathotypes include uropathogenic *E. coli* (UPEC) isolated from patients with urinary tract infections or bloodstream infections as well as neonatal meningitis-associated *E. coli* (NMEC) (Leimbach et al., 2013).

1.3. Enterohaemorrhagic *E. coli*

1.3.1. Origin and prevalence of EHEC

Enterohaemorrhagic *E. coli* (EHEC) is a subset of the STEC pathotype, first identified in 1982 after an outbreak of haemorrhagic colitis was associated with the ingestion of uncooked meat in the United States (Riley et al., 1983). With a remarkably low infectious dose (10-100 cells), EHEC colonises the distal portion of the large intestine and is responsible for causing diarrhoeal disease and in extreme circumstances, acute renal failure (Correa-Martinez et al., 2022). The clinical manifestations of EHEC infection are therefore extremely variable, with the onset of mild symptoms (abdominal cramps, non-bloody diarrhoea) occurring 3-4 days post ingestion and developing there onwards (Cramer, 2014). Subsequently, EHEC has caused major public concern, particularly as diarrhoeal disease causes much mortality worldwide (Croxen et al., 2013). EHEC continues to cause sporadic outbreaks that exert a public health burden, exemplified by the occurrence of novel hybrid strains including that responsible for a significant outbreak of bloody diarrhoea and haemolytic uremic syndrome (HUS) in Germany in 2011 (Prager et al., 2014). This outbreak led to a total of 53 deaths, > 800 cases of HUS and ~3,000 reports of gastroenteritis (Prager et al., 2014). It was later identified that sprouts were the vehicle of infection (Mellmann et al., 2012), and notably, EHEC outbreaks since have been more greatly associated with contaminated vegetables as opposed to meat.

The impact of EHEC is accentuated by the broad spectrum of clinical symptoms that vary in severity and a complete absence of specific therapeutic treatments (Goldwater & Bettelheim, 2012). Resultantly, treatments are mainly supportive, and limited to rehydration, analgesics, and adequate nutrition (Blount, 2015; Goldwater & Bettelheim, 2012).

1.3.2. EHEC reservoirs

Ruminant livestock are capable of harbouring EHEC, whether that to be to act as spill over hosts or dead ends (Persad & LeJeune, 2014). However, epidemiological studies, as reviewed by Persad and LeJeune (2014), have long-established cattle as the primary reservoir of EHEC. In cattle, the pathogen persists, and asymptomatically colonises the recto-anal junction (Naylor et al., 2003). Transmission of EHEC to humans therefore largely occurs after the ingestion of food and/or water contaminated with faeces that can contain from 100 to 10^6 CFU of O157:H7 per gram (Croxen et al., 2013). The process of EHEC transmission between reservoir and hosts is summarised in **Figure 1-2**.

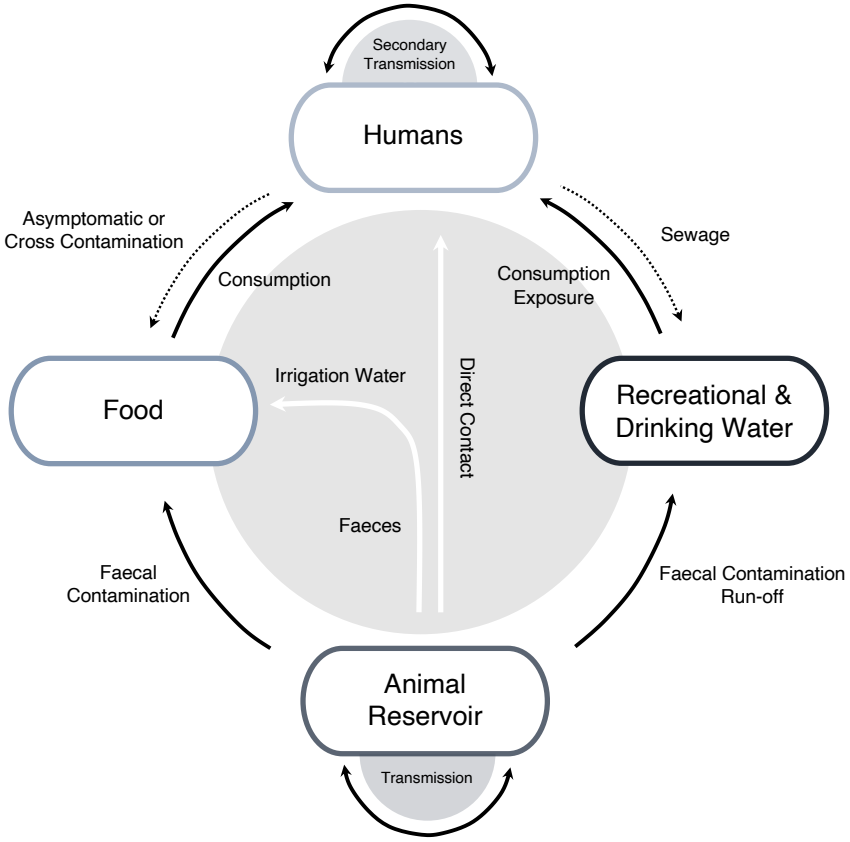


Figure 1-2. Overview of reservoirs and modes of transmission in EHEC. The primary reservoir of EHEC is typically animals such as cattle where they colonise asymptomatically. The bacterium is transferred to the environment via animal faeces which contaminates food (i.e., crops) and water (both recreational and drinking water). Human infection with EHEC can therefore occur indirectly by ingestion of contaminated food or water, or directly through contact with animals. Secondary transmission can then occur between humans. Figure taken and adapted from Croxen *et al.*, (2013).

1.3.3. EHEC O157:H7

E. coli O157:H7 is the serotype predominantly isolated from infected individuals of the Western world (U.S., UK, Japan), and has been subject to extensive genomic analyses. In the UK alone, the serotype is responsible for approximately 800 cases/year (Byrne et al., 2018). In the U.S., the incidence of O157:H7 infections is much higher and is estimated by the Centre for Disease Control to be around 95,000 cases/year (36 % of 265,000 total STEC infections). The prototypical strain, EDL933, has a 5.5 Mb chromosome and a 90 Kb virulence plasmid (Perna et al., 2001; Warr et al., 2019). Various comparative studies looking to determine differences in the genetic content of non-pathogenic *E. coli* and EHEC have helped to shed light on genetic factors contributing to pathogenesis. Mainly, EHEC employ two major virulence strategies: (1) the production of a Shiga toxin (Stx) and (2) the formation of attaching and effacing (A/E) lesions (Wick et al., 2005). Both strategies are the outcome of horizontally acquired genomic changes over evolutionary time, with O157:H7 being thought to have evolved in a stepwise manner from EPEC O55:H7; a recent ancestor to a non-toxigenic progenitor (Reid et al., 2000; Wick et al., 2005) (**Figure 1-3**).

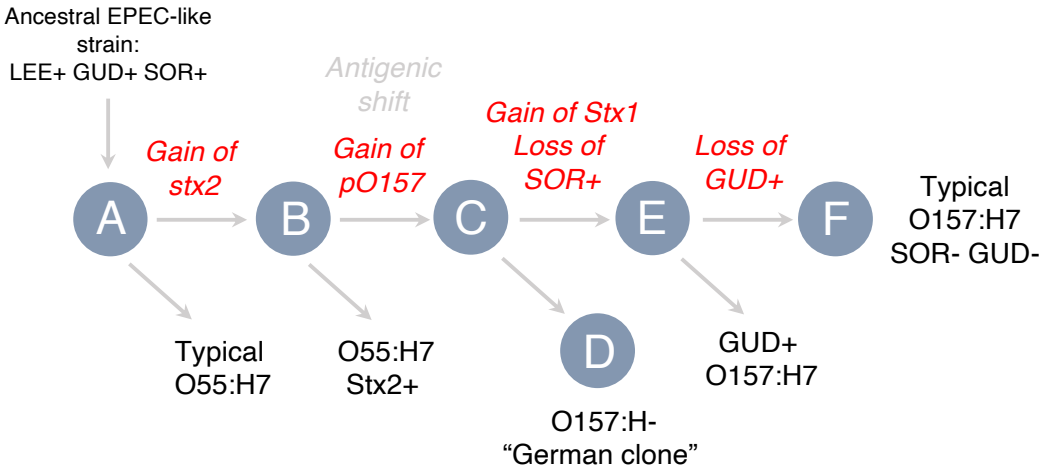


Figure 1-3. Stepwise model of evolution in O157:H7. Through a series of sequential evolutionary steps, O157:H7 emerged from an ancestral EPEC-like strain by gain or loss of genomic traits. SOR (sorbitol fermentation). GUD (ß-glucuronidase activity). (+) denotes strains positive for a trait. (-) denotes strains negative for a trait. Figure adapted from Wick *et al.*, (2005).

In a seminal study by Perna *et al*. in 2001, the complete genome sequencing of O157:H7 strain EDL933 allowed for direct comparison to K-12, elucidating the specific genetic traits of EHEC (Perna et al., 2001). Evidence of HGT within the genome of O157:H7 is extensive, with clusters spanning 1,387 novel genes unique to EHEC identified (Perna et al., 2001). These gene clusters, defined as O-islands (OIs), represent 26 % of the O157:H7 genome (**Figure 1-4**) (Perna et al., 2001). Of those OI genes characterised, functional analyses have revealed a majority to have roles in pathogenesis, encoding both VFs and regulatory proteins (Jiang et al., 2021). This has been functionally substantiated and the deletion of several OIs have significant effects on the virulence profile of EDL933 (Flockhart et al., 2012). Although less common, multiple OI genes have also been associated with biological processes (Jiang et al., 2021) and are predicted to encode systems specific for nutrient transport (Perna et al., 2001) suggesting a role in fitness as opposed to virulence. Therefore, whilst there is a strong association with OI genes and virulence this is not exclusive, especially considering the vast majority of genes within OIs (> 90%) remain to be characterised (Jiang et al., 2021). It is important to note that the described OI traits are EDL933-specific and variation in the diversity and carriage of OIs varies across EHEC/STEC strains (Cooper et al., 2014), again adding to the genotypic diversity amongst lineages.

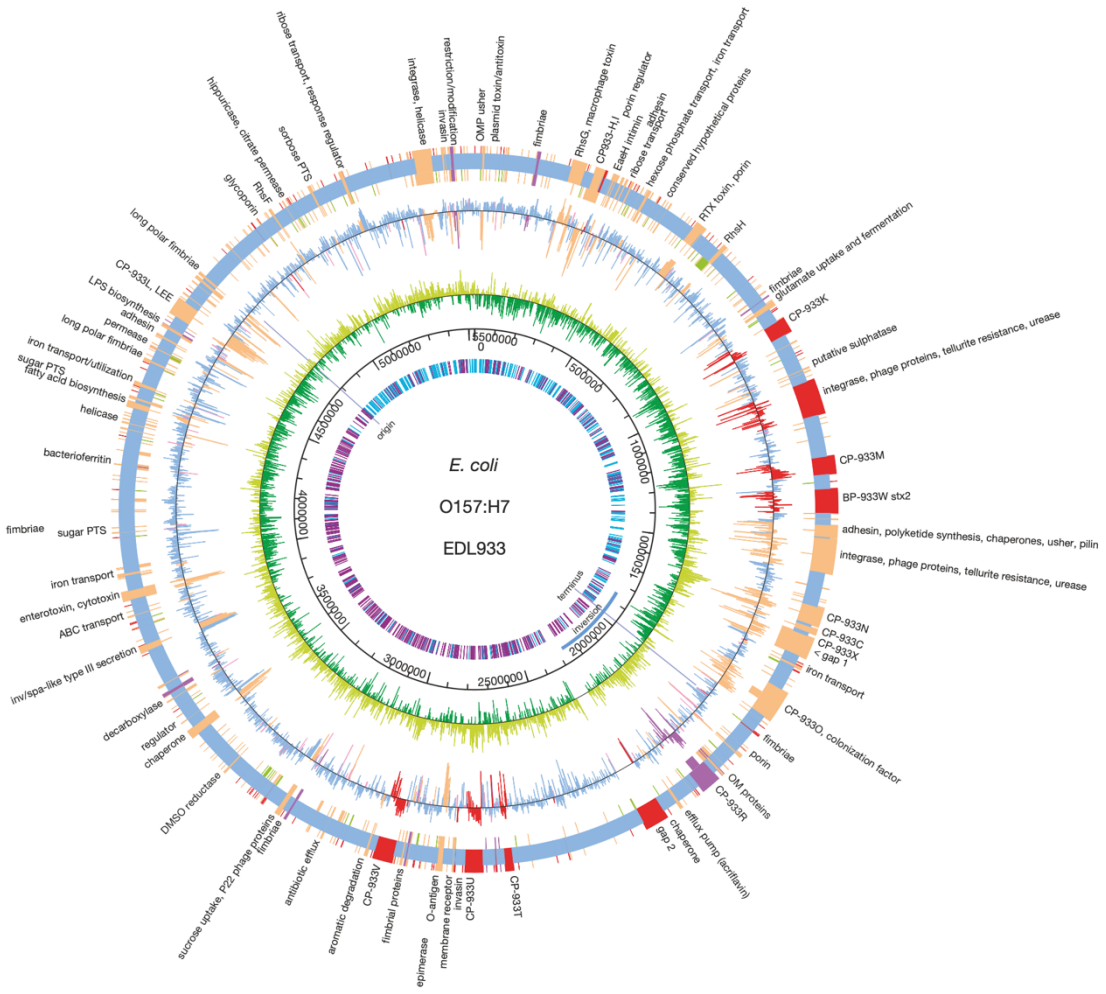


Figure 1-4. Comparative genome map of *E. coli* O157:H7 EDL933 with *E. coli* MG1655. The outer ring shows the co-linear backbone shared between strains (blue), with the distribution of EDL933 specific regions (O-islands) and MG1655 specific regions (K-islands) highlighted in red and green respectively. Those highlighted in tan represent overlapping regions of O- and K-islands. Purple regions denote an area of hypervariability. The origin and terminus of replication is also shown. Figure taken from Perna *et al.*, (2001).

1.4. Pathogenesis of EHEC

1.4.1. Locus of Enterocyte Effacement (LEE)

The locus of enterocyte effacement (LEE) is a large chromosomally located PAI. The LEE of EHEC spans 35 Kb, covering 41 open reading frames (ORFs) (**Figure 1-5**) that encode the necessary proteins for A/E lesion formation on host intestinal epithelial cells (IECs). A/E formation is a major hallmark of EHEC infection and is characterised by actin rich membrane protrusions known as pedestals and destruction of the intestinal villi (Croxen et al., 2013; Egan et al., 2019; Kaper et al., 2004).

As discussed in *section 1.2.2.*, PAIs are recognised as being important for virulence and can be used as discriminatory genetic factors in the identification of pathogenic *E. coli* strains, particularly those of an A/E phenotype (Rumer et al., 2003). Similarly, the lower G/C content of the LEE (38.8 %) against the G/C content of the chromosome (50.8 %) is indicative that the element was acquired following HGT (Perna et al., 1998). The LEE of EHEC has a conserved core displaying 93 % sequence similarity with that of its close relative EPEC (Perna et al., 1998). High sequence conservation in the LEE genes likely reflects their essentiality for the pathogenic lifestyle. As with other PAI, the LEE contrasts mobile elements such as plasmids and bacteriophage due to the absence of an origin of replication, an inability to self-mobilise and having restricted replication (Desvaux et al., 2020). Subsequently, acquisition of the LEE by EHEC is hypothesised to have been aided by bacteriophage (Desvaux et al., 2020). Variation observed in PAIs can be attributed to the mobile elements that they carry, as these have important roles in recombination and therefore genetic rearrangements (Desvaux et al., 2020). Genomic and evolutionary studies have localised the LEE insertion point to regions of hypervariability, such as the *selC* gene, which encodes a selenocysteine tRNA (Perna et al., 1998).

Structurally, genes of the LEE are organised into five main polycistronic operons (LEE1-5) (**Figure 1-5**). The bulk of LEE1-3 encode apparatus needed to form a functional Type Three Secretion System (T3SS) (Described further in *section 1.4.2.*), whilst the major secreted proteins intimin and translocated intimin receptor (Tir), required for the adherence of EHEC to IECs, are encoded by LEE5 (Lara-Ochoa et al., 2023). Genes encoding the proteins EspB and EspD, responsible for forming the translocation pore in host cells, are in LEE4 (Lara-Ochoa et al., 2023) (**Figure 1-5**).

Importantly, the master regulator of the LEE, defined as Ler (LEE encoded regulator), is encoded as the first ORF in LEE1 and is responsible for controlling the expression of the entire island (Elliott et al., 2000). Additionally, the LEE harbours two bicistronic operons (*grlRA* and *espG-orf1*) (**Figure 1-5**) also with regulatory functions (Egan et al., 2019; Lara-Ochoa et al., 2023), discussed further in *section 1.5.2*.

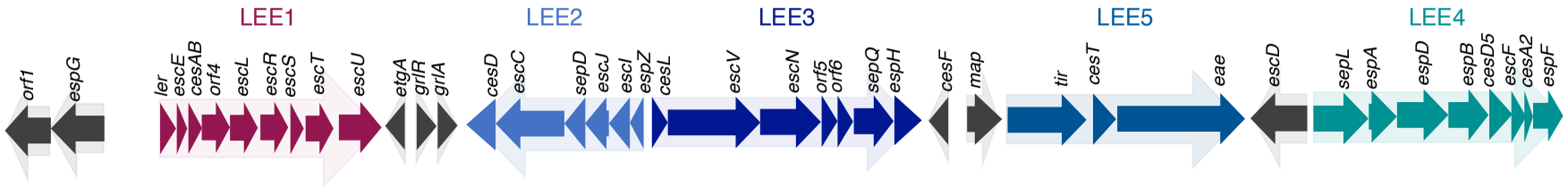


Figure 1-5. Organisation of the Locus of Enterocyte Effacement (LEE) in EHEC. The genomic island comprises of five main polycistronic operons denoted as LEE1, LEE2, LEE3, LEE5 and LEE4, spanning a 35 Kb region of the genome. The genes encoded within each of these operons is shown. Transcriptional direction is also indicated by the arrow direction.

1.4.2. Type III Secretion System (T3SS)

Bacterial secretion systems are large macromolecular complexes that allow the translocation of proteins across the cell envelope and into the surrounding environment (Filloux, 2022). Additionally, bacterial secretion systems can be used to establish direct contact with both prokaryotic and eukaryotic cells, facilitating the cell-to-cell transfer of proteins (Filloux, 2022). Their presence across the bacterial kingdom is widespread and at least 11 types (Type I-XI) have now been described, albeit to varying degrees (Filloux, 2022).

The LEE encodes a T3SS which comprises of > 15 different proteins that collectively span the inner membrane (IM), periplasmic space, and outer membrane (OM) (Costa et al., 2015) (**Figure 1-6**). The T3SS was initially investigated using genetic and biochemical techniques, however, advancements in structural biology have allowed a more complete understanding of the components that constitute the system (Schraidt & Marlovits, 2011). The primary function of the T3SS is to act as a conduit for effector protein (i.e., toxin) translocation into host cells and was the first example of a mechanism essential for bacterial pathogenesis (Filloux, 2022). The T3SS is not exclusive to *E. coli* and diverse forms of this VF are present amongst other clinically relevant Gram-negative pathogens such as *Shigella, Yersinia* and *Salmonella* spp. (Green & Mecsas, 2016). Though nine known core components of the T3SS display conservation, the secretion system has been described to differ in their assembly, regulation, and effector repertoire (Deng et al., 2017; Du et al., 2016; Green & Mecsas, 2016). For example, the number of T3SS effectors in both *Pseudomonas* and *Yersinia* spp. are much lower than in *Shigella* and EHEC (Green & Mecsas, 2016). However, despite being a close relative *E. coli*, the T3SS of *Shigella* displays greater similarity to the T3SS of *Salmonella* (Green & Mecsas, 2016). These differences are also reflected in the pathogenicity islands within which they are encoded.

In the case of EHEC, the T3SS is primarily responsible for the intimate attachment to IECs (Gaytán et al., 2016). This attachment is shortly followed by effacement of the microvilli border and formation of pedestal-like structures that are rich in actin, resulting in A/E lesion formation (Gaytán et al., 2016). These changes to cellular physiology are mediated via the delivery of effector proteins into the host cell cytoplasm (See *section 1.5.*). Reliance on the T3SS to form A/E lesions on host cells is also observed for some

specific pathogens outside of the human host, such as the murine pathogen *Citrobacter rodentium* (Ruano-Gallego et al., 2021). *C. rodentium* possesses a highly homologous T3SS to EHEC and is therefore frequently used as a surrogate model to study EHEC infection mechanisms *in vivo*, particularly as EHEC is unable to naturally colonise mice (Collins et al., 2014). That being said, the effector repertoire of *C. rodentium* does differ in both the number and therefore function compared to EHEC (Deng et al., 2012).

Architecturally, the T3SS can be divided into three recognisable substructures:

(1) Extracellular segments

The extracellular component is formed by the needle protein (EspF), filament (EspA) and translocation pore (EspB/D) (**Figure 1-6**). EspA is polymerised to form extensive filaments that adjoin the bacterial cell cytoplasm to host cell membrane (Crepin et al., 2005). EspD caps the EspA filament and constitutes the main component of the translocation pore embedded in the host cell membrane (Crepin et al., 2005). EspD interaction with EspB is also required to form a heterooligomeric pore for the translocation of effector proteins into target cells. Interactions between the translocation pore and filament are founded by EspA and EspD, not EspB (Tejeda-Dominguez et al., 2017). Notably, *espA* gene expression is gradually downregulated over time, indicative of its crucial role in the attachment and effector translocation phase of infection (Crepin et al., 2005). Finally, EspA associates with the basal body of the T3SS embedded in the bacterial OM via EscF (Crepin et al., 2005), which is reliant upon the chaperones EscE and EscG (Sal-Man et al., 2013).

(2) Basal body

The basal body with extracellular segments is termed the 'needle complex' (Makino et al., 2016). In its entirety, the body comprises three concentric rings (EscD, EscJ, EscC) that encircle the needle and inner rod (EscI) which anchor to the IM (Deng et al., 2017). It is within these regions that the export apparatus (EscRSTUV) is located (**Figure 1-6**). The export complex is recognised to possess the highest conservation of all the T3SS substructures, indicating its essentiality for T3SS function (Gaytán et al., 2016).

(3) Cytoplasmic components

T3SS cytoplasmic components, as with the export apparatus, are less understood (Makino et al., 2016). The region is a highly complex and dynamic substructure noted to undergo large-scale confirmational changes, with some components seemingly able to be lost (Makino et al., 2016). An ATPase complex (EscN, EscO, EscL) is required for the active transport of effector proteins across the translocon and localised to the cytoplasm (Gaytán et al., 2016). Infant rabbit colonisation studies, whereby the *escN* gene in EHEC was deleted, led to a defective T3SS *in vivo* (Ritchie & Waldor, 2005). This likely reflects the inability to hydrolyse ATP and therefore generate the energy to transport substrates across the translocon into target cells. Non-functionalisation of the T3SS via EscN has made it an interesting target in the design of small inhibitors that block T3SS activity (Bzdzion et al., 2017). The ATPase also interacts with the annular C-ring, composed of EscQ, thought to act as a sorting platform for effectors prior to their translocation (Gaytán et al., 2016).

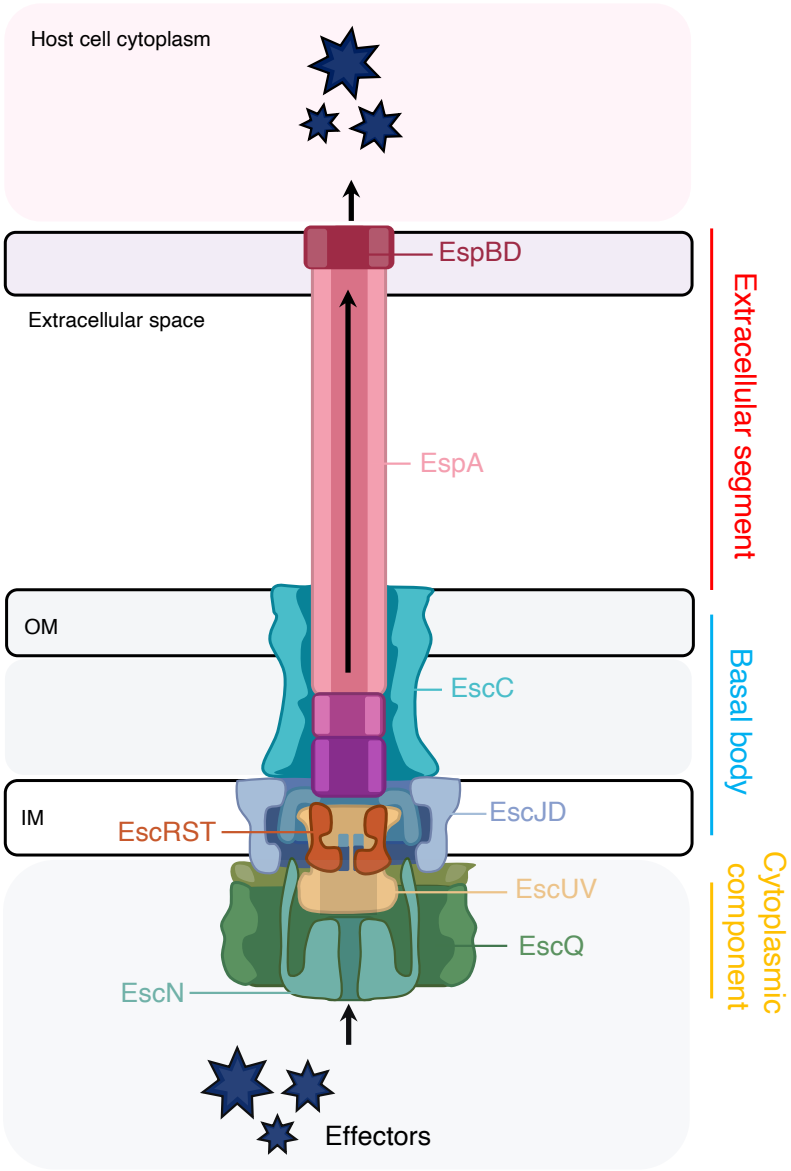


Figure 1-6. Schematic of the T3SS in EHEC. The T3SS is a multiprotein complex forming a needle structure that spans the bacterial IM, periplasmic space, and OM, divided into extracellular segments (red), basal body (blue) and cytoplasmic components (yellow). The filament (EspA) spans the extracellular space to allow for pore formation in the host membrane via translocator proteins (EspD, EspB). Effectors proteins are then delivered via the translocon into target host cells. Figure made in part using BioRender.

1.4.3. T3SS effectors and A/E lesion formation

T3SS effector proteins can be defined as early, intermediate, and late substrates (Deng et al., 2017). Spatiotemporally controlled, the secretion of substrates is highly dynamic and has been shown to be dependent on specialised chaperones to ensure their hierarchical secretion (Serapio-Palacios & Finlay, 2020). Substrates are recruited in an unfolded state following the recognition of an N-terminally located secretion signal, specific for the T3SS (Slater & Frankel, 2020)

Prior to employing the T3SS, EHEC must first form an initial adherence to target IECs. The mechanisms which underly this process are poorly understood compared to the intimate attachment conferred by the T3SS that proceeds. Adherence is thought to occur in a localised manner that is reliant on type IV pili known as haemorrhagic coli pili (HCP) (Gaytán et al., 2016). Once adhered, the T3SS is expressed, and intimate attachment is established by the secretion of Tir into host cells (Kenny, 2001; Kenny et al., 1997). Tir is then embedded into the host cell membrane where it acts as a receptor for the bacterial OM adhesin, intimin (**Figure 1-7**) (Kenny, 2001). Following interaction with intimin, Tir interacts directly with host cell insulin receptor tyrosine-kinase substrate (IRTKS; IRSp53) that recruits the non-LEE encoded (NLE) effector protein, EspFu (**Figure 1-7**) (Martins et al., 2020). After localisation to the bacterial attachment site, EspFu activates the mammalian protein, Neuronal Wiskott-Aldrich syndrome protein (N-WASP) following binding (**Figure 1-7**) (Martins et al., 2020). These signalling events terminate with the polymerisation of actin for pedestal formation, which induces inflammatory responses to the detriment of host cells (Martins et al., 2020).

Additional LEE encoded effector proteins include: EspF, EspG, MAP, EspH and EspZ (Dahan et al., 2005). Both MAP and EspG have been thought to be dispensable for colonisation (Warr et al., 2019), whilst EspF, EspH and EspZ have roles in intestinal barrier disruption, filopodia repression (Dahan et al., 2005) and effector translocation (Frankel et al., 2012), respectively.

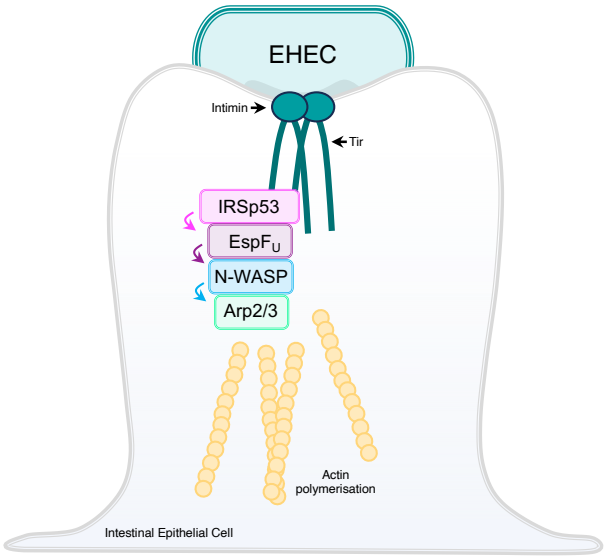


Figure 1-7. Pedestal formation on host intestinal epithelial cells by EHEC. Following the initial attachment of EHEC to host cells via HCP, the T3SS is expressed and effectors such as Tir are secreted into the target cell. Tir is responsible for the intimate attachment to host cells through interaction with intimin, expressed on the bacterial cell surface. Intracellularly, Tir interacts with host insulin receptor tyrosine kinase substrate (IRSp53) precedent to the recruitment of the NLE effector EspFU, also secreted into the cell. EspFU then activates N-WASP, which via the Actin Related Protein 2/3 (Arp2/3) complex, induces the polymerisation of host cell actin and formation of A/E lesions on the host cell surface.

1.4.4. Non-LEE encoded (NLE) effectors

As alluded to by EspFu, not all effectors of the T3SS are encoded by the LEE. Instead, NLE effectors are carried within lambdoid prophages distributed throughout the genome (Tobe et al., 2006). The role of these NLE effectors is largely immunoregulatory, acting to promote inflammatory responses and apoptotic effects within host cells (Dean & Kenny, 2009). The function of NLE effectors has been reviewed in more detail by Wong et al., (2011). Acquisition of these effectors once again highlights the highly adaptive and plastic nature of the EHEC genome.

The number of NLE effectors secreted via the T3SS greatly exceeds the number encoded by the LEE and account for more than 2/3 of EHECs known effector repertoire (**Figure 1-8**). What is noticeable between A/E pathogens (e.g., EPEC, *C. rodentium*) is that the range of NLEs is variable, with some being found to be dispensable for infection (i.e., NleC-F) (Sanchez-Garrido et al., 2022). For example, Ruano-Gallego *et al.*, (2021) demonstrated that the effector repertoire of *C. rodentium* can undergo a 60

% contraction and remain similarly pathogenic. This study also identified a further two novel NLE effectors, NleN and NleO, adding to the increasing number of effectors present in the genomes of A/E pathogens (Ruano-Gallego et al., 2021). The caveat to the identification of these effectors is that the majority remain unstudied *in vivo*, attributable to EHEC being human restricted (Sanchez-Garrido et al., 2022). Similarly, the advances in predictive *in silico* methodologies has meant that the identification of novel effectors has surpassed effector characterisation (Slater & Frankel, 2020). Trying to determine a core of NLE effectors essential for pathogenesis is therefore extremely complex due to the lack of functional data available, as well as the varying importance they have under different environments, termed context-dependent effector essentiality (Ruano-Gallego et al., 2021; Sanchez-Garrido et al., 2022).

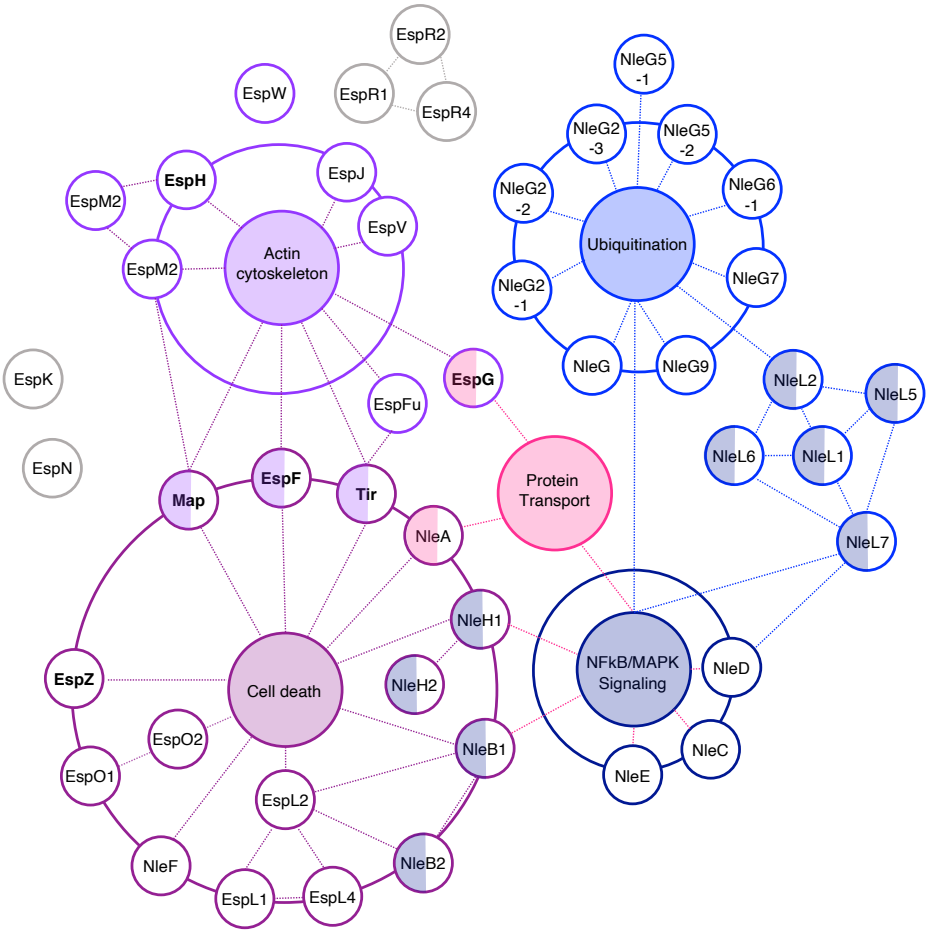


Figure 1-8. Network overview of T3SS effectors expressed in EHEC. The T3SS effector repertoire targets several common host cellular pathways: cell death, ubiquitination, protein transport, NFkB/MAPK signalling, and actin cytoskeleton organisation. Most T3SS effectors are not encoded by the LEE. Those effectors encoded by the LEE are shown in bold. Effectors with overlapping functions are also filled with the colour corresponding to the cellular pathways with which they are associated with. For example, EspG is involved in actin cytoskeleton organisation (purple border) and protein transport (pink fill). Effectors with no colour assigned (grey) are those with unknown functions. Interactions between effectors are denoted by the coloured edges. Figure modified from Sanchez-Garrido *et al.*, (2021).

1.4.5. Shiga toxin (Stx) production

Stx (also known as Verocytotoxin) are potent cytotoxins and key VFs of EHEC encoded within lysogenic lambdoid phages integrated into their genome (Yara et al., 2020; Zhang et al., 2000). Expression and release of Stx in the human colon results from activation of the SOS response to DNA damage, resulting in phage-mediated lysis of the bacterial cell envelope (Zhang et al., 2000). Consequently, antibiotics have been discouraged in the treatment of EHEC infections due to their association with SOS induction (Kakoullis et al., 2019).

The effects of Stx follow their binding to globotriaosylceramide-3 (Gb_3) expressed on host endothelial cells via the B_5 domain of their AB_5 structure (Fraser et al., 2004). Gb_3 expression therefore determines the tropism of Stx pathophysiology, with variable expression observed across animals (Pruimboom-Brees et al., 2000). Subsequently, Gb_3 has been the target of several potential therapeutics in the treatment of EHEC infection (Goldwater & Bettelheim, 2012). However, in the human colon, the mechanism used to cross the epithelial barrier remains unknown due to an absence of Gb_3 (Schüller, 2011). Cattle on the other hand display a complete lack of vasculature Gb_3 expression, hence them being asymptomatic carriers of EHEC (Pruimboom-Brees et al., 2000). Upon breaching the epithelial barrier of humans, Stx enter the circulation for dissemination to extra-intestinal sites, where there is endothelial expression of Gb_3, such as the microvasculature of the kidneys (Kaper et al., 2004) (**Figure 1-9**). Stx-receptor binding in the kidney drives significant changes to cellular processes such as protein synthesis and apoptosis (Yara et al., 2020) (**Figure 1-9**). Dysregulation of the host cell results in HUS, culminating as renal failure.

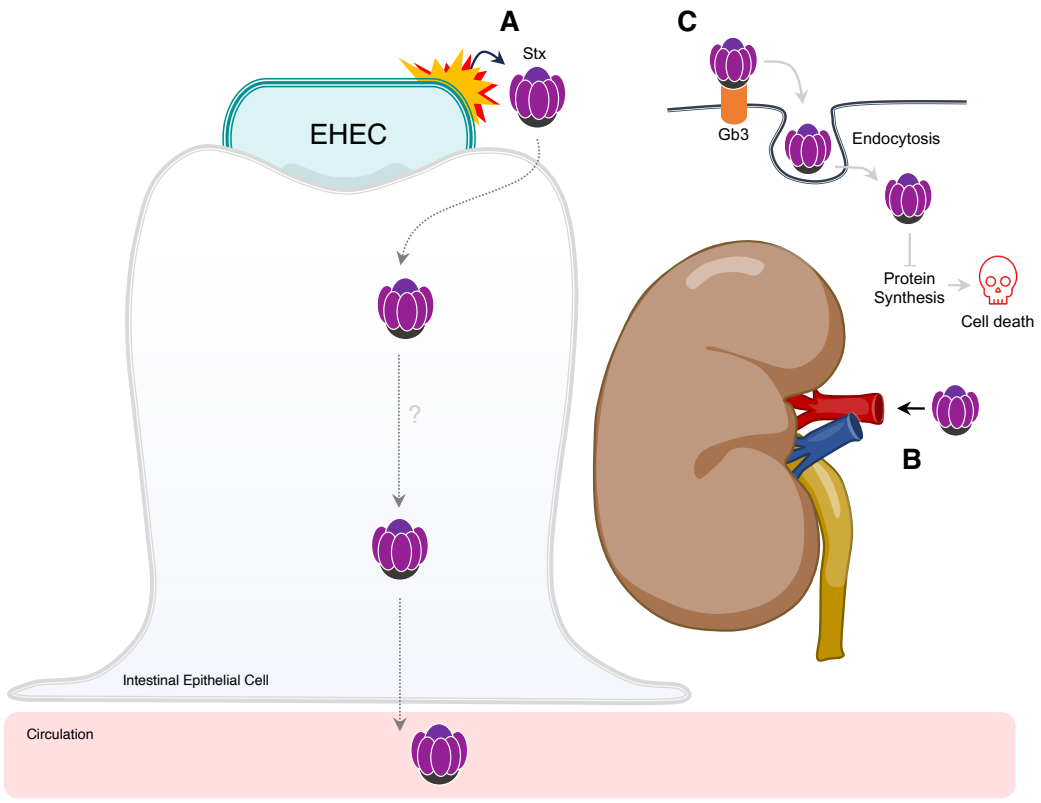


Figure 1-9. Overview of the Stx mechanism of action in human hosts during EHEC infection. (**A**) Stx is released into the lumen of the colon following lambdoid phage-mediated lysis of EHEC. Stx then breaches the intestinal epithelial barrier and enters the circulation, where it is disseminated to extraintestinal sites. (**B**) Extraintestinal sites include the kidney where the Stx-specific receptor (Gb_3) is highly expressed. (**C**) In the vasculature of the kidney, Stx binds Gb_3 expressed on the surface of endothelial cells, resulting in clathrin-mediated endocytosis and changes to cellular processes such as the inhibition of protein synthesis, leading to cell death.

Two major types of Stx are produced amongst STEC: Stx1 and Stx2. Stx1 comprises of four subtypes (Stx1a-e) and Stx2 comprises of seven subtypes (Stx2a-g) (Mühlen & Dersch, 2020). Evolutionarily, Stx2 is thought to have been acquired first via transduction, later followed by Stx1 (Wick et al., 2005). Distinct variants of Stx2 have been observed over time, typically defined by their biological activity and impact on clinical outcome (Yara et al., 2020). For example, STEC that possess Stx2a either independently or in concert with additional variants, significantly enhance development of HUS (Yara et al., 2020). Additionally, other Stx2 subtypes, Stx2c and Stx2d, have been described as most relevant in human infection (Joseph et al., 2020). In contrast, Stx2e and Stx2f occurrence in human disease happens almost never (Joseph et al.,

2020). Subsequently, there is a clear association between subtypes and host, which might reflect the tropism of these toxins. It is important to note that not all strains capable of producing Stx are EHEC, but rather EHEC define Stx positive strains also carrying the LEE (Kaper et al., 2004). Hence, EHEC being a subset of STEC.

1.4.6. pO157 plasmid

Though suggested to be less crucial for virulence in EHEC than the T3SS and Stx (Pilla & Tang, 2018), the carriage of a large F-like plasmid (pO157) is observed amongst all O157:H7 strains. Following its sequencing in 1998, pO157 was found to comprise 100 ORFs of the same directionality and have an overall size of ~93 Kb (Burland et al., 1998). Significantly, a number of these genes were predicted to encode VFs including an extracellular serine protease (EspP), catalase peroxidase (KatP), haemolysin (HlyA), metalloprotease (StcE) and a Type 2 Secretion System (T2SS; EtpC-O) (Burland et al., 1998; Lim et al., 2010). The identification of such genes have supported previous claims of pO157 being important for haemolytic activity and the adherence to host cells (Burland et al., 1998). Further, a study whereby the O157:H7 strain, Sakai, was cured of pO157, displayed significantly reduced microcolony formation on Caco-2 cells compared to its non-cured counterpart (Tatsuno et al., 2001). Introduction of a minimal pO157 construct later attributed this loss of adherence to the absence of ToxB (Tatsuno et al., 2001), a predicted adhesin (Kaper et al., 2004). However, the biological significance of pO157, and its role in pathogenicity, is not clearly defined as conflicting results have been obtained *in vivo*, across different animal hosts (Lim et al., 2010).

1.5. Regulation of virulence in EHEC

1.5.1. Master regulation of the LEE

The expression of the LEE in EHEC must be tightly controlled to allow for coordinated expression of its components. Regulatory mechanisms that control LEE expression, whether that is to activate or repress, are underpinned by a complex network of interacting signals and transcription factors. At its core, LEE regulation is driven by an interplay between the LEE-encoded regulator, Ler, and Ler activation regulator, GrlA (Connolly et al., 2015) (**Figure 1-10**). However, it is important to highlight that the regulation of the LEE is not exclusively driven by these regulators, with a diverse range of environmental factors such as temperature, oxygen, pH, osmolarity and nutrients

each having their own impact on LEE expression during host colonisation (Sharma et al., 2005). Therefore, due to binding via the T3SS not being dependent upon a tissue-receptor interaction, environmental signals act as the primary trigger signalling expression of the T3SS and host-cell attachment.

As previously mentioned, Ler is a 15 kDa protein encoded within LEE1 (Bustamante et al., 2011) and acts to positively control the expression of LEE1-5 as part of a regulatory cascade (Elliott et al., 2000). In addition to its largely positive regulatory role, Ler also negatively autoregulates the LEE1 promoter (Berdichevsky et al., 2005). Specifically, Ler has been shown to bind upstream of the LEE1 operon under both *in vitro* and *in vivo* conditions, with an affinity that is sufficient to optimally activate LEE expression (Berdichevsky et al., 2005). Within this region there are said to be two promoter sites: P1 (distal) and P2 (proximal). The designated importance of these sites in the activation of the LEE in not entirely clear, as some groups suggest P1 to be the major promoter, whilst others suggest P2 to have the bigger role (Islam et al., 2011). A mechanism of autoregulation and steady-state kinetics exemplified by Ler highlights the sophisticated means by which LEE expression is maintained and balanced accordingly during infection. Its crucial importance in virulence expression and central position in the hierarchy of LEE expression has led to Ler being recognised as the 'master' regulator of the LEE (Islam et al., 2011). It is important to note that Ler has also been shown to regulate several non-LEE encoded genes such as *espC*, *tagA*, *stcE* and *nleA* (Bustamante et al., 2011). Studies have demonstrated the crucial role of Ler for virulence phenotypically, whereby its deletion (Δ*ler*) impaired effector secretion and A/E lesion formation (Elliott et al., 2000). More recently, a Δ*ler* mutant in *C. rodentium* was shown to be avirulent during murine infection *in vivo* (Connolly et al., 2018).

To exert its regulatory effect on the LEE, Ler acts as an antagonist to nucleoid-associated proteins (NAP), specifically the histone-like nucleoid structuring protein (H-NS) (Lara-Ochoa et al., 2023). H-NS is a global regulator of transcription that displays a bias for binding AT-rich regions of the chromosome and as such acts as a xenogeneic silencer of foreign DNA (Lara-Ochoa et al., 2023; Singh & Grainger, 2013). For example, H-NS can bind to and regulate > 900 genes in EHEC alone (Wan et al., 2016). When bound, H-NS forms DNA-bridges that ultimately block, or trap RNA

polymerase (RNAP) required for the transcription of these target genes (Rangarajan & Schnetz, 2018) (**Figure 1-10A**). Subsequently, expression of the H-NS targeted gene is said to have been 'silenced'. Ler is a member of the H-NS family of NAPs (Lara-Ochoa et al., 2023) and alleviates the repressive effects of H-NS by displacing the protein pre-bound at promoter sites across the LEE due to the regulator having a stronger binding constant (Shin, 2017). In addition to Ler, H-NS binding can also be displaced by the small transcriptional regulator, Pch, which is required for maximal LEE expression in EHEC (Fukui et al., 2016) (**Figure 1-10B**). A plasmid-encoded *pch* homologue (Plasmid-encoded regulator; *per*) in close relative EPEC has also been demonstrated to be required for full LEE activation (Bustamante et al., 2011).

1.5.2. GrlRA regulation of the LEE

Encoded within a bicistronic operon (*grlRA*) located between LEE1 and LEE2, GrlA is integral for the full activation of the LEE due to its role in promoting *ler* expression (Alsharif et al., 2015; Islam et al., 2011). Then, with Ler acting as a positive regulator of GrlA, the two proteins form a positive feedback loop that maintains balance between *ler* autoregulation and LEE activation (Barba et al., 2005). However, as with any regulator, GrlA activity must also be controlled to sustain an equilibrium between activation and repression of the LEE. GrlA activity is modulated by the global regulator of *ler* repression (GrlR) (**Figure 1-10**) via interaction with the helix-turn-helix (HTH) DNA binding motif located at the N-terminus of GrlA (Lara-Ochoa et al., 2023). The introduction of point mutations in the HTH motif interferes with GrlA function and its interaction with GrlR (Lara-Ochoa et al., 2023). To exert its effect, dimeric GrlR physically represses GrlA by directly binding the protein such that it is no longer able to bind regulatory regions of DNA for *ler* activation (Padavannil et al., 2013) (**Figure 1-10**). Above all, GrlA and GrlR are co-expressed, so how GrlA evades suppression by GrlR has been questioned. In EHEC, the protease ClpXP has been reported to degrade GrlR, thereby freeing GrlA during the stationary phase of growth (Lara-Ochoa et al., 2023) (**Figure 1-10**). In support, deletion of either ClpP or ClpX in EHEC abolished the secretion of key effector proteins, whilst GrlR levels increase (Iyoda & Watanabe, 2005).

The de-repression of GrlA following ClpXP degradation of GrlR acts as a pre-requisite to complete virulence gene transcription by GrlA. This is due to the required

mechanical stimulation of GrlA (Sirisaengtaksin et al., 2020). In its free but inactive state, GrlA is membrane-associated and re-locates to the cytoplasm following mechano-sensing of stimuli such as attachment to host cells and fluid sheer in the intestinal environment (Alsharif et al., 2015). Similar phenomena have been observed in *Pseudomonas* spp. (Persat et al., 2015), however, the exact mechanism in EHEC remains unclear (Sirisaengtaksin et al., 2020). GrlA has also been demonstrated to negatively regulate flagella expression via *flhDC*, whilst positively regulating the haemolytic capacity of EHEC via the *ehxCABD* operon (Platenkamp & Mellies, 2018).

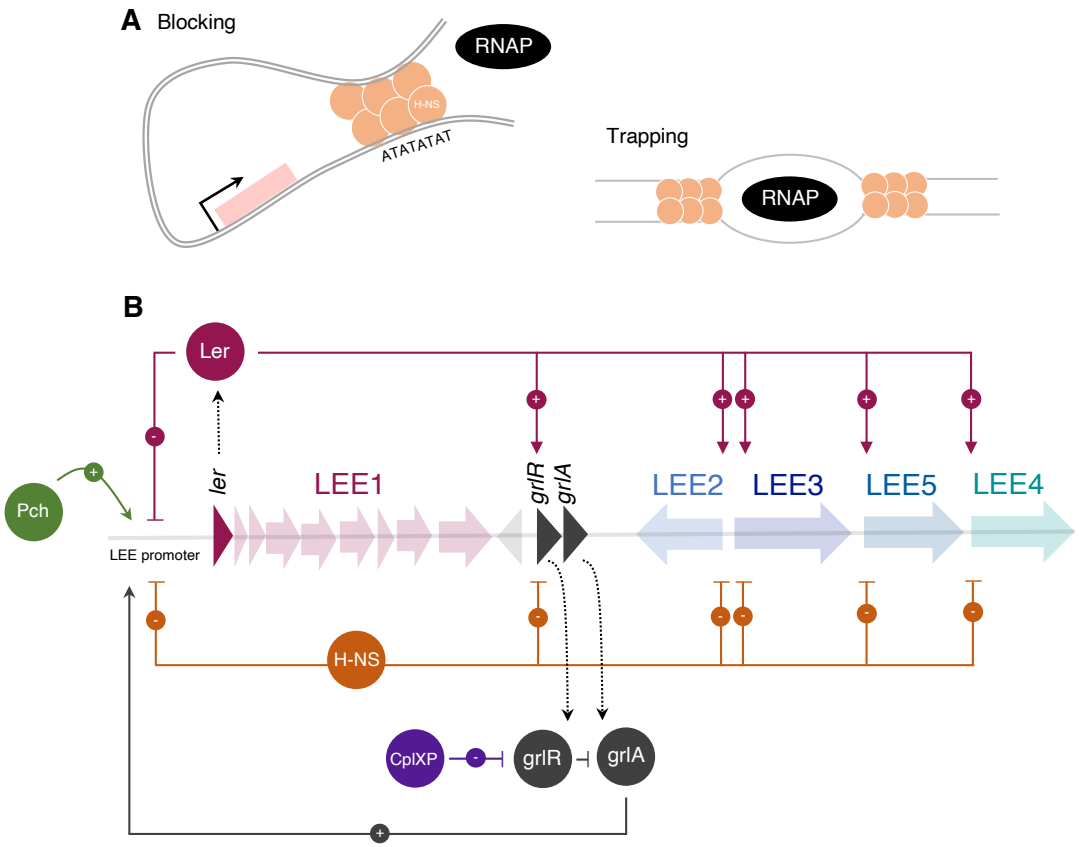


Figure 1-10. Master regulation of the LEE in EHEC. Under unfavourable conditions expression of the LEE is repressed by H-NS binding to AT-rich regions of DNA. (**A**) H-NS can either block RNAP access to target promoter sites through DNA looping or alternatively trap RNAP itself. (**B**) Under favourable conditions, the master regulator of the LEE, Ler, acts to positively regulate the expression of LEE2-5 and GrlRA by antagonising H-NS and displacing the protein. Ler also negatively autoregulates its own expression. GrlA acts to also regulate LEE expression positively through binding to the *ler* regulatory region. GrlA activity is negatively regulated through GrlR. Repression of *grlA* by GrlR is alleviated by ClpXP which degrades GrlR. The extrinsic regulatory factor Pch is required for the full activation of the LEE through direct binding to the *ler* regulatory region. Dotted lines represent translation. Transcriptional activation and repression are indicated by (+) and (-) respectively.

1.5.3. Other core regulatory mechanisms of the LEE

Aside from the key intrinsic regulators (Ler, GrlRA) of the LEE, and the associated regulators described above, regulators encoded elsewhere in the genome also have important roles in regulating LEE expression. These regulators similarly work by targeting regulatory regions of LEE1 as way of affecting *ler* expression. For example, the global regulator of virulence A (GrvA) activates and enhances LEE expression alongside the secondary regulator, RcsB, part of a phosphorelay system (RcsDCB) (Morgan et al., 2016) (**Figure 1-11**). The upregulation of LEE expression conferred by GrvA occurs following the downregulation of *gadE*, a known repressor of LEE1 via *ler* (Morgan et al., 2016). Characterised in O157:H7 strain Sakai, GadE is responsible for protecting the cell against low pH by activating the expression of the glutamate-dependent (GAD) acid-resistance system (Vanaja et al., 2009). On the other hand, additional regulators, prophage-encoded secretion regulator PsrA and PsrB negatively affect LEE expression by promoting GadE expression (Jiang et al., 2021) (**Figure 1-11**). For these reasons, GadE constitutes a core regulatory component in the control of LEE expression. Further details on the role of GadE and other components of the GAD system in the direct and indirect regulation of the LEE have been described in a review by Gelelcha *et al*., (2022).

Other notable transcriptional regulators described in the literature to have influence over LEE expression include hemolysin expression modulating protein (Hha), integration host factor (IHF), EtrA, and EivF (**Figure 1-11**). Hha, a member of the Hha family of NAPs (Sharma & Casey, 2014), is responsible for repressing expression of *ler* and therefore expression of the LEE, with Δ*hha* strains displaying enhanced adherence to Hep-2 cells (Sharma et al., 2005). EtrA and EivF are also capable of downregulating LEE expression in EHEC (**Figure 1-11**). Both these regulators belong to a cryptic non-functional secondary T3SS locus (ETT2) identified in EHEC (Luzader et al., 2016). It is by the activity of a third regulator encoded by ETT2 (EtrB), that the activity of these repressors is suppressed (Luzader et al., 2016). Subsequently, EtrB can indirectly and directly (via interaction with the *ler* promoter) increase LEE expression. Similar to H-NS and Hha, IHF belongs to a family of NAPs but instead positively regulates the LEE by aiding the displacement of H-NS within the *ler* regulatory region (Connolly et al., 2015). Collectively, transcriptional regulation of the LEE is inherently complex and encompasses many other regulatory proteins which will

not be discussed here but include additional NAPs such as Fis and sigma factors, RpoS and RpoN (Gelalcha et al., 2022; Platenkamp & Mellies, 2018).

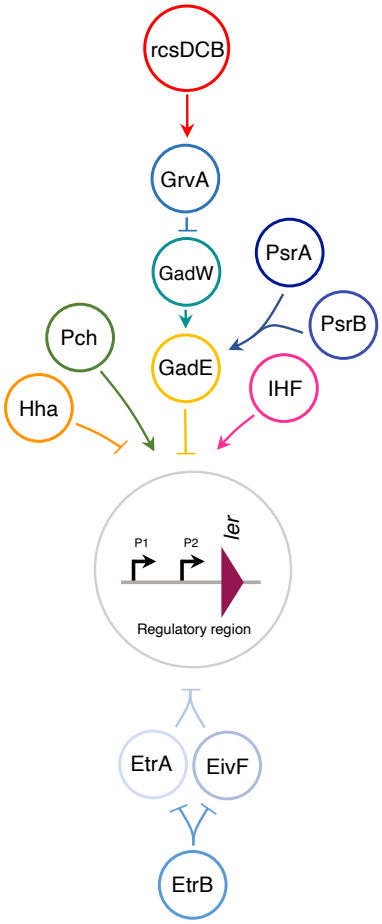


Figure 1-11. Overview of LEE regulation by known global transcriptional regulators. Regulatory network showing the direct and indirect interactions of various transcriptional regulators with the *ler* regulatory region, defined as the LEE promoter. Pointed and flat arrows represent transcriptional activation and repression, respectively.

1.5.4. Post-transcriptional control of the LEE

Most known regulatory processes that control LEE expression in EHEC occur at the transcriptional level (Gelalcha et al., 2022). In recent years, attention has been given to molecules that oversee post-transcriptional regulation of the LEE (**Figure 1-12**). Subsequently, there has been increased identification and understanding of small regulatory RNAs (sRNA) and RNA chaperones (Bhatt et al., 2017). RNA chaperones act in conjunction with sRNAs to aid their complementary base-pairing with mRNA targets, typically in the 5′ untranslated region (Bhatt et al., 2011; Hansen & Kaper, 2009). An example of chaperone-mediated control of gene expression is the

repression of *hns* by the chaperone, Hfq, and the sRNA, DsrA, which together decrease the stability of *hns* mRNA (Hansen & Kaper, 2009; Lease et al., 2004).

In EHEC str. EDL933, Hfq is able to temporally regulate LEE expression via *ler* through two independent mechanisms (**Figure 1-12**) based upon the phase of growth (Hansen & Kaper, 2009). During exponential growth, Hfq blocks *ler* expression via GrlA by destabilising the *grlRA* transcript (Hansen & Kaper, 2009). As a result, *grlA* mRNA was found to be more stable in a Δ*hfq* mutant compared to the WT (Hansen & Kaper, 2009). During stationary phase, Hfq instead represses *ler* expression independent of GrlRA (Hansen & Kaper, 2009). Interestingly, however, there does appear to be strain-specific differences in the effect of Hfq on LEE expression. For example, in EHEC str. 86-24, *ler* expression is instead enhanced by the chaperone (Kendall et al., 2011). Work also undertaken in 86-24 has demonstrated post-transcriptional regulation by Hfq-dependent sRNAs, GlmY and GlmZ, which have been shown to negatively regulate LEE4 and LEE5 (**Figure 1-12**) by also destabilising transcripts (Gruber & Sperandio, 2015).

More recently, a greater number of sRNAs have been identified and localised to various OIs interspersed throughout the EDL933 genome, including OI-43 (Esr41), OI-93 (Esr055) and the LEE (Arl, sRNA350) (Jiang et al., 2021). These findings support the idea that OIs are not just randomly associated with pathogenic *E. coli* but have functional roles in the regulation of virulence. Ultimately, post-transcriptional regulation via sRNAs and associated small chaperone proteins in EHEC adds to the highly plastic yet coordinated regulatory response to the environment (Bhatt et al., 2011). This is especially advantageous as these are much faster, and less energetically demanding to synthesise than transcription factors (Bhatt et al., 2011) which depend upon translation and associated proofreading processes. Spatially, post-transcriptional regulation has been demonstrated by the RNA-binding protein, CsrA, which through its antagonism by the T3SS chaperone, CesT, controls effector secretion in response to host cell attachment in A/E pathogens (**Figure 1-12**) (Katsowich et al., 2017).

1.5.5. Post-translational control of the LEE

As with post-transcriptional regulation, little is known about the post-translational control of virulence expression. Post-translational regulation typically involves protein-

protein interactions with the goal of modifying the activity of or the amount of protein in cells (Sauder & Kendall, 2018). In contrast to post-transcriptional regulation, which is energetically cheap, the machinery required to modify the abundance of intracellular protein is expensive but does have the added benefit of being extremely rapid, robust, and even sometimes reversible (Sauder & Kendall, 2018). The best-known example of post-translational regulation of virulence in EHEC, is the increased transcription of *ler* by ClpXP (Bhatt et al., 2011) (**Figure 1-12**), as discussed previously in *section 1.5.2*.

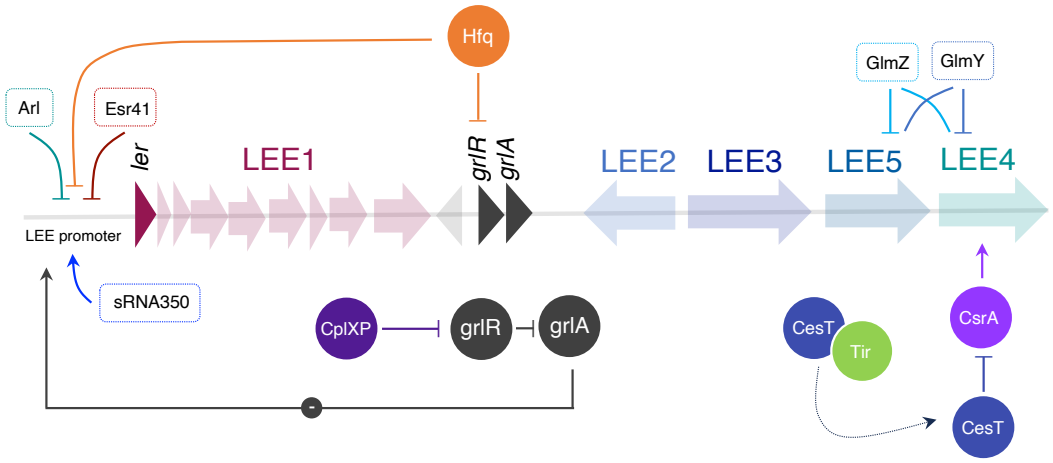


Figure 1-12. Overview of post-transcriptional and post-translational control of LEE expression. Expression of the LEE is extensively regulated at the post-transcriptional level by sRNAs, highlighted in the coloured boxes. Hfq regulates *ler* expression directly and indirectly through GrlRA repression. Arl, Esr41 and sRNA350 regulate the LEE via the *ler* regulatory region, whilst GlmZY affects LEE4-5 transcript stability. The T3SS chaperone, CesT, also post-transcriptionally regulates expression of LEE4 by antagonising CsrA, when no longer bound with T3SS effectors such as Tir. Pointed and flat arrows represent activation and repression, respectively.

1.6. Environmental cues in the regulation of virulence

1.6.1. Chemical sensing

Whilst the control of virulence expression at the single-cell level is essential for coordinating a fine-tuned response to the environment, gene regulation also occurs at the population level via quorum sensing (QS). QS is a form of cell-to-cell communication that relies on the exchange of chemical signals (i.e., small diffusible molecules) between bacteria at both the intra- and interspecies level (Oliveira et al., 2023).

In EHEC, several small diffusible chemical molecules, otherwise known as autoinducers (AI), have been found to act as key mediators in regulating fundamental aspects of virulence (**Figure 1-13**). Three main classes of AI (1-3) have been identified and reported to play a role in the virulence of *E. coli* (Gatsios et al., 2021). Both AI-1 and AI-2 are involved in interbacterial communication. However, unlike AI-2, *E. coli* is unable to self-synthesise AI-1 (Acyl-homoserine lactone; AHL) but can sense the molecule via the AI-1 receptor, SdiA (Gatsios et al., 2021). It is via SdiA-AHL signalling that LEE gene transcription is repressed in EHEC, whilst increasing tolerance to acid stress via the Gad system (Hughes et al., 2010). Conversely, *E. coli* is able to synthesise AI-2 (via LuxS) which can be utilised via the *lsrACDGFGE* and divergent *lsrRK* operon (Gatsios et al., 2021). AI-2 acts to promote motility and biofilm formation in EHEC, as well as increase the expression of virulence genes and attachment to host IECs (Bansal et al., 2008). How AI-2 drives these changes is not entirely understood and the outcome on LEE expression following exposure to AI-2 is sometimes conflicted, with some data to suggest downregulation of the LEE. Though, this is possibly attributable to strain specific differences. AI-3 is produced by the intestinal microbiota, but its effect on virulence regulation is better understood. Sensed by a two-component system (TCS) comprising a sensor histidine kinase (HK; QseC) and response regulator (RR; QseB), AI-3 is integrated into a regulatory network responsible for regulating the LEE (Sperandio et al., 2003) (**Figure 1-13**). These TCS have also evolved to sense signals of host origin, such as host-produced neurotransmitters, epinephrine (Epi) and norepinephrine (NEpi) (Hughes et al., 2009). Epi is synthesised in the adrenal medulla and reaches the gut via the blood, whilst NEpi is synthesised locally within the enteric nervous system (Hughes et al., 2009).

In addition to QseCB, a second TCS (QseEF) has been identified to sense Epi but not NEpi (at least directly) and AI-3. This system appears to be exclusive to enteric bacteria compared to QseCB, which is more widespread across the bacterial kingdom (Hughes et al., 2009). As outlined in **Figure 1-13**, the two HKs (QseC, QseE) sit in the IM and are auto phosphorylated in response to distinct signalling molecules prior to the transfer of its phosphate to the RR (QseB, QseF), which activates its regulatory activity (Njoroge & Sperandio, 2012). Crosstalk between the two systems exists as *qseE* expression is activated by QseC. Both QseC and QseE enhance LEE expression via QseF in a *ler*-dependent manner (Hughes et al., 2009), as well as increase *espFu*

transcription needed for actin polymerisation within host cells (Reading et al., 2007). Further, QseC induces the expression of a third RR, KdpE, which has been previously demonstrated to increase LEE transcription in a similarly *ler*-dependent manner (Njoroge et al., 2012). Whilst not directly regulating virulence via the LEE, QseB is responsible for antagonistically regulating the expression of the flagella regulon via *flhDC* (Hughes et al., 2009). More recently, the host-derived endocannabinoid, 2-arachidonoyl glycerol (2-AG), has been demonstrated to modulate *C. rodentium* infection by antagonising pro-virulence QseC, and therefore downregulate LEE expression (Ellermann et al., 2020). These examples demonstrate EHEC to have the ability to sense and communicate not only with each other but also with the host, important for regulating their virulence programme.

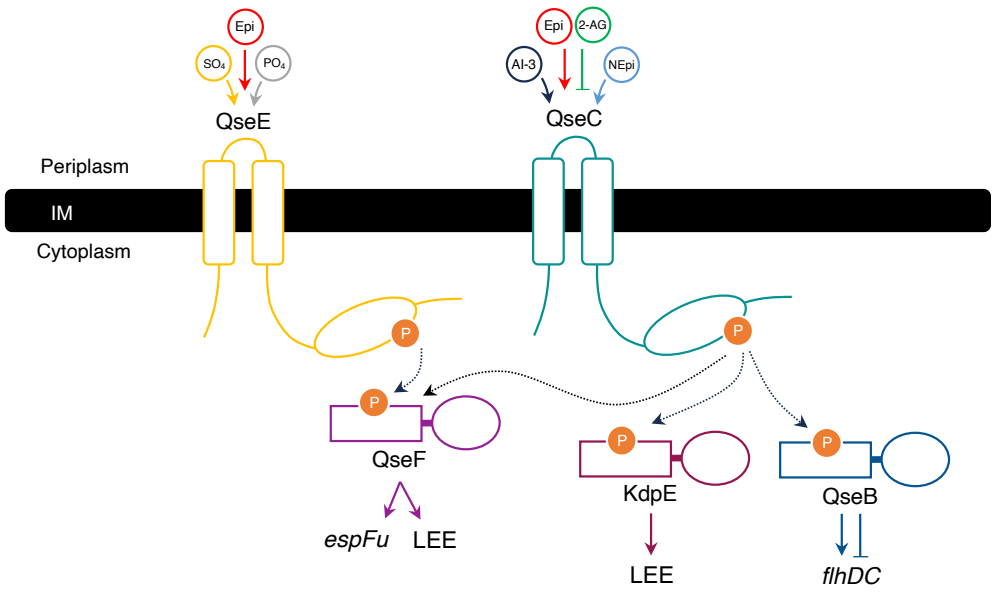


Figure 1-13. Chemical sensing in EHEC. Both bacterial- (AI-3) and host-derived (Epi, NEpi, 2-AG) chemicals are sensed via TCS. The two HKs, QseE and QseC, in the IM, are autophosphorylated on their cytoplasmic domains in response to distinct chemical signals. The phosphate is then transferred to their cognate RR in the cytoplasm of the cell: QseF and QseB. QseC is also able to interact with KdpE and phosphorylate the regulator. Phosphorylation of these regulators activates their regulatory activity such that QseF goes on to activate expression of the LEE and espFu, KdpE activates LEE expression, and QseB exerts control over flagella expression via the master regulator *flhDC*. Dotted lines represent the phosphorylation event between HK and RR. Pointed and flat arrows represent transcriptional activation and repression, respectively.

1.6.2. Nutrient sensing

The intestinal environment of the host is a highly complex and dynamic metabolic landscape, displaying exceptional temporal and spatial heterogeneity (Liang & Vallance, 2021). Within this ecosystem, the diversity of nutrients present can be derived from the diet, the host, and the gut microbiota. Notably, these factors can depend on one another, with host diet greatly impacting the bacterial composition of the intestine (Hernandez-Doria & Sperandio, 2013). Further, the abundance of nutrients is heavily dependent upon geographical location along the GI axis. For example, differential expression of mucins (glycoprotein constituent of mucus) along the gut axis have been described to act as 'GI tract signposts' (Carlson-Banning & Sperandio, 2016). Thus, the presence or absence of nutrients is informative of biogeographical location and environmental suitability. This is especially important in preventing energy being wasted expressing expensive VFs (i.e., the T3SS) at sites unfavourable for colonisation.

Amongst enteric pathogens such as EHEC, the ability to determine the differential concentrations of a given nutrient has been coined 'nutrient sensing'. As with chemical signals such as those described in *section 1.6.1.*, sensing involves transcriptional regulators that integrate environmental cues into complex regulatory networks, required for the spatio-temporal expression of virulence.

Fucose signalling represents one of the best understood examples of nutrient sensing by EHEC for the co-ordination of metabolism and LEE expression. Fucose is sensed via a TCS (FusKR), whereby FusK and FusR are the HK and RR, respectively (Pacheco et al., 2012) (**Figure 1-14A**). Following phosphorylation by FusK, active FusR represses LEE expression, as well as the genes necessary for the utilisation of fucose (*fuc* operon) (Pacheco et al., 2012). The downregulation of *fuc* genes is advantageous to EHEC as it prevents competition for fucose with commensal *E. coli* residing in the mucus layer that preferentially use the sugar (Hernandez-Doria & Sperandio, 2013). Upon breaching the mucus layer, at the epithelial interface, competition is substantially reduced and the mechanism of LEE and *fuc* gene repression is reversed (Pacheco et al., 2012). Differential nutrient preference has been described extensively in the literature, and it is known that pathogenic *E. coli* opt to exploit nutrient sources not favourably used by commensals such as mannose, ribose,

galactose, hexuronates (**Figure 1-14B**) and ethanolamine (Fabich et al., 2008; Kendall et al., 2012). This phenomenon is especially important for establishing a niche in the highly competitive gut environment and for overcoming colonisation resistance (CR) (See *section 1.7.2.*).

Galacturonate is a sugar acid present in the human intestine recently shown to act as an environmental cue in the regulation of virulence (Jimenez et al., 2019). Galacturonate is sensed by the transcriptional regulator ExuR, which during early infection promotes pathogen proliferation by upregulating genes for its metabolism (Jimenez et al., 2019). As infection proceeds, ExuR later transitions to a regulator of virulence by driving expression of the LEE by directly binding to the *ler* regulatory region (Jimenez et al., 2019) (**Figure 1-14A**). Subsequently, mice infected with Δ*exuR* display absence of histopathological damage and reduced faecal shedding compared to those infected with WT *C. rodentium* (Jimenez et al., 2019). ExuR therefore displays a bifunctionality which is dependent on the stage of infection. Similar mechanisms of nutrient sensing have also been observed for amino acids. For instance, L-arginine exerts regulatory control over the LEE via ArgR by again binding directly to the *ler* regulatory region to instead promote LEE expression (Menezes-Garcia et al., 2020) (**Figure 1-14A**). In contrast, D-serine indirectly downregulates LEE expression by differentially modulating the expression of pre-existing LEE transcriptional regulators, IHF and YhiF (GAD acid stress response regulator), through transcriptional repression and activation respectively (Connolly et al., 2014) (**Figure 1-14A**).

Unlike L-arginine and D-serine, sugars such as those described above are often not freely found in the gut. Instead, these sugars are bound within complex glycan structures and must be liberated before EHEC can utilise them (Pacheco & Sperandio, 2015). This requirement arises from the absence of specialised glycoside hydrolases (GH) needed to hydrolyse the diverse array of glycosidic bonds that interlink and adjoin sugars within polysaccharides (Pacheco & Sperandio, 2015). Subsequently, EHEC and other *E. coli* (pathogenic or commensal) have forged a dependence on distinct members of the gut microbiota that encode the necessary enzymatic machinery to liberate monosaccharides (Pacheco & Sperandio, 2015). EHEC is then able to scavenge these substrates for energy and regulation of their virulence repertoire via the sensing mechanisms described, thereby exploiting the positive metabolic activity

of the microbiota. For example, galacturonate is a microbiota-derived sugar generated from the degradation of pectin present in plant material ingested as part of the diet (Jimenez et al., 2019). Similarly, fucose is released into the intestinal environment from mucus by the action of fucosidases secreted by *Bacteroides thetaiotaomicron (Bt)*, a prominent member of the gut microbiota (Hernandez-Doria & Sperandio, 2013). Subsequently, the gut microbiota can be seen to greatly influence the pathogenic fate of EHEC (See *section 1.7.*), whether that be advertent or inadvertent, is less clear.

The role of the microbiota in virulence expression has been additionally evidenced by their ability to produce fatty acids (FA) that alter LEE expression. Adding to the complexity of microbiota-pathogen interactions, some FAs act to enhance LEE transcription, whilst others repress. For example, the short chain FA (SCFA), butyrate, significantly enhances LEE expression via the leucine-responsive regulator, Lrp, through promoting the activity of Pch (Nakanishi et al., 2009) (**Figure 1-14A**). A follow up study also revealed butyrate to induce expression of the leucine biosynthesis operon regulator (LeuO) by Lrp, which is also able to activate Pch (Takao et al., 2014). Further related to the microbiota, *Bt* generates succinate as a by-product of fermentative metabolism under gluconeogenic conditions (Curtis et al., 2014). In *C. rodentium* and EHEC, succinate has been demonstrated to have a pro-virulence effect through activation of the sugar-sensitive regulator, Cra, able to bind regulatory sites upstream of *ler* (Curtis et al., 2014) (**Figure 1-14A**). The promotion of virulence has also been extended to other succinate-producing members of the gut such as *Enterococcus faecalis* (Curtis et al., 2014). However, the view that a nutrient or metabolite acts as a sole regulatory cue, important in the decision of whether to express the LEE or not is an over-simplification. The mixture of substrates simultaneously available in the gut environment represents a unique situation whereby both nutrients known to activate LEE expression are present, in addition to those that repress. This is also likely to be reflected in metabolism whereby the flux of metabolites through the pathways of central carbon metabolism (e.g., Glycolysis, pentose phosphate pathway (PPP) and tricarboxylic acid (TCA) cycle) (**Figure 1-14B**) generate a dynamic pool of substrates that confer contrasting effects on LEE expression. Similarly, this represents an opportunity for substrates known to act on the LEE to confer regulatory effects despite not being present in the extracellular environment. For example, although liberated by the gut microbiota, succinate is also generated via

the flux of glycolytic substrates into and through the TCA cycle. Consequently, the regulation of LEE expression is able to be inherently shaped by those metabolites downstream and the flux through these pathways at the system level. This therefore highlights the multifaceted nature of nutrient sensing in regulating the LEE through metabolism.

Finally, whilst the nutrient composition of the gut can largely be attributed to diet and the microbiota, it is necessary to highlight several host-derived metabolites that act as cues in regulating virulence. One such example is the enterocyte phospholipid component, ethanolamine. Following enterocyte turnover, free ethanolamine is sensed by the EutR regulator and enhances virulence expression by direct interaction with the LEE in EHEC (Kendall et al., 2012; Luzader et al., 2013) (**Figure 1-14A**). In *C. rodentium* the mechanism appears to be highly conserved *in vitro* and *in vivo* (Rowley et al., 2020), suggesting importance of the metabolite amongst A/E pathogens. Similarly related to host cell membranes, arachidonic acid is a long chain FA (LCFA) liberated from membrane associated phospholipids (Ellermann et al., 2021). Arachidonic acid in its acyl-CoA form suppresses LEE activation through binding of the LCFA regulator, FadR, reducing affinity for DNA-binding (Ellermann et al., 2021) (**Figure 1-14A**). Ordinarily, in the absence of LCFAs, FadR binds the LEE1 regulatory region to induce expression (Ellermann et al., 2021).

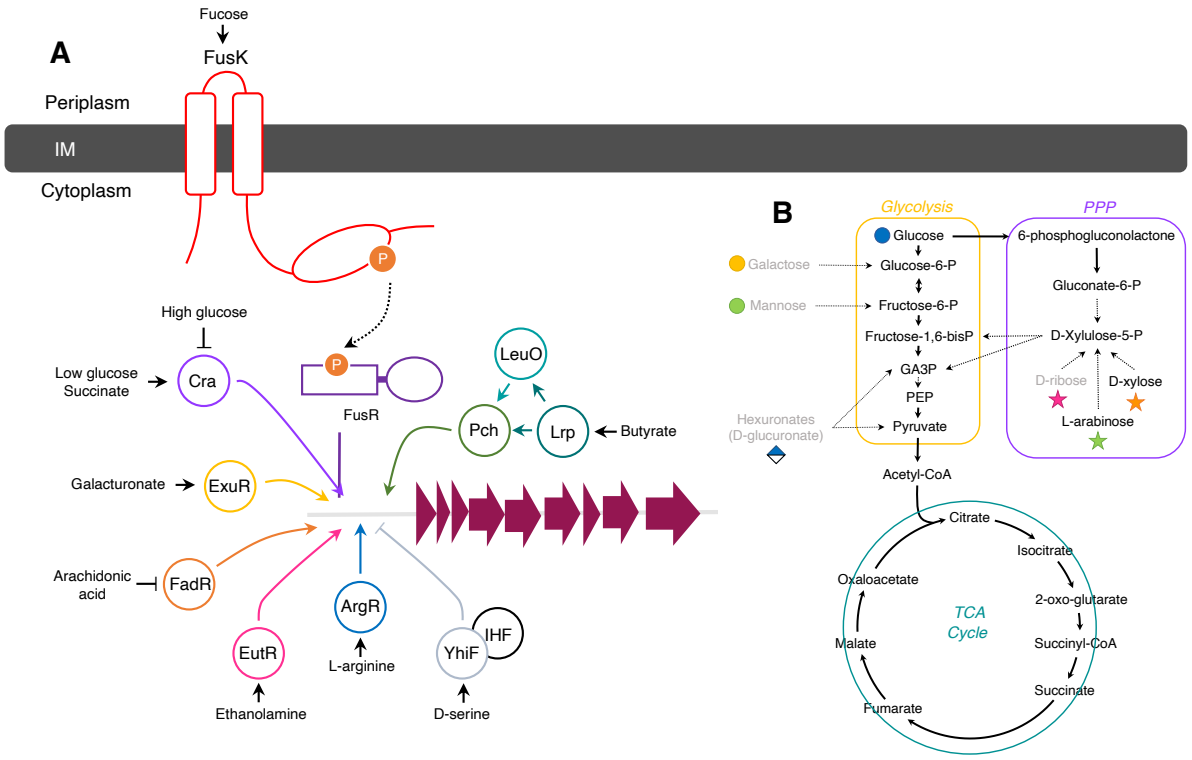


Figure 1-14. Overview of nutrient sensing in EHEC. (**A**) A plethora of differentially sourced nutrients (host-, diet- and microbiota-derived) are integrated into a complex network of transcriptional regulators that regulate LEE expression, predominantly via direct interaction with the LEE1 promoter. Typically, nutrients are sensed in the periplasm via TCS such as FusK, activating a cascade of phosphorylation events, that end with a terminal regulator capable of modifying LEE expression. Alternatively, nutrients are bound by their cognate transcriptional regulators in the cytoplasm and in turn directly or indirectly (via additional regulators) regulate LEE expression. Dotted lines represent the transfer of phosphate. Pointed and flat arrows represent transcriptional activation and repression, respectively. (**B**) Basic overview of central carbon metabolism in *E. coli*, encompassing glycolysis (yellow), pentose phosphate pathway (PPP; purple) and tricarboxylic acid cycle (TCA cycle; green). Dashed lines represent those pathways where not all substrates are shown. Sugars known to be used exclusively by EHEC and not commensal *E. coli* strains are labelled in grey and their entry point into metabolism specified.

1.7. Pathogen-microbiota interactions

1.7.1. Defining the human gut microbiota

The gut microbiota plays a fundamental role in the health and disease of humans, with disturbances associated with inflammatory bowel disease, cancer, obesity, and diabetes (Durack & Lynch, 2019). In its composition, organisms span all three domains of life (Archaea, Bacteria and Eukarya), although being more unproportionally enriched with bacteria (from 10^8 in the ileum to 10^{11} in the caecum) (Sender et al., 2016), with Bacteroidetes and Firmicutes dominating (Baümler & Sperandio, 2016). Due to its overall genetic and metabolic capacity, the gut microbiota is widely recognised as an 'organ' within itself, founded by the consortia of microbes that inhabit the GI tract (Pacheco & Sperandio, 2015). At the community level, the gut ecosystem is inherently complex, attributable to the dynamic and sophisticated network of interactions that microbes employ to communicate with each other, and with the host. Evolutionary studies have shown the forging of host-microbial symbiosis to be the outcome of co-evolution over millions of years, reflected in the stability of the human microbiome (Lee et al., 2013).

1.7.2. The microbiota as a protective barrier

An important role of the microbiota is to exclude invading pathogens from the gut environment (Shealy et al., 2021). Largely, this is achieved through the maintenance of high microbial densities and the limitation of freely available nutrients such that pathogens are unable to establish a nutrient niche in the highly competitive environment (Ng et al., 2013). The microbiota therefore acts as a biological barrier to prevent pathogen colonisation of the GI tract, in a phenomenon known as CR (Sorbara & Pamer, 2019). The term CR was first coined following work that recognised streptomycin-mediated displacement of the microbiota increased host susceptibility to infection by *Salmonella* (Bohnhoff et al., 1954). Notably, CR is recognised as an acquired trait due to the microbiota being established after birth (Litvak & Bäumler, 2019). Subsequently, pathogens have had to evolve novel strategies to overcome CR, being mainly tailored towards exploiting alternative nutrient sources not already being used by resident gut microbes (Litvak & Bäumler, 2019; Pacheco & Sperandio, 2015). The importance of nutrients in defining niches along the GI tract led to the proposal of the nutrient niche hypothesis by Rolf Freter (1983) (Freter et al., 1983). This theory

recognises that for a species to colonise they must be the most proficient at using a given nutrient compared to its competitors (Freter et al., 1983). With that, divergence in nutrient preferences and a hierarchy of nutrient utilisation has been observed (Chang et al., 2004; Fabich et al., 2008). On the contrary, there are studies such as those already described in the previous section that suggest the microbiota can indeed promote the colonisation and virulence of enteric pathogens. The relationship between microbiota and pathogens is therefore inherently complex, and intrinsically linked. Other than acting as a blockade to pathogens, the microbiota has functional implications in the initiation of protective immune responses (Willing et al., 2011), niche modification (Curtis et al., 2014) and post-transcriptional control of virulence in pathogens (Cameron et al., 2018).

1.7.3. High levels of virulence regulation by the microbiota

Recent work has highlighted the ability of the microbiota to regulate virulence at the post-translational level through proteases that target key components of the T3SS (Cameron et al., 2018). Both *E. faecalis* and *Bt* secrete proteases that aid pore formation in host IECs through cleavage of the translocon protein, EspB, required for the delivery of effectors into cells (Cameron et al., 2018, 2019). In contrast to EspP, an endogenous protease native to EHEC that limits T3SS activity, microbiota-derived proteases act to positively regulate T3SS activity and A/E lesion formation (Cameron et al., 2018). Subsequently, whilst the microbiota as a collective can be seen to be protective through mechanisms such as CR, distinct members are able to enhance the virulence of invading pathogens. It is likely that several context-dependent factors determine the nature of this relationship.

1.8. Novel nutrient transporters in EHEC infection

Whether derived from the diet, microbiota, or host, EHEC is an expert at exploiting nutrients to overcome CR, establish novel niches and critically regulate their virulence repertoire. Genome sequencing revealed EHEC to encode additional nutrient-related systems across the genome on OIs (Perna et al., 2001). These systems (typically transporters or TCS) likely provide EHEC with a competitive edge, allowing the rapid sensing or uptake of nutrients highly competed for. However, as most of these systems are putatively annotated, further effort is needed to characterise them genetically and biochemically, to confirm their function. In support of this, recent work has shown that

a novel TCS (LmvRK), ABC transporter (LmuZYX) and associated metabolic enzymes (LmuKAI) carried on OI-167 contributed to the utilisation of mannose and lyxose, as well as enhanced LEE expression (Yang et al., 2023). Deletion of *lmvR* and *lmvK* resulted in reduced adherence of EHEC to human colonoids and a competitive disadvantage in mice when compared to the WT (Yang et al., 2023). Further, instances exist where OIs are annotated to encode transporters specific for nutrients already with canonical systems, suggesting these substrates to perhaps be of importance to the pathogen.

1.9. Project aims

The importance of nutrient-related systems for virulence and infection has recently been demonstrated in the EHEC murine model pathogen, *C. rodentium*. In one study, the transcriptome of *C. rodentium* during murine infection, across two different sites (caecum and rectum), was probed to identify differentially expressed genes (DEGs) relevant for infection (Connolly et al., 2018). Though largely being virulence-specific, a subset of genes related to nutrient utilisation were amongst the most highly upregulated DEGs *in vivo*, when compared to *in vitro* cultures. These DEGs suggest that certain nutrients are available *in vivo* and signal regulation of associated genes. These included genes for 1,2-propanediol catabolism (*pduC*), ribose utilisation (*rbsD*) and glutamate/aspartate transport (*gltI*) (Connolly et al., 2018). As such, the study demonstrated that 1,2-propanediol catabolism enhanced the infective capacity of *C. rodentium* via the indirect regulation of its T3SS, acting as proof of concept that *in vivo* induced DEGs related to nutrient metabolism can play crucial roles during interaction with the host.

Additionally, several genes with only putative function were amongst these: ROD_24811, 32431, 21511 and 13781 (Connolly et al., 2018). Notably, it was only ROD_28411 that was significantly upregulated at both caecum (15.44-fold) and rectum (18.79-fold) at the peak of infection (Connolly et al., 2018). This gene was predicted to encode a periplasmic binding protein specific for D-ribose, corresponding with the upregulation of *rbsD*. These results suggest that D-ribose utilisation, and possibly ROD_24811, are likely important for *C. rodentium* and EHEC pathogenicity. Given that EHEC relies on the metabolism of simple sugars and can preferentially use particular

sugars over commensal strains, the role of sugar metabolism in EHEC formed the overarching questions of this thesis.

The aims of this project were therefore:

1. To bioinformatically characterise ROD_24811 and its EHEC homologs.
2. To elucidate the physiological significance of EHEC ROD_24811 homologs.
3. To investigate the broader implications of these systems and their substrates on virulence regulation and expression in EHEC.

2. Materials and methods

2.1. General bacterial growth

Overnight cultures were prepared by inoculating 5 mL of LB with a single bacterial colony and incubating at 37 °C shaking (200 rpm; New Brunswick Scientific shaking incubator) for 16-18 h. Antibiotics were supplemented where necessary for selection. When growth in M9 minimal media was required, cultures were centrifuged at 3,500 rpm for 10 mins following overnight growth. Supernatants were discarded and pellets were washed in the equivalent volume of 1 x phosphate buffer saline (PBS) to remove any carryover of spent LB media. This step was repeated two more times prior to a final resuspension.

2.2. Storage of bacterial strains and DNA stocks

Overnight cultures of bacterial strains were prepared. Following growth overnight, 500 μL of culture was mixed with 500 μL 50 % (v/v) glycerol (1:1) in a 2 mL cryovial tube and stored at -80 °C. Plasmids and gDNA were routinely stored at -20 °C in nuclease-free water.

2.3. Bacterial strains and plasmids

Table 2-1. Bacterial strains used in this study

Strain	Description	Source
TUV93-0	Wild type EHEC O157:H7 str. EDL933 Stx^-	Connolly inventory
TUV93-0 Δ*rbs*	TUV93-0 *rbs* knockout; Kan^R	This study
Δ*rbsACB*	TUV93-0 *rbsACB* knockout; Kan^R	This study
Δ*rbsR*	TUV93-0 *rbsR* knockout; Kan^R	This study
Δ*rbsK*	TUV93-0 *rbsK* knockout; Kan^R	This study
Δ*rbsD*	TUV93-0 *rbsD* knockout; Kan^R	This study
Δ*araC*	TUV93-0 *araC* knockout; Kan^R	This study
Δ*araBAD*	TUV93-0 *araBAD* knockout; Kan^R	This study
Δ*araFGH*	TUV93-0 *araFGH* knockout; Kan^R	This study
Δ*araE*	TUV93-0 *araE* knockout; Cm^R	This study
Δ*araE*/Δ*araFGH*	TUV93-0 *araE/FGH* knockout; Cm^R Kan^R	This study
Δ*araE*/Δ*araFGH*/ΔZ0415-19	TUV93-0 *araE/FGH*/Z0415-19 knockout; Cm^R Kan^R	This study
ΔZ0415-19	TUV93-0 Z0415-19 knockout; Kan^R	This study
Δ*pdhR*	TUV93-0 *pdhR* knockout	Roe lab
Δ*bssR*	TUV93-0 *bssR* knockout; Cm^R	This study
Δ*bssS*	TUV93-0 *bssS* knockout; Kan^R	This study
TUV93-0$^{Z0417\text{-}3xFLAG}$	TUV93-0 *Z0417* with C-terminus 3xFLAG epitope tag	This study
Sakai	Wild type EHEC O157:H7 Stx^-	Connolly inventory
Sakai Δ*araC*	Sakai *araC* knockout; Kan^R	Connolly inventory
ZAP193	Wild type EHEC O157:H7 Stx^-	Connolly inventory
ZAP193 ΔZ0415-19	ZAP193 Z0415-19 knockout; Kan^R	This study
CE10	Wild type MNEC O7:K1	Connolly inventory
CE10 ΔZ0415-19	CE10 Z0415-19 knockout; Kan^R	This study
CE10$^{RS01535\text{-}3xFLAG}$	CE10 *RS01535* with C-terminus 3xFLAG epitope tag	This study
CFT073	Wild type UPEC O6:H1:K2	Connolly inventory
MG1655	Lab strain K-12 *E. coli*	Connolly inventory
ICC169	Wild type *C. rodentium*	Connolly inventory
ICC168 Δ*rbs*	ICC168 *rbsDACBKR* knockout; Kan^R	This study

ICC169 Δ*araBAD*	ICC169 *araBAD* knockout; Kan^R	This study
ICC168 Δ*rbl*	ICC168 *rbl* knockout; Kan^R	This study
ICC168 Δ*ROD24811-41*	ICC168 *ROD24811-41* knockout; Kan^R	This study
ICC168 Δ*ROD24851*	ICC168 *ROD24851* knockout; Cm^R	This study
ICC168 Δ*ROD24861*	ICC168 *ROD24861* knockout; Kan^R	This study
ICC168 Δ*bssS*	ICC168 *bssS* knockout; Kan^R	Connolly inventory
VPI-5482	WT *B. thetaiotaomicron*	Lowe lab

Table 2-2. Plasmids used in this study

Plasmid	Description	Reference
pMK1*lux*	pBR322 with *luxCDABE* and MCS; Amp^R	Karavolos *et al.,* (2008)
pMK1*lux*-P_{LEE1}	pMK1*lux* with TUV93-0 LEE1 promoter cloned into MCS; Amp^R	This study
pMK1*lux*-P_{LEE1}	pMK1*lux* with ICC168 LEE1 promoter cloned into MCS; Amp^R	This study
pMK1*lux*-P_{Z0415}	pMK1*lux* with *Z0415* promoter cloned into MCS; Amp^R	This study
pMK1*lux*-$P_{ROD24811}$	pMK1*lux* with *ROD24811* promoter cloned into MCS; Amp^R	This study
pMK1*lux*-P_{araB}	pMK1*lux* with TUV93-0 *araB* promoter cloned into MCS; Amp^R	This study
pMK1*lux*-P_{bssS}	pMK1*lux* with TUV93-0 *bssS* promoter cloned into MCS; Amp^R	This study
pMK1*lux*-P_{bssR}	pMK1*lux* with TUV93-0 *bssR* promoter cloned into MCS; Amp^R	This study
pACYC184	p15A ori multicopy plasmid; Cm^R, Tet^R	Connolly inventory
pACYC184-*araC*	pACYC184 with TUV93-0 araC cloned into MCS; Cm^R, Tet^R	This study
pSUPROM	Cloning vector for expression under the Tat promoter; Kan^R	Jack *et al*., (2004)
pSUPROM-*araC*	pSUPROM with TUV93-0 *araC* cloned into MCS; Kan^R	This study
pSUPROM-*araE*	pSUPROM with TUV93-0 *araE* cloned into MCS; Kan^R	This study
pSUPROM-*araBADE*	pSUPROM with TUV93-0 *araBAD/araE* cloned into MCS; Kan^R	This study
pSUPROM-*araA*	pSUPROM with TUV93-0 *araA* cloned into MCS; Kan^R	This study
pSUPROM-*araBA*	pSUPROM with TUV93-0 *araBA* cloned into MCS; Kan^R	This study
pSUPROM-*Z0417*^{3xFLAG}	pSUPROM with Z0417 and C-terminal 3xFLAG tag cloned into MCS; Kan^R	This study

pSUPROM-*Z0416-17*^{3xFLAG}	pSUPROM with Z0416-17 and C-terminal 3xFLAG tag cloned into MCS; Kan^R	This study
pSUPROM-*RS01535*^{3xFLAG}	pSUPROM with RS015135 and C-terminal 3xFLAG tag cloned into MCS; Kan^R	This study
pSUPROM-*rbl*	pSUPROM with *rbl* locus cloned into MCS; Kan^R	This study
p*rpsM*:GFP	rpsM promoter translational fusion of pAJR70 to eGFP; Cm^R	Roe et al., (2003)
pKD46	LRed recombinase expressing plasmid; Amp^R; temperature sensitive	Datsenko and Wanner., (2000)
pKD3	Template plasmid for LRed mutagenesis; Cm^R	Datsenko and Wanner., (2000)
pKD4	Template plasmid for LRed mutagenesis; Kan^R	Datsenko and Wanner., (2000)
pCP20	FLP recombinase expressing plasmid; Amp^R; temperature sensitive	Datsenko and Wanner., (2000)
pDOC-F	Template plasmid for amplifying 3xFLAG cassette	D. J. Lee *et al.,* (2009)

2.4. Chemicals and molecular biology reagents

All chemicals and molecular biology reagents were purchased from Merck, Fisher Scientific and New England Biolabs (NEB) unless stated otherwise.

2.5. Growth media, solutions, and buffers

All growth media, solutions and buffers were prepared using milliQ H_2O or ddH_2O, and typically sterilised by autoclaving at 121 °C for 30 mins. Heat sensitive components and antibiotics were filter sterilised by passing through a 0.2 μM PES filter and added to media, buffers, and solutions post-autoclaving.

Table 2-3. LB media recipe (400 mL; pH 7.5)

Reagent	Amount
Tryptone	4 g
Yeast extract	2 g
NaCl	4 g
ddH_2O	400 mL

*When LB agar was required 6 g (1.5 %) of agar was added.

Table 2-4. M9 minimal media recipe (500 mL; pH 7.5)

Reagent	Amount
M9 salts (5x)	100 mL
$CaCl_2$ (1 M)	0.05 mL
$MgSO_4$	1 mL
ddH_2O	400 mL

Table 2-5. Brain heart infusion (BHI) media recipe (1 L; pH 7.2)

Reagent	Amount
BHI Broth	37 g
ddH_2O	1 L

Table 2-6. *Bacteroides* minimal media (BMM) recipe (100 mL)

Reagent	Amount
Vitamin K_3 (1 mg/ml)	0.1 mL
Vitamin B_{12} (0.01 mg/ml)	0.05 mL
$FeSO_4$ (0.4 mg/ml)	1 mL
Mineral salt solution	5 mL
Resazurin (0.25 mg/ml)	0.4 mL
NH_4SO_4	0.1 g
Na_2CO_3	0.1 g
Cysteine (Free base)	0.05 g
KPO_4 (1 M; pH 7.2)	10 mL
ddH_2O	85 ML

*0.1% hematin-histidine was added post-autoclaving

Table 2-7. Mineral salt solution for BMM (1 L)

Reagent	Amount
NaCl (300 mM)	18 g
$CaCl_2$ (3.6 mM)	0.53 g
$MgCl_2$ (4.2 mM)	0.4 g
$MnCl_2$ (1 mM)	0.2 g
$CoCl_2$ (0.84 mM)	0.2 g

Table 2-8. Super optimal broth (SOB) recipe (200 mL; pH 7.0)

Reagent	Amount
Tryptone	4 g
Yeast extract	1 g
NaCl	0.1 g
KCl (1 M)	2 mL

*1 mL $MgCl_2$ (2 M) and 4 mL (1 M) glucose was added for SOB with catabolite repression (SOC) post-autoclaving.

Table 2-9. Antibiotics used in this study

Antibiotic	Working concentration
Ampicillin (Amp)	100 μg/mL
Chloramphenicol (Cm)	20 μg/mL
Kanamycin (Kan)	50 μg/mL
Gentamycin (Gent)	200 μg/mL
Nalidixic acid (Nal)	50 μg/mL

2.6. Molecular biology techniques

2.6.1. Polymerase Chain Reaction (PCR)

All PCR primers (**Table 2-10**) were designed using Primer3 (Untergasser et al., 2012) or the NEBuilder assembly tool (https://nebuilder.neb.com/#!/). Generally, primers were designed to have a Tm of 55-65 °C except for Lambda red primers which worked best at 72 °C due to the polymerase used. Those primers used in restriction enzyme cloning were designed to contain the relevant restriction cut sites. Lyophilised primers were purchased from Life Technologies Ltd. and reconstituted with nuclease-free water to a concentration of 100 μM. For use in PCR, stocks were diluted 1:10 and stored at 4 °C.

PCR reactions were set up to have a 20 μL total volume and carried out in a BioRad thermocycler (S1000™) under conditions shown in **Table 2-11**. For PCR products to be used for Lambda red and cloning, Q5® High Fidelity (HF) 2 X Master Mix was generally used (**Table 2-12**). GoTaq Green 2 X Master Mix was used for verification of successful knockouts by colony PCR (**Table 2-13**). For colony PCR, single colonies were picked, patched on LB agar (supplemented with the appropriate antibiotic), and mixed with 50 μL nuclease-free water. 1 μL of sample was then added to the PCR mixture as substitute for gDNA.

Table 2-10. Primers used in this study

Primer	Description	Sequence
Knockout Primers		
EHEC_*rbsDACBKR*_LRed_Fwd	EHEC *rbsDACBKR* KO forward primer	gtcaggattaaactgcgggtcagcgaaacgtttcgctgatggagaaaaaagtgtaggctggagctgcttc
EHEC_*rbsDACBKR*_LRed_Rev	EHEC rbsDACBKR KO reverse primer	gcttctgaaaacaaccgatggtaataatttgatcaaaagacagcgtaaatcatatgaatatcctccttag
EHEC_*rbsDACBKR*_Check_Fwd	EHEC rbsDACBKR KO check forward primer	cgtcacgatggttttccc
EHEC_*rbsDACBKR*_Check_Rev	EHEC rbsDACBKR KO check reverse primer	gcttcagcagcaatgttca
Crod_*rbsDACBKR*_LRed_Fwd	*C. rodentium rbsDACBKR* KO forward primer	cttaagattaaactgacgccagcgaaacgtttcgctggtggagcagaaaagtgtaggctggagctgcttc
Crod_*rbsDACBKR*_LRed_Rev	*C. rodentium rbsDACBKR* KO reverse primer	gctgttgagggcgaaagatggccgcaacctgatcaaagaccggcacaagccatatgaatatcctccttag
Crod_*rbsDACBKR*_Check_Fwd	*C. rodentium rbsDACBKR* KO check forward primer	aactgacgccagcgaaac
Crod_*rbsDACBKR*_Check_Rev	*C. rodentium rbsDACBKR* KO check reverse primer	acgatcgtatccgcgcta
*Z0415-19*_LRed_Fwd	Z0415-19 KO forward primer	gcgcgctaatttgggcgaacacttcctgactaccctgcaatgaggctgaagtgtaggctggagctgcttc
*Z0415-19*_LRed_Rev	Z0415-19 KO reverse primer	cgcctgatatgtcatcgcggcaaaacgcgtccattgaatatagccaatatcatatgaatatcctccttag
*Z0415-19*_Check_Fwd	Z0415-19 KO forward check primer	tctctccagcgcgctaat
*Z0415-19*_Check_Rev	Z0415-19 KO reverse check primer	atgtcatcgcggcaaaac
*ROD24811-51*_LRed_Fwd	ROD24811-51 KO forward primer	ttgctaacctcgtttcgtgacatgccctggtcccattaaaaggaacgacagtgtaggctggagctgcttc
*ROD24811-51*_LRed_Rev	ROD24811-51 KO reverse primer	cgcttcgctatagagccgccaggtggcggtggcctgctgccatgcgctgccatatgaatatcctccttag
*ROD24811-51*_Check_Fwd	ROD24811-51 KO check forward primer	tgccctggtcccattaaa
*ROD24811-51*_Check_Rev	ROD24811-51 KO check reverse primer	attaacgcctgccactgc
EHEC_*araE*_LRed_Fwd	EHEC araE KO forward primer	attgttcacgtatttttttcactatgtcttactctctgctggcaggaaaaagtgtaggctggagctgcttc
EHEC_*araE*_LRed_Rev	EHEC araE KO reverse primer	ctctattaacgaaaaaagggccggatgtacagcacatccggcccgtgaaacatatgaatatcctccttag
EHEC_*araE*_Check_Fwd	EHEC araE KO check forward primer	aatatccatcacataacggcatg
EHEC_*araE*_Check_Rev	EHEC araE KO check reverse primer	attcccagctcattcctccc
EHEC_*araFGH*_LRed_Fwd	EHEC araFGH KO forward primer	ttttgccctgcacaaaacgacactaaagctggagagaaccgtgtaggctggagctgcttc

EHEC_*araFGH*_LRed_Rev	EHEC araFGH KO reverse primer	tgtggtgggaaaaaacgttaaattgttg tggaaaaaagcacatatgaatatcctc cttag
EHEC_*araFGH*_Check_Fwd	EHEC araFGH KO check forward primer	tcccgctaaatttatgcacgt
EHEC_*araFGH*_Check_Rev	EHEC araFGH KO check reverse primer	ttgcaacgaagaacagccaa
EHEC_*araBAD*_LRed_Fwd	EHEC araBAD KO forward primer	gcaactctctactgtttctccatacccgtt tttttggatggagtgaaacggtgtaggc tggagctgcttc
EHEC_*araBAD*_LRed_Rev	EHEC araBAD KO reverse primer	aaaaaaccaggcttgattatagcctgg tttcatttgattggctgtggttttatacagtc acatatgaatatcctccttag
EHEC_*araBAD*_Check_Fwd	EHEC araBAD KO check forward primer	cgtcacactttgctatgcca
EHEC_*araBAD*_Check_Rev	EHEC araBAD KO check reverse primer	aagataaaacctgcctgcgc
EHEC_*araC*_LRed_Fwd	EHEC araC KO forward primer	tgcaatatggacaattggtttcttctctga atggcgggagtatgaaaagtgtgtag gctggagctgcttc
EHEC_*araC*_LRed_Rev	EHEC araC KO reverse primer	caaaccctatgctactccgtcaagccg tcaattgtctgattcgttaccaacatatg aatatcctccttag
EHEC_*araC*_Check_Fwd	EHEC araC KO check forward primer	tcttctctgaatggcgggag
EHEC_*araC*_Check_Rev	EHEC araC KO check reverse primer	atggacgaagcagggattct
EHEC_*rbsD*_LRed_Rev	EHEC rbsD KO reverse primer	ggaaggctttatcgatgcctttaagctg aagtaatgcttccatgacggcccatat gaatatcctccttag
EHEC_*rbsD*_Check_Fwd	EHEC rbsD KO check forward primer	ccagacgcctcctttcttca
EHEC_*rbsD*_Check_Rev	EHEC rbsD KO check reverse primer	tcgcgagtatagatgccagt
EHEC_*rbsACB*_LRed_Fwd	EHEC rbsACB KO forward primer	tccgtatgcgaatatcattctctgtgctg gcgtgacgttcttgaggccgtcgtgtag gctg
EHEC_*rbsACB*_LRed_Rev	EHEC rbsACB KO reverse primer	ctgtcgatgacgtatttatgtcaccatca ggtcatacaacctgattaaaacatatg aata
EHEC_*rbsACB*_Check_Fwd	EHEC rbsACB KO check forward primer	ttcgcagcggagaatgtt
EHEC_*rbsACB*_Check_Rev	EHEC rbsACB KO check reverse primer	cgtgccctgcttttcttt
EHEC_*rbsK*_LRed_Fwd	EHEC rbsK KO forward primer	gggcacgcgccaccctaacacggtg gcgcattttatggacatcccgaatgtgt aggctg
EHEC_*rbsK*_LRed_Rev	EHEC rbsK KO reverse	gaggtagaaacgcccgccaggcag gcgggcaacatctttcattgtagccaa gcgcatatgaata
EHEC_*rbsK*_Check_Fwd	EHEC rbsK KO check forward primer	ggtggcgcattttatgga
EHEC_*rbsK*_Check_Rev	EHEC rbsK KO check reverse primer	gcttcactgacgaagcga
EHEC_rbsR_LRed_Fwd	EHEC rbsR KO forward primer	taccgtggcgtgaagagatcgacgca tttttagacaggcagaggtgacgcgtgt aggctg
EHEC_*rbsR*_Check_Fwd	EHEC rbsR KO check forward primer	cgtaaaggcgcacaacct

*ROD24811-41*_LRed_Fwd	ROD24811-41 KO forward primer	ttgctaacctcgtttcgtgacatgccctg gtcccattaaaaggaacgacagtgta ggctggagctgcttc
*ROD24811-41*_LRed_Rev	ROD24811-41 KO reverse primer	ttcctacatcgacaccaataaaataac tcgccatcatttttctcccgaaacatatg aatatcctccttag
*ROD24811-41*_Check_Fwd	ROD24811-41 KO check forward primer	acatgccctggtcccattaa
*ROD24811-41*_Check_Rev	ROD24811-41 KO check reverse primer	ggtaaatttcaatggcgcgc
*ROD24851*_LRed_Fwd	ROD24851 KO forward primer	cctctttatcgcattacagaatcgtaaa gcctgatttcgggagaaaaatggtgta ggctggagctgcttc
*ROD24851*_LRed_Rev	ROD24851 KO reverse primer	cgcttcgctatagagccgccaggtggc ggtggcctgctgccatgcgctgccatat gaatatcctccttag
*ROD24851*_Check_Fwd	ROD24851 KO check forward primer	atacggcgagtcctatctgc
*ROD24851*_Check_Rev	ROD24851 KO check reverse primer	ttatcatcaggtcgtcggca
*ROD24861*_LRed_Fwd	ROD24861 KO forward primer	gagatgtatcaggatcacatgaagtac cgtcagctgatgcaggaggcgttgtgt aggctggagctgcttc
*ROD24861*_LRed_Rev	ROD24861 KO reverse primer	tattttctgcatatcggaaaaagccccg tctatgggacggggccaggccacata tgaatatcctccttag
*ROD24861*_Check_Fwd	ROD24861 KO check forward primer	cgcagaccaaccgcattaag
*ROD24861*_Check_Rev	ROD24861 KO check reverse primer	taacgtcaggattgcagggg
Crod_*araBAD*_LRed_Fwd	*C. rodentium* araBAD KO forward primer	cccactcactactgtttctccatacccgt atttctggatggagtgaaacggtgtag gctggagctgcttc
Crod_*araBAD*_LRed_Rev	*C. rodentium* araBAD KO reverse primer	tgtgttccggaataaaaatacgcgcca ctgtcgggacgcgtattttgcatcatatg aatatcctccttag
Crod_*araBAD*_Check_Fwd	*C. rodentium* araBAD KO check forward primer	acaacggcagaaatgtccac
Crod_*araBAD*_Check_Rev	*C. rodentium* araBAD KO check reverse primer	ctttcattcgctggagggc
EHEC_*bssS*_LRed_Fwd	EHEC bssS KO forward primer	gcattgaacctcgaataacgttgtctag taacacgaattaggggccatggtgta ggctggagctgcttc
EHEC_*bssS*_LRed_Rev	EHEC bssS KO reverse primer	aatggtaaaggcaccggtgaggtgcc ttttgggtggatggtcatgtcatgcatat gaatatcctccttag
EHEC_*bssS*_Check_Fwd	EHEC bssS KO check forward primer	aaaacgcattgaacctcgaat
EHEC_*bssS*_Check_Rev	EHEC bssS KO check reverse primer	agcacgtttaaaccggcg
EHEC_*bssR*_LRed_Fwd	EHEC bssR KO forward primer	atagccatactatttaattgcaacaagg ctggaaagaggaggatcgaagtgtgt aggctggagctgcttc
EHEC_*bssR*_LRed_Rev	EHEC bssR KO reverse primer	tttacaaaaacttaaacatgaagggg gagacgctttctccccctttagtttcatatg aatatcctccttag
EHEC_*bssR*_Check_Fwd	EHEC bssR KO check forward primer	tacaggaaaggtcagggcag
EHEC_*bssR*_Check_Rev	EHEC bssR KO check reverse primer	cccagagagcggaagaaaga

FLAG-tag primers		
TUV93-0$^{Z0417\text{-}3xFLAG}$_Fwd	Forward primer for introducing 3xFLAG tag at Z0417 C-terminus	tcctgcctgacagcagttctcaggcca ggattgcggaggtggtaaatggcgac tacaaagaccatgacgg
TUV93-0$^{Z0417\text{-}3xFLAG}$_Rev	Reverse primer for introducing 3xFLAG tag at Z0417 C-terminus	acagcaccacaatcagtaatgccagc cagaattcatggcgttttttcagttcagc cataatatcctccttagttcc

Cloning Primers		
pMK1*lux*-PLEE1_EHEC_Fwd	Forward primer for cloning EHEC LEE1 promoter with EcoRI	cccgaattcctgtaactcgaattaagt
pMK1*lux*-PLEE1_EHEC_Rev	Reverse primer for cloning EHEC LEE1 promoter with BamHI	cccggatccaatctccgcatgctttaat a
pMK1*lux*-PLEE1_Crod_Fwd	Forward primer for cloning *C. rodentium* LEE1 promoter with EcoRI	cccgaattccaaatcgggtacgcgatc
pMK1*lux*-PLEE1_Crod_Rev	Reverse primer for cloning *C. rodentium* LEE1 promoter with BamHI	cccggatccaatctcctcatacttttata
pMK1*lux*-PZ0415_Fwd	Forward primer for cloning Z0415 promoter with EcoRI	cccgaattcattcaccagaaatggac g
pMK1*lux*-PZ0415_Rev	Reverse primer for cloning Z0415 promoter with BamHI	cccggatccatttcagcctcattgcag
pMK1*lux*-PROD24811_Fwd	Forward primer for cloning ROD24811 promoter with EcoRI	cccgaattcctgccgcgactgctggca
pMK1*lux*-PROD24811_Rev	Reverse primer for cloning ROD24811 promoter with BamHI	cccggatccattgtcgttccttttaat
pMK1*lux*-ParaB_Fwd	Forward primer for cloning araB promoter with EcoRI	cccgaattccgggaccaaagccatga c
pMK1*lux*-ParaB_Rev	Reverse primer for cloning araB promoter with XbaI	gcgctctagacgtttcactccatccaaa
pMK1*lux*-PbssS_Fwd	Forward primer for cloning bssS promoter with EcoRI	gcggaattctcagcgaataatatgcag tgattt
pMK1*lux*-PbssS_Rev	Reverse primer for cloning bssS promoter with BamHI	ggatcccatggccccctaattcgtgtta cta
pMK1*lux*-PbssR_Fwd	Forward primer for cloning bssR promoter with EcoRI	ccgaattcgttacaggaaaggtcagg gca
pMK1*lux*-PbssR_Rev	Reverse primer for cloning bssR promoter with BamHI	ccggatcccatacttcgatcctcctctttc
pMK1*lux*_Check_Fwd	Forward primer to check pMK1*lux* cloning	ctataaaaataggcgtatcac
pMK1*lux*_Check_Rev	Reverse primer to check pMK1*lux* cloning	ctggccgttaataatgaatg
pACYC184-*araC*_Fwd	araC Gibson assembly forward primer	tgaagtcagccccatacgattgcaatc gccatcgtttca
pACYC184-*araC*_Rev	araC Gibson assembly reverse primer	caatccatgccaacccgttcttatgaca acttgacggct
pACYC184_Check_Fwd	Forward primer to check pACYC184 cloning	gacgctcaaatcagtggtgg
pACYC184_Check_Rev	Reverse primer to check pACYC184 cloning	gcattcacagttctccgcaa
pACYC184_Linear_Fwd	pACYC184 linearisation forward primer	gaacgggttggcatggattg

pACYC184_Linear_Rev	pACYC184 linearisation reverse primer	atcgtatggggctgacttca
pSUPROM-*araC*_Fwd	Forward primer for cloning araC with BamHI	ggccggatcctcttctctgaatggcgg gag
pSUPROM-*araC*_Rev	Reverse primer for cloning araC with XbaI	ggcctctagaatggacgaagcaggg attct
pSUPROM-*araE*_Fwd	Forward primer for cloning araE with BamHI	ggccggatcctgtcttactctctgctggc a
pSUPROM-*araE*_Rev	Reverse primer for cloning araE with XbaI	ggcctctagaaacgagacaaacgcc tcaac
pSUPROM-*araBADE*_F1_Fwd	araBAD fragment Gibson assembly forward primer	tctaccacagaggaggatccatggcg attgcaattggc
pSUPROM-*araBADE*_F1_Rev	araBAD fragment Gibson assembly forward primer	cagcagagagttactgcccgtaatatg cc
pSUPROM-*araBADE*_F2_Fwd	araE fragment Gibson assembly forward primer	cgggcagtaactctctgctggcagga aaaaatg
pSUPROM-*araBADE*_F2_Rev	araE fragment Gibson assembly forward primer	ctcgaggggtcgactctagatcagac gccgatatttctcaac
pSUPROM-*araA*_Fwd	Forward primer for cloning araA with XbaI	ggcctctagaatgacgatttttgataatt atgaagtgtgg
pSUPROM-*araA*_Rev	Reverse primer for cloning araC with XhoI	ctcgagtacttagcggcgaaacccgt aatacactt
pSUPROM-*araBA*_Fwd	Forward primer for cloning araBA with BamHI	gcgggatccatggcgattgcaattggc
pSUPROM-*araBA*_Rev	Reverse primer for cloning araBA with XbaI	gcgctctagattagcggcgaaacccg t
pSUPROM-*Z0417*^{3xFLAG}_Fwd	Forward primer for cloning Z0417 with 3xFLAG tag with BamHI	ggatccatggtgaatcactatcaccac ggtctgt
pSUPROM-*Z0417*^{3xFLAG}_Rev	Reverse primer for cloning Z0417 with 3xFLAG tag with XbaI	gcg tctagactatttatcgtcgtcatctttgtag tc
pSUPROM-*Z0416-17*^{3xFLAG}_Fwd	Forward primer for cloning Z0416-17 with 3xFLAG tag with BamHI	ggatccatggaaaccttcctttcccttcg tc
pSUPROM-*Z0416-17*^{3xFLAG}_Rev	Reverse primer for cloning Z0416-17 with 3xFLAG tag with XbaI	tctagactatttatcgtcgtcatctttgtag tc
pSUPROM-*rbl*_Fwd	Rbl Gibson assembly forward primer	tctaccacagaggaggatccatgaaa ttcaaactcgcattactac
pSUPROM-*rbl*_Rev	Rbl Gibson assembly reverse primer	ctcgaggggtcgactctagatcataac gcctcctgcatc
pSUPROM_Check_Fwd	Forward primer to check pSUPROM cloning	ctcttcgctattacgccagc
pSUPROM_Check_Rev	Reverse primer to check pSUPROM cloning	accctcatcagtgccaacat
pSUPROM_Linear_Fwd	pSUPROM linearisation forward primer	tctagactcgacccctcg
pSUPROM_Linear_Rev	pSUPROM linearisation reverse primer	ggatcctcctctgtggtag

RT-qPCR Primers

*escT*_Fwd	escT RT-qPCR forward primer	tttgggctatagatgcggct
*escT*_Rev	escT RT-qPCR reverse primer	ggatgaatcgcttatagacggg
*escC*_Fwd	escC RT-qPCR forward primer	gctgaagtgagtgctcgttt

*escC*_Rev	escC RT-qPCR reverse primer	cctcaagcgggtcaataacg
*escV*_Fwd	escV RT-qPCR forward primer	ctaaaagttctccagtacgtgc
*escV*_Rev	escV RT-qPCR reverse primer	tcgccagagaaatcatcattca
*espA*_Fwd	espA RT-qPCR forward primer	ttcctgtaaatccgatgcgc
*espA*_Rev	espA RT-qPCR reverse primer	tggttgacgctttagatgcc
*tir*_Fwd	tir RT-qPCR forward primer	ttcctgtaaatccgatgcgc
*tir*_Rev	tir RT-qPCR reverse primer	atcgagcggaccatgatcat
*Z0415*_Fwd	Z0415 RT-qPCR forward primer	tggtgtcttcgctgttattagg
*Z0415*_Rev	Z0415 RT-qPCR reverse primer	cacggcataccatcgacttta
*Z0417*_Fwd	Z0417 RT-qPCR forward primer	tggaagtttccgaccgtattt
*Z0417*_Rev	Z0417 RT-qPCR reverse primer	tcatcaggaaaccgagttgtt
*Z0418*_Fwd	Z0418 RT-qPCR forward primer	gccttactggttaatcgcctac
*Z0418*_Rev	Z0418 RT-qPCR reverse primer	gtacagccacacacctttactc
Housekeeping_GroEL_Fwd	GroEL RT-qPCR forward primer	accgctgcagttgaagaa
Housekeeping_GroEL_Rev	GroEL RT-qPCR reverse primer	ctacggtttcgtcggagttag
Housekeeping_GapA_Fwd	GapA RT-qPCR forward primer	cggtaccgttgaagtgaaaga
Housekeeping_GapA_Rev	GapA RT-qPCR reverse primer	acttcgtcccatttcaggttag

Table 2-11. Overview of the protocol used for all PCR reactions in this study

Step	Temperature	Time	
Initial denaturation	98 °C	5 mins	
Denaturation	98 °C	20 sec	35
Annealing	50-72 °C	30 sec	cycles
Extension	72 °C	30-60 sec*	
Final extension	72 °C	10 mins	

*Extension time varied with polymerase: Q5® (30 sec/1 Kb) and GoTaq (60 sec/ 1Kb)

Table 2-12. Components for PCRs using Q5® HF 2 X Master Mix

Reagent	Volume
Q5® HF 2 X Master Mix	10 μL
Primers (10 μM)	1 μL
Template DNA	0.5 μL
Nuclease free H_2O	7.5 μL

Table 2-13. Components for PCRs using GoTaq Green Master Mix

Reagent	Volume
GoTaq Green 2 X Master Mix	10 μL
Primers (10 μM)	0.5 μL
Template DNA	0.5 μL
Nuclease free H_2O	8.5 μL

2.6.2. Gel electrophoresis

1 % (w/v) agarose gels containing 10,000 X GelRed (Biotium) were cast, and DNA electrophoresed at 100 V for 60 mins in 1 x Tris-Acetate-EDTA (TAE) buffer (**Table 2-14**). Prior to loading, DNA was mixed with 6 X purple loading dye. All samples were then loaded into wells along with a 1 Kb plus DNA ladder. Gels were then visualised using a BioRad gel documentation XR^+ system.

Table 2-14. Tris-acetate-EDTA (TAE) buffer (1 L)

Reagent	Amount
TAE (50x) (Formedium)	20 mL
ddH_2O	980 mL

2.6.3. Plasmid isolation

5 mL of LB was inoculated with a single colony harbouring the desired plasmid and incubated overnight at 37 °C, shaking (200 rpm). Cells were harvested by centrifugation at 3,000 rpm for 10 mins. Plasmid DNA was then isolated using the Monarch® Plasmid MiniPrep Kit (NEB) according to the manufacturer's instructions. Plasmid DNA was eluted in 50 μL nuclease-free water.

2.6.4. PCR clean-up and DNA gel extraction

For PCR product clean-up and gel extraction the respective Monarch® kits were used following the manufacturer's instructions provided. All DNA was eluted in a 30 μL final volume of nuclease-free water.

2.6.5. Ligations

Purified inserts and linearized plasmid DNA were ligated to have a final insert:vector ratio of 3:1. Ligation reactions contained 1 μL T4 ligase and 1 μL T4 ligase buffer, made up to a total volume of 10 μL. All reactions were incubated overnight at 16 °C. 2-5 μL of ligation mixture was transformed into commercial DH5α cells (NEB) by heat shock. Colonies were screened by colony PCR and positively identified clones subject to plasmid isolation. Plasmids (50 ng/μL) were then sent for sequencing (Eurofins) to confirm cloning.

2.6.6. Gibson Assembly (NEBuilder)

Ligation reactions were set up to contain 0.25 μL of each purified insert, 2.5 μL NEBuilder® HiFi DNA assembly mix and nuclease-free water to make a total volume of 4 μl. Reactions were incubated at 50 °C for 20 mins and then placed immediately on ice. For later use, reactions were stored at -20 °C.

2.6.7. Heat shock transformation

50 μL aliquots of commercial chemically competent cells (NEB) were thawed on ice, mixed with 2-5 μL DNA, and incubated on ice for 30 mins. Cells were then heat-shocked at 42 °C for 30 sec and immediately placed back on ice for 5 mins. Cells were recovered in 950 μL pre-warmed SOC media and incubated at 37 °C, shaking (200 rpm) for 2 h. Suspensions were then plated (1:1, 1:10) on LB agar supplemented with the appropriate antibiotic.

2.6.8. Electroporation

Overnight culture was used to inoculate 10 mL SOB 1:100 and incubated for 2 h at 37 °C, shaking (200 rpm) until an OD_{600} of 0.4 was reached. Cells were then incubated on ice for 20 mins prior to harvesting by centrifugation at 3,000 rpm at 4 °C for 10 mins. Supernatants were discarded, pellets resuspended in 1 mL ice-cold sterile ddH_2O and transferred into 1.5 mL Eppendorf tubes. Cells were centrifuged at 9,000 rpm at 4 °C for 3 mins. The wash step was repeated twice more (3X total) and cells resuspended in 100 μL of ice-cold H_2O. For electroporation, 50 μL of cells were mixed with 1 μL DNA (typically, 100 ng/μL). The DNA/cell mix was transferred to a pre-chilled 2 mm electroporation cuvette (Flowgen) and shocked at 2500 volts (V) using an Eporator electroporator (Eppendorf). Cells were immediately recovered in 950 μL pre-warmed SOC media and incubated at 37 °C, shaking (200 rpm) for 2 h. Suspensions were then plated (1:1, 1:10) on LB agar supplemented with the appropriate antibiotic.

2.6.9. RNA extraction and storage

Total RNA extraction was done using a Monarch® Total RNA Miniprep Kit (NEB) following the manufacturer's instructions. Prior to extraction, 10 mL cultures of TUV93-0 in MEM-HEPES supplemented with and without L-arabinose (5 mg/mL) were grown for 9 h (late exponential). After 9 h, 1.5 mL of culture was incubated on ice for 5 mins, then centrifuged at 9,000 rpm for 3 mins. The supernatant was discarded, and pellet resuspended in 1 volume of RNA protect reagent (Qiagen) then incubated at room temperature for 5 mins, shaking (50 rpm). Pellets were harvested by a final centrifugation at 9,000 rpm for 3 mins. For short- and long-term storage, pellets were kept at -20 and -80 °C respectively.

Pellets were defrosted on ice and resuspended in 400 μL lysozyme (1 mg/mL) and incubated for 5 mins at 25 °C. Two volumes of RNA lysis buffer were then added to the cells and vortexed vigorously for 10 sec before centrifugation at 13,000 rpm for 2 mins to pellet cellular debris. The supernatant was transferred to a gDNA removal column and centrifuged at 13,000 rpm for 30 sec. The flowthrough was retained as this fraction contained the RNA. Equal volumes of > 95 % EtOH were added to the flowthrough and transferred to an RNA purification column for centrifugation at 13,000 rpm for 30 sec. Bound RNA was then washed with 500 μL RNA priming buffer by centrifugation at 13,000 rpm for 30 sec, followed by 2 washes with 500 μL RNA wash

buffer at 13,000 rpm for 30 sec, and a final wash with another 500 μL RNA wash buffer at 13,000 rpm for 2 mins. The column was transferred to an RNAse-free tube and eluted into 30 μL pre-warmed nuclease-free water by centrifugation at 13,000 rpm for 30 sec.

The extracted RNA was treated to remove any contaminant DNA using DNAse TURBO (Invitrogen) using the manufacturer's instructions provided. The concentration and quality of extracted RNA samples were analysed using a Nanodrop 2000 (Thermo Scientific). Samples with a 260/280 absorbance of 1.8-2.0 and 260/230 absorbance of 2.0-2.2 were considered high quality and free of contaminants. To check for degradation 1 μL of RNA was electrophoresed on a 1.6 % (w/v) agarose gel at 100 V for 80 mins.

2.6.10. cDNA synthesis

cDNA was prepared from purified RNA prior to RT-qPCR using the LUNA Script Kit (NEB). All samples were normalised, and reactions contained 3 μL of RNA, 2 μL LUNA script mix and 5 μL NF H_2O. No reverse transcriptase and no template controls, whereby the specified component was replaced with nuclease-free water, were used to detect the presence of contaminant DNA that may impact the results. The reaction conditions are specified in **Table 2-15**.

Table 2-15. Protocol used for synthesis of cDNA from RNA

Step	Temperature	Time
Primer annealing	25 °C	2 mins
cDNA synthesis	55 °C	10 mins
Heat inactivation	95 °C	60 sec

2.6.11. Reverse transcriptase quantitative PCR (RT-qPCR)

RT-qPCR was performed using previously prepared cDNA as described above. Reactions were set up to have a total volume of 10 μL containing 5 μL Luna Universal qPCR Master Mix, 1 μL of cDNA, 0.25 μL forward primer, 0.25 μL reverse primer and 3.5 μL nuclease-free water. All reactions were performed in triplicate for each of the biological replicates using the LightCycler® 96 Instrument (Roche) under the conditions specified in **Table 2-16**. The data was then analysed using the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen, 2001).

Table 2-16. 2-step amplification RT-qPCR protocol

Step	Cycles	Temperature	Time
Preincubation	1	95 °C	60 sec
2 step amplification	45	95 °C	15 sec
		60 °C	30 sec
Melting	1	95 °C	10 sec
		60 °C	60 sec
		97 °C	1 sec
Cooling	1	37 °C	30 sec

2.7. Genetic techniques

2.7.1. Lambda Red

Genetic knockouts were made using the protocol described by Datsenko & Wanner., (2000), using PCR products. PCR products consisting of an antibiotic resistance cassette were amplified from pKD3 or pKD4 with the addition of 50 bp overhangs that directly flank the gene or region to be deleted.

Briefly, overnight cultures of strains harbouring pKD46 were grown in LB-Amp at 30 °C, shaking (200 rpm). 100 μL of culture was used to inoculate 10 mL SOB-Amp and incubated 30 °C, shaking (200 rpm). After 90 mins, cultures were spiked with 0.1 M L-arabinose to induce the expression of the bacteriophage Lambda Red genes (*exo,* γ and β) encoded on pKD46 under the control of an L-arabinose inducible promoter (P_{araB}). Exo displays 5′ to 3′-dsDNA exonuclease activity that generates 3′ overhangs on linear PCR products (Sharan et al., 2009). Beta then binds to the cut 3′ overhangs to promote single stranded annealing and recombination with target DNA (Sharan et al., 2009). Gam is responsible for preventing the degradation of dsDNA fragments by RecBCD (Sharan et al., 2009).

After induction, cultures were incubated for a further 90 mins. Cells were then harvested and made electrocompetent as described in *section 2.6.8*. To the electrocompetent cells, 100-200 ng of PCR product specific for the gene or region to be deleted was electroporated into cells (*Section 2.6.8.)*. Transformed cells were then recovered at 37 °C, shaking (200 rpm) to encourage the loss of pKD46 prior to plating on LB containing the appropriate antibiotic (Cm for pKD3 and Kan for pKD4) and incubated at 37 °C, overnight. The following day colonies were checked for successful

deletions by colony PCR (*section 2.6.1.)*. Successful mutants were then re-streaked and used to prepare glycerol stocks for long term storage.

2.7.2. Chromosomal tagging of genes

In brief, target genes were tagged such that the translated protein would contain an in-frame C-terminal 3xFLAG tag by replacement of the native STOP codon. A replacement cassette containing both Kan resistance cassette and 3xFLAG epitope tag was PCR amplified from pDOC-F. Cassettes were then cleaned-up as described in *section 2.6.4.* and quantified using a Nanodrop 2000 (Thermo Scientific). The same protocol as used for making chromosomal gene deletions (See *section 2.7.1.*) was then performed to introduce the epitope tag.

2.7.3. Elimination of antibiotic resistance cassettes using pCP20

As described by Datsenko & Wanner., (2000), FRT-flanked Kan and Cm resistance cassettes used to make mutants were eliminated by thermal induction of a flippase recombinase encoded on temperature-sensitive, pCP20. In brief, pCP20 was electroporated into cells (*Section 2.6.8.*) and plated onto LB-Amp, grown at 30 °C overnight. Single colonies were then re-streaked onto LB-agar and incubated at 42 °C overnight for the induction of FLP and select for loss of pCP20. Single colonies were then sequentially streaked onto LB-agar > LB-Kan/Cm > LB-Amp and incubated at 37 °C overnight. Elimination of the antibiotic resistance cassette was checked by colony PCR (*Section 2.6.1.*).

2.8. Biochemical techniques

2.8.1. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

All protein samples were prepared by resuspension in 1 x lithium dodecyl sulphate (LDS) buffer, boiling for 10 mins and then centrifuging at 9,000 rpm for 5 mins. Typically, 10 μL of sample was loaded into the wells of a pre-cast 4-12 % Bis-Tris NuPAGE gel (Invitrogen) and run for 90 mins at 150 V in 1 x NuPAGE MES SDS Running Buffer (Invitrogen) (**Table 2-17**). PageRuler^{TM} plus pre-stained protein ladder (10-250 kDa) was run alongside all samples. SDS-PAGE gels were then stained for 1 h in Coomassie blue stain (**Table 2-18**) on an orbital shaker and destained by 3 x 30

mins of washing in destain solution (**Table 2-19**). Alternatively, gels were destained in dH_2O overnight.

Table 2-17. MES SDS running buffer (1 L)

Reagent	Amount
20 x MES running buffer (Invitrogen)	50 mL
ddH_2O	950 mL

Table 2-18. Coomassie blue stain (1 L)

Reagent	Amount
Methanol	400 mL
Acetic acid	100 mL
Coomassie blue	0.5 g
ddH_2O	500 mL

Table 2-19. Coomassie blue de-stain (1 L)

Reagent	Amount
Methanol	400 mL
Acetic acid	100 mL
ddH_2O	500

2.8.2. Western blotting

Proteins for Western blot analysis were transferred from a 4-12 % Bis-Tris NuPAGE gel (ThermoFisher) to a 0.45 μM nitrocellulose membrane (GE Healthcare) in 1 x transfer buffer (**Table 2-20**), using an XCell II Blot module (Invitrogen) at 30 V for 90 mins. Blots were then blocked with 5 % skimmed milk made with PBS-Tween (0.1 %) (**Table 2-21**) at room temperature for 1 h. The membrane was then incubated with the primary antibody at the relevant concentration (**Table 2-21**) in 10 mL PBS-T overnight at 4 °C. The membrane was washed three times before incubation with a horseradish-peroxidase conjugated secondary antibody at the appropriate concentration (**Table 2-22**) in 10 mL PBS-T for 60 mins at room temperature. The membrane was washed a final three times and developed using SuperSignal West Pico chemiluminescent ECL substrate (1:1; stable peroxide:luminol/enhancer) for 5 mins. All washes were done using 20 mL PBS-T for 10 mins at room temperature. Blots were then visualised using the G:Box Chemi system (Syngene).

Table 2-20. Western transfer buffer (1 L)

Reagent	Amount
25 x Transfer buffer (Invitrogen)	40 mL
ddH_2O	960

Table 2-21. PBS-Tween (PBS-T) (400 mL)

Reagent	Amount
1 x PBS	400 mL
Tween-20	0.4 mL

Table 2-22. Antibody concentrations used in this study

Antibody	Source	Primary conc.	Secondary conc.
FLAG	Mouse	1:5000	1:10000
EspD	Mouse	1:2500	1:5000
GroEL	Rabbit	1:25000	1:30000

2.8.3. Thin layer chromatography (TLC)

Cell free extracts were prepared by passing 1 mL of culture through a 0.2 μM PES filter. To a TLC silica-coated foil plate (Merck), 3-6 μL spots were then equidistantly spotted 1 cm above the bottom edge. The plates were placed in a tank equilibrated with 1-butanol:acetic acid:water (2:1:1) (**Table 2-23**) and left to run until the solvent was 1 cm from the top of the plate. The plates were then briefly dried in a fume hood, before being immersed with 0.4 % orcinol in 8.0 % sulphuric acid, and further dried for 10 mins at 110 °C to visualise the carbohydrate products. The appropriate sugar standard was run in tandem with samples at a concentration of 5 mM.

Table 2-23. TLC running buffer

Reagent	Amount
Acetic acid	250 mL
1-Butanol	500 mL
ddH_2O	250

2.9. Phenotypic assays

2.9.1. Microplate reader growth assays

Overnight cultures of strains were centrifuged at 3,500 rpm for 10 mins, supernatants discarded and washed as described in *section 2.1*. All bacterial growth assays were done using a clear, flat-bottom, 96-well polystyrene microtiter plate (Greiner). Wells contained a maximum of 200 μL of media, typically supplemented with sugar in the concentration range of 0 to 5 mg/mL. Each well was inoculated with the strain of interest 1:100. Growth was carried out at 37 °C with shaking (200 rpm) for 12-16 h using a FLUOstar Omega microplate reader. Measurements (OD_{600}) were taken every 20 mins.

2.9.2. Manual growth assays

For manual growth, 10 mL of media in 150 mL baffler flasks were inoculated with overnight culture to a starting OD_{600} 0.05. Every 1-2 h 500 μL of culture was removed and absorbance (OD_{600}) measured using a spectrophotometer (Biochrom Colourwave WPA, CO7500 colorimeter). For OD_{600} reading > 1.2, cultures were diluted 1:10.

2.9.3. *Bacteroides* growth assays

For *Bacteroides* growth assays, overnight cultures were prepared by inoculating 5 mL of BHI supplemented with 0.1 % hematin-histidine with 30 μL of a *B. thetaiotaomicron* VPI-5482 glycerol stock. Cultures were grown anaerobically at 37 °C statically using an A35 anaerobic cabinet (Don Whitley Scientific). Following overnight growth, cultures were centrifuged at 3,500 rpm for 10 mins and supernatant discarded. Pellets were then resuspended in the equivalent volume of PBS and used to inoculate freshly prepared BMM (supplemented with 0.1 % hematin-histidine and 20 mM $NaNO_3$) to a starting OD_{600} of 0.1. Media was also supplemented with and without 5 mg/mL L-arabinose. Cultures were grown under anaerobic conditions at 37 °C for 12 h, with the OD_{600} being measured every 1-2 h using a spectrophotometer.

For *Bacteroides*-EHEC co-cultures, overnight cultures were prepared for each as previously described. Following centrifugation at 3,500 rpm for 10 mins and resuspension in PBS, 5 mL of freshly prepared BMM (supplemented with 0.1% hematin-histidine and 20 mM $NaNO_3$) was inoculated with each bacterium to a starting OD_{600} of 0.1. Monocultures were also set up to act as controls. All cultures were grown

anaerobically at 37 °C for 12 h. At each time point the OD_{600} of monocultures was measured using spectrophotometer and 50 μL aliquots were taken from all cultures (mono- and cocultures) for serial dilutions in PBS. For each dilution, 20 μL was spotted onto LB and LB-Gent. To select for only EHEC, LB agar plates were incubated aerobically at 37 °C overnight, whilst LB-Gent agar plates were incubated anaerobically at 37 °C for two days to select for only *Bt*. Colonies were later counted to obtain the CFU/mL using the calculation below:

(Number of colonies x dilution factor) / volume of culture plated (mL)

2.9.4. Microplate reader luminescence assays

Promoter activity was determined using the same set-up as growth assays, measuring both cell density (OD_{600}) and absolute luminescence (LUX) with gain set to 2000. Assays were conducted in white walled, clear flat-bottom, 96-well polystyrene microtiter plates (Corning). The relative luminescence units (RLU) were calculated by dividing the LUX values by OD_{600}.

2.9.5. Endpoint luminescence measurements

Promoter activity during manual growth was determined by inoculating 5 mL MEM-Amp with overnight culture 1:50. Cultures were then grown at 37 °C shaking (200 rpm) for 2 h, before they were spiked with 5 mg/mL L-arabinose and grown for a further 5 h. A total culture volume of 200 μL was then transferred to a white walled, clear flat-bottom, 96-well polystyrene microtiter plate. A single measurement of cell density (OD_{600}) and LUX was taken using a FLUOstar Omega microplate reader to determine the RLU at that endpoint. Non-spiked (no L-arabinose) controls were included for all cultures.

2.9.6. Biofilm assays

Biofilm formation was assayed using a 96-well flat-bottom polystyrene microtiter plate (Greiner). Each well contained 200 μL of media supplemented with either 0, 0.5 or 5 mg/mL L-arabinose. Wells were inoculated 1:100 with the strain of interest prepared from an overnight culture and grown statically at 37 °C for 24 h. After 24 h, waste media (containing planktonic bacteria) was discarded, and plates were sequentially washed in sterile PBS to remove any remaining unbound bacteria. Biofilms were then fixed by

drying at 50 °C for 20 mins, followed by staining with 1 % crystal violet for 15 mins at room temperature. The crystal violet stain was then removed, and plates rinsed with sterile ddH_2O. To qualitatively assess biofilm formation, images were taken using spImager (S&P Robotics). For quantitative measurements, crystal violet was solubilised by addition of 100 % EtOH to each well for 15 mins at room temperature. The contents of each well was briefly mixed and then transferred to a clean 96-well flat-bottom polystyrene microtiter plate (Greiner). The absorbance of each well was measured at 575 nm (OD_{575}) using a FLUOstar Omega microplate reader. Assays were conducted in technical triplicates for each biological replicate.

2.9.7. Secretion assays

Overnight cultures were used to inoculate 50 mL pre-warmed MEM-HEPES 1:100. Media was supplemented with or without 5 mg/mL L-arabinose prior to incubation at 37 °C, shaking (200 rpm) for 9 h (OD_{600} 0.8-0.9; late exponential). Cultures were harvested by centrifugation at 3,500 rpm for 15 mins at 4 °C. The supernatant was carefully removed, and filter sterilised by passing through a 0.45 μM filter. In a fume hood, trichloroacetic acid (TCA) was added to supernatant fractions to a final concentration of 10 % (v/v) prior to storage at 4 °C overnight. Following overnight precipitation, samples were centrifuged at 3,750 rpm for 60 mins at 4 °C to pellet secreted proteins. Supernatants were carefully discarded to leave a residual volume of approximately 1 mL for resuspension of secreted protein pellets, which were then transferred to a 1.5 mL Eppendorf. Samples were centrifuged at 14,000 rpm for 15 mins at 4 °C and supernatant removed. Pellets were dried at 50 °C for 20 mins and then resuspended in 150 μL 1 x LDS buffer for analysis by SDS-PAGE and Western blot (See *section 2.8.1.* and *2.8.2.*).

2.10. *In vitro*, *ex vivo* and *in vivo* infection models

2.10.1. HeLa cell infection assays

HeLa cells were routinely cultured in high glucose DMEM supplemented with 10 % foetal calf serum (FBS) at 37 °C with 5 % CO_2. Glass coverslips coated with rat tail collagen were seeded with 4 x 10^4 HeLa cells in a 12-well clear tissue culture-treated plate (Corning) and incubated overnight at 37 °C with 5 % CO_2 (Prepared by Helen Glenwright; TC facility manager, Newcastle University). Seeded cells were washed twice and replaced with 500 μl MEM-HEPES with or without 5 mg/mL L-arabinose. For

infection, bacterial cultures were grown in MEM-HEPES with or without 5 mg/mL L-arabinose for 9 h (OD_{600} 0.8-0.9) at 37 °C shaking (200 rpm). Cells were infected at a multiplicity of infection (MOI) of 100 (40 μL of bacterial culture adjusted to OD_{600} 0.1). Plates were then centrifuged at 400 rpm for 3 mins and incubated at 37 °C with 5 % CO_2 for 2 h. Wells were then washed with fresh media and grown for a further 3 h under the same conditions as stated previously.

Following infection, wells were washed three times and samples fixed with 4 % paraformaldehyde (PFA) for 15 mins at room temperature. Wells were then washed a further two times and permeabilised with 0.1 % Triton X-100 for 5 mins at room temperature. Wells were washed twice more before incubation with ActinRed™ 555 ReadyProbes™ reagent (Rhodamine phalloidin) (Invitrogen) as per the manufacturer's instructions. Tissues were then washed for a final time and mounted onto glass slides using Fluoroshield™ with DAPI. Unless stated otherwise, all washes were done using sterile PBS. Slides were imaged using a Zeiss Axioimager at x40 magnification. Images were randomly taken from 20 fields of view and processed using Zen Pro (Zeiss).

2.10.2. Intestinal organoids

All media and reagents used for intestinal monolayers were prepared by Dr Jonathan Chapman (Stewart Lab; Newcastle University). Frozen human intestinal organoids generated from adult ileum tissue were partially thawed by immersing in a 37 °C water bath. Organoid cells were then carefully mixed with 1 mL of cold CMGF- and centrifuged at 1,000 rpm for 5 mins to pellet. The supernatant was then removed, and pellet carefully resuspended in 120 μL ice cold Matrigel (Corning). To the wells of a prewarmed 24-well plate 2 droplets of organoid-Matrigel mix was added. Plates were briefly incubated at 37 °C (5 % CO_2) for 15 mins to allow the Matrigel to solidify and fix. Each well containing Matrigel was then submerged in 500-600 μL of CMGF+ with Wnt. The plate was then returned for incubation at 37 °C with 5 % CO_2. Cells were fed every 2 days by removing waste media and replacing with the equivalent volume of fresh media.

Upon passaging of the cells (6-9 days after the start), the old media was removed, and organoids resuspended in 300 μL (0.05 %) trypsin at 37 °C for 5 mins. To neutralise,

350 μL FBS (10 %) was immediately added to each well after incubation. The trypsin-Matrigel mix was then centrifuged at 1,000 rpm for 5 mins at 4 °C. The supernatant was removed prior to resuspension of the organoid pellet in X μL of ice-cold Matrigel, where X corresponded to the number of wells to seed multiplied by 30. Again, two droplets were added to each well and incubated for 5 mins at 37 °C to allow the Matrigel to solidify. Each well containing Matrigel was then submerged in 500-600 μL of CMGF+ with Wnt.

To generate monolayers, transwell inserts (0.4 μM pore size; Corning) were coated with 100 μL of Matrigel diluted in PBS (1:40) and incubated at 37 °C. Waste media was removed, and organoids were resuspended in 0.5 mM EDTA and combined. Organoids were then centrifuged at 1,200 rpm for 5 mins to pellet and then incubated with 1 mL trypsin (0.05 %)/EDTA (0.5 mM) at 37 °C for 5 mins. The cells were then passed through a 40 μM cell strainer and centrifuged at 1,500 rpm for 5 mins to pellet. Waste was discarded and pellets were resuspended in 1.2 mL CMGF+ supplemented with 10 μM ROCKi (Merck). Excess PBS was removed from transwells, and inserts were seeded with 4 x 10^5 cells. A total volume of 500 μL CMGF+ supplemented with 10 μM ROCKi was added to the basolateral compartment. Cells were incubated at 37 °C with 5 % CO_2 until confluent and a trans-epithelial electrical resistance (TEER) of 300 (Ω) was reached (~ 2 days). TEER of monolayers was measured daily. Apical and basolateral media were discarded prior to replacement with differentiation media and CMGF+, respectively. Monolayers were then incubated for a further 2 days at 37 °C with 5 % CO_2, after which differentiation media was changed.

EHEC-monolayer co-culture experiments were carried out using the organoid anaerobe co-culture (OACC) model (Ty et al., 2019). In brief, the console comprised of a gas permeable tissue culture plate (Corning) and oxygen containing compartment. Each well contained 600 μL of differentiation media. In a Coy Type B anaerobic chamber, the apical media of transwells was removed and replaced with pre-warmed MEM-HEPES supplemented with and without 5 mg/mL L-arabinose. Monolayers were infected with EHEC at an MOI of 50 for 6 h at 37 °C rotating (30 rpm). Apical media was collected and transwells excised before resuspension in 1 mL PBS by pulse vortexing. EHEC counts were obtained by serially diluting apical and transwell fractions onto LB-agar plates and incubating at 37 °C overnight.

2.10.3. Murine *in vivo* infection with *C. rodentium*

Both single and co-infection of BALB/c mice was done following the protocol described by Crepin *et al*., (2016). Cages of 4-6 adult female mice (8-10 weeks) were orally gavaged with 200 μL sterile PBS containing 3 x 10^9 CFU *C. rodentium*. For co-cultures, an equal volume (1:1) of WT and mutant *C. rodentium* were combined. The starting inoculum was confirmed by retrospective serial dilution plating. Mice were scored for changes in weight daily and faecal pellets collected every other day for 17 days. Every 0.1 g of faeces was resuspended in 1 mL PBS and serially diluted. Serial dilutions were plated on LB agar containing the necessary antibiotic to determine the CFU/g of faeces.

2.11. Bioinformatics

2.11.1. Sequence retrieval and similarity searches

The nucleotide sequences for the genes were retrieved from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2002) in this study. For protein domain analysis, amino acid sequences were searched in InterPro Scan (Jones et al., 2014) using the default parameters. Percent identity scores for sequences of interest were determined using either BLASTN or BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) dependent on the sequence type.

2.11.2. Multiple sequence alignments and phylogenetic analysis

Nucleotide and amino acid sequences were aligned using Clustal Omega (Sievers et al., 2011) and MUSCLE (Edgar, 2004) respectively, under the default parameters. For alignment summarisation sequence logos were generated using WebLogo3 (Crooks et al., 2004). Amino acid sequences used for phylogenetic analysis were aligned in MEGA X (Kumar et al., 2018) using the MUSCLE alignment tool. The model tool was used to find the best DNA/protein model prior to phylogenetic reconstruction.

Phylogenomic analysis of Z0415-19 and LEE carriage presented in this thesis was performed in collaboration with Dr Rhys White (University of Queensland). In brief, paired-end sequence read data for 1,067 distinct strains of *E. coli* and *Shigella* sp. (Connolly et al., 2014) were retrieved from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA). Raw reads were filtered to remove low-quality bases. All quality-trimmed paired-end reads were then mapped to the complete chromosome of strain EDL933.

2.11.3. Protein homology modelling and annotation

The full-length amino acid sequences for Z0415-19 and ROD_24811-51 were run through AlphaFold2 using the multimer model to generate a predicted 3D model by Dr Rhys Grinter (Monash University). The models for ROD_24851 were generated using the Phyre2 engine (Kelley et al., 2015). All images and annotation of predicted protein structures were done using PyMOL (v.2.5.0).

2.11.4. RNA-sequencing analysis and data visualisation

Enriched mRNA samples were processed for RNA-sequencing by the Newcastle Genomics Core Facility. Briefly, ribosomal depletion and library assembly was carried out using an illumina Ribo-Zero Tru-seq kit according to the manufacturer's specifications. Sequencing was carried out using a mid-range run on the Illumina Next-seq platform, generating 75-bp single-end reads. Raw read quality was checked using FASTQC (https://github.com/s-andrews/FastQC). To estimate transcript abundance, SALMON (Patro et al., 2017) was used under the default parameters for the mapping of reads to the *E. coli* O157:H7 strain EDL933 reference sequence (Accession: GCA_000006665), retrieved from Ensembl (https://www.ensembl.org/index.html). The transcript-level counts outputted from SALMON were then summarised at the gene-level using tximport. DESeq2 (v.1.28.1) (Love et al., 2014) was then used to normalise RNA-seq count data and identify differentially expressed genes between conditions. Genes were considered to be differentially expressed if they displayed an absolute fold change of 1.5 and a false discovery rate (FDR)-adjusted $P \leq 0.05$.

Data was visualised using the enhanced volcano package (v.1.6.0) (https://github.com/kevinblighe/EnhancedVolcano) in R studio (v.1.3.1073). (https://www.r-project.org/).

2.11.5. Additional bioinformatic and online tools

Table 2-24. Additional bioinformatic/online tools

Tool	Description/use/reference
Integrated microbial genomes & microbiomes (IMG/M)	Community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context (Chen et al., 2023)
MicrobesOnline	Comparative genomics website (Alm et al., 2005)
RegulonDB	Primary database on transcriptional regulation in *E. coli* K-12 (Tierrafría et al., 2022)
BioCyc	Collection of pathway/genome databases for model eukaryotes and prokaryotes (Karp et al., 2019)
Bprom	Bacterial promoter prediction (Solovyev V and Salamov A., 2011)
SignalP5.0	Prediction of the presence of signal peptides and the location of their cleavage sites (Almagro Armenteros et al., 2019)
ShinyGO 0.77	Graphical tool for gene enrichment analysis (Ge et al., 2020)
STRING	Protein-protein interaction networks and functional enrichment analysis (Szklarczyk et al., 2023)
ProtParam	Tool for the computation of various physical and chemical parameters for a given protein (Walker et al., 2005)
Protein Data Bank (PDB)	Global archive of experimentally determined 3D structures (Burley et al., 2019)
DeepTHMM	Deep learning model for transmembrane topology prediction and classification (Hallgren et al., 2022)
Jalview v.2.11.1.0	Free cross-platform for multiple sequence alignment editing, visualisation, and analysis (Waterhouse et al., 2009)

2.12. Statistical analysis

All statistical analyses were conducted in Prism v10, and *P*-values were determined using the stated statistical tests in each figure legend. Significance levels are shown on figures as *, **, ***, **** and NS, corresponding to $P < 0.05$, $P < 0.01$, $P < 0.001$, $P < 0.0001$ and no significance.

3. Bioinformatic, genetic and functional analysis of ROD_24811 and its associated genes in *C. rodentium* and homologues in EHEC

3.1. Introduction

It has been previously shown that ROD_24811, a predicted periplasmic binding protein (PBP) in *C. rodentium*, is significantly upregulated during murine infection. PBPs, also known as solute-binding proteins, are key components of ABC transporters (Maqbool et al., 2015).

ABC transporters constitute the largest and most ancient protein superfamily, being responsible for the transport of a diverse range of substrates across biological membranes (Thomas & Tampé, 2020). Structurally, ABC transporters comprise two cytoplasmic nucleotide-binding domains (NBDs) and two variable transmembrane domains (TMDs) spanning the membrane, displaying a modular architecture (Ter Beek et al., 2014). To drive the transport of substrates across membranes, whether that be inward or outward, ATP binding and hydrolysis is required (Srikant, 2020). Specifically, it is within the NBD that ATP is bound and hydrolysed to generate the sufficient energy for substrate translocation (Srikant, 2020). In Gram negative bacteria, the PBPs which interact with the TMDs of ABC transporters by delivering the target substrates, typically exist as untethered diffusible proteins (Ortega et al., 2022). Subsequently, the specificity of these proteins is not restricted to a single substrate and can instead bind more than one target or be recognised by more than one transporter (Ter Beek et al., 2014). For example, YtfQ is a PBP in *E. coli* K-12 able to bind both D-galactose and L-arabinose (Horler et al., 2009). Moreover, some PBPs act to stimulate a variety of signalling proteins such as chemoreceptors, sensor kinases and diguanylate cyclases/phosphodiesterases (Ortega et al., 2022). The significant upregulation of ROD_24811 therefore suggests a role in transporting, or at least responding to, nutrients present during infection. These nutrients may be important for host colonisation by *C. rodentium*, as well as in regulating the virulence programme of the pathogen. This is substantiated by recent evidence to suggest that ABC transporters are essential in the virulence of various clinically important pathogens such as *E. coli* (Akhtar & Turner, 2022; Tang & Saier, 2014).

This chapter aims to bioinformatically characterise ROD_24811 and associated genes. Additionally, using bacterial genetics the physiological role of ROD_24811 in *C. rodentium* strain ICC168 was explored to better understand the reasons for its significant upregulation during infection. Furthermore, a search for homologues in EHEC identified a novel locus and its carriage across the *E. coli* phylogeny was assessed.

3.2. ROD_24811-61 encodes a predicted ABC transporter and associated enzymes

Initial investigations aimed to provide support for the annotation of ROD_24811 as a PBP using *in silico* methods. As PBPs are found in the periplasm of the cell, these proteins must be translocated from the cytoplasm where they are expressed, across the IM. The process of protein translocation typically relies on the presence of short signal peptide sequences that target extracytoplasmic proteins for secretion. ROD_24811 was predicted to carry a signal peptide of the Sec/SPI type, with a cleavage site located between residues 24 and 25, supporting the likelihood that this protein is periplasmic. Furthermore, modelling of the protein using AlphaFold2 revealed ROD_24811 to have a bilobal appearance, characteristic of PBPs (**Figure 3-1A**). A search of the protein in InterPro also returned distinct domains found in PBPs such as the SPB_2_domain (IPR025997), corresponding to PBP family 2. Though unintegrated, signatures for D-xylose binding PBPs and LsrB QS-like PBPs were also identifiable. In support of the LsrB QS-like signature, a BLASTP search of ROD_24811 returned hits for PBPs specific for the QS molecule, AI-2, in other *Citrobacter* spp. (> 90 % identity).

As PBPs are usually associated with transporters or receptors as part of a signalling pathway, the genomic context of ROD_24811 was explored to decipher which genes were found in the same neighbourhood. Located at position 2610245 to 2617029 on the chromosome, ROD_24811 clustered with the genes annotated ROD_24821, 24831, 24841, 24851 and 24861 (**Figure 3-1BC**). Collectively, the six genes including ROD_24811 spanned 6.7 Kb, forming an apparent operon with a G/C content of 58.7 %. The KEGG database was next used to gather functional information on the proteins predicted to be encoded by ROD_24821-61. ROD_24821-41 with ROD_24811 were predicted to encode the components of a simple sugar transport system. ROD_24821 and ROD_24831/41 were predicted to encode an ATP binding protein and permease proteins, respectively (**Figure 1-3B**). Modelling of ROD_24811-41 in AlphaFold2 structurally supported the inferred functions of encoded gene products, displaying a modular structure and conserved folding pattern, commonly observed for ABC transporters (**Figure 3-1D**). Assessment of the associated genes, ROD_24851 and 24861, revealed them to encode a predicted putative carbohydrate kinase and isomerase, respectively (**Figure 3-1B**). These findings were indicative of ROD_24851

and 24861 having metabolically associated functions, as opposed to transport roles, predicted for ROD_24811-41.

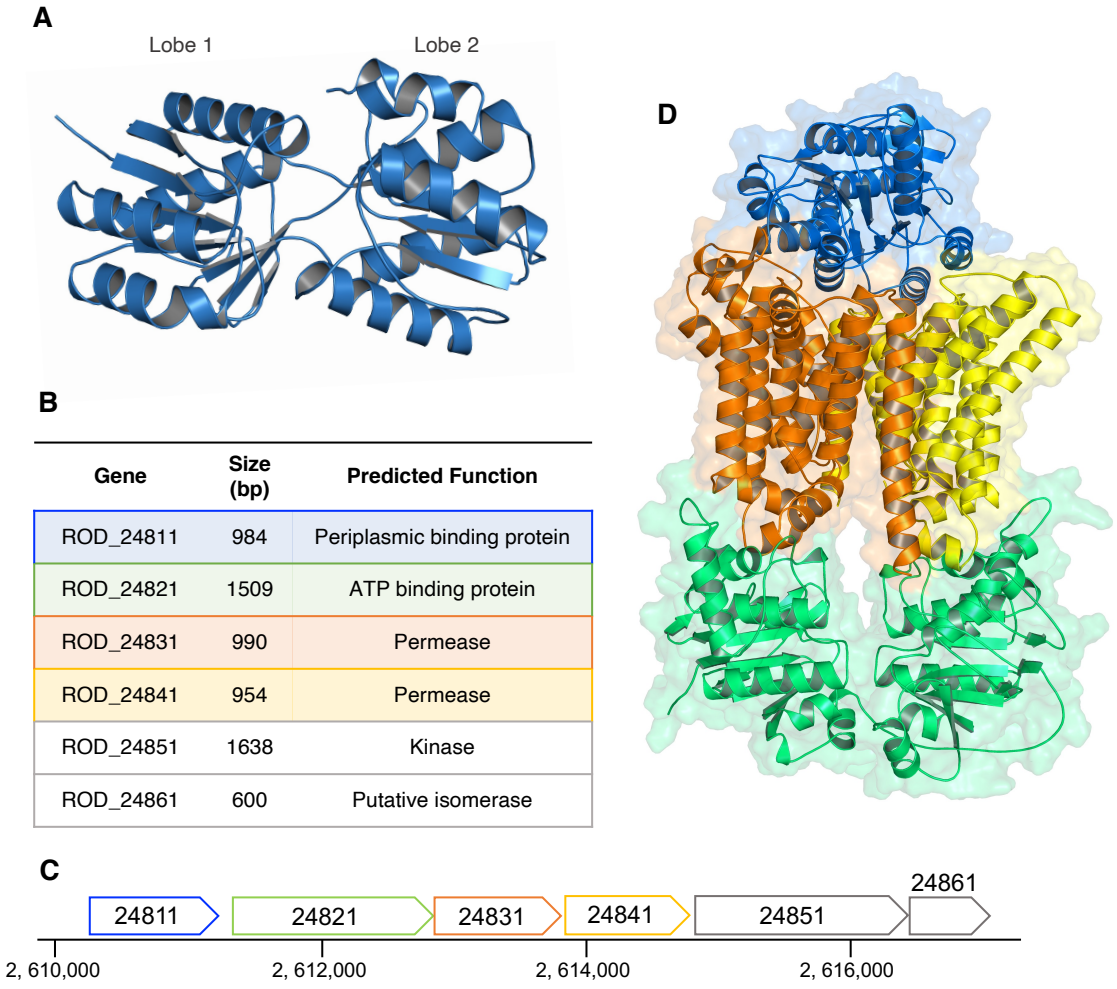


Figure 3-1. ROD_24811-61 in *C. rodentium* ICC168 encodes a predicted ABC transporter and associated enzymes. (**A**) AlphaFold2 model of ROD_24811; (**B**) Size and predicted functions for each gene of the ROD24811-61 locus; (**C**) Arrangement and chromosomal positions of ROD_24811-61; (**D**) AlphaFold2 model of the predicted ROD_24811-41 ABC transporter, highlighting periplasmic binding protein (PBP; blue), permeases (Orange, yellow) and ATP binding protein (Green). Colours also correspond with the annotations in panel **B**. Models were generated using AlphaFold2 and images produced in PyMol.

3.3. Transcriptional responsiveness of ROD_24811-61 to sugar substrates

The bioinformatic analysis supported the hypothesis that ROD_24811-61 encodes the apparatus for the uptake and metabolism of an unknown substrate in *C. rodentium*. All six genes of the locus were computationally predicted to lie within the same operon with high confidence (**Table 3-1**). To determine the transcriptional responsiveness of the locus to potential substrates a luminescence (LUX)-based reporter system that allows promoter activity to be measured in response to specific stimuli was used.

In brief, the LUX-promoter fusion construct was generated by fusing the promoter region of ROD_24811 upstream of the *luxCDABE* operon from *Photorhabdus luminescens* in pMK1*lux*. The genes of the *luxCDABE* operon encode a LCFA reductase (*luxC*), transferase (*luxD*) and synthetase (*luxE*) to form the FA reductase complex, as well as a heterodimeric bacterial luciferase (*luxAB*) responsible for the emission of blue-green light (**Figure 3-2**). By cloning a promoter of interest into the multiple cloning site of pMK1*lux* the expression of the promoter can be measured as a bioluminescent readout.

Table 3-1. *In silico* operon predictions for ROD_24811-61. The probability that both genes lie in the same operon (bOp), estimated probability that the pair are in the same operon (pOp) and separated distance between gene pairs (Sep) predict ROD_24811-61 to lie in the same operon. pOp values close to 1 indicate high confidence that the genes lie in the same operon.

Gene pair	bOp	pOp	Sep
24811-21	TRUE	0.924	106
24821-31	TRUE	0.997	-7
24831-41	TRUE	0.998	2
24841-51	TRUE	0.874	17
24851-61	TRUE	0.961	-3

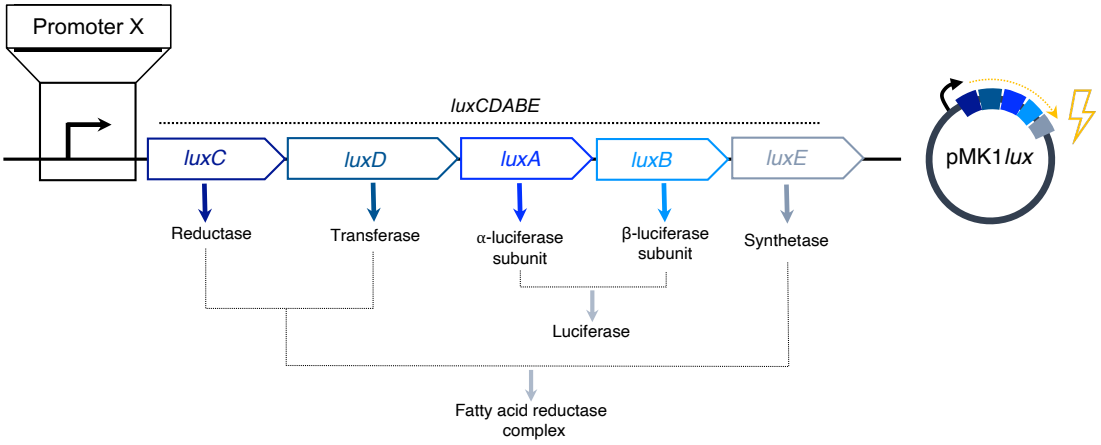
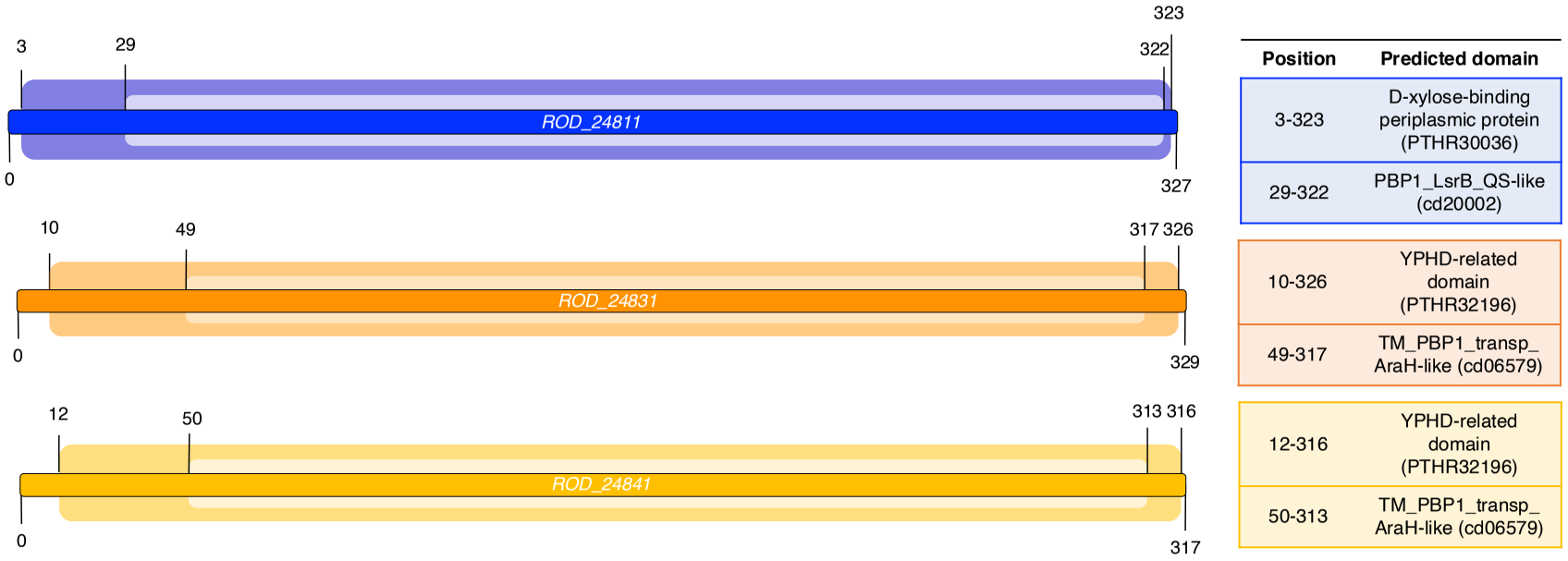


Figure 3-2. Schematic of the Lux operon of *P. luminescens* carried on pMK1lux. The Lux operon encodes five enzymes that allow for bioluminescence. Following the introduction of a promoter of choice (Promoter X), its activity can be measured in response to various stimuli via the bioluminescent output.

A 300 bp stretch of sequence immediately upstream of ROD_24811 was cloned into pMK1*lux* (pMK1*lux*-P_{24811}) as the region was likely to encompass the native promoter. Promoter activity was then measured following exposure to various sugar substrates via the bioluminescent readout generated as a result of the operon being expressed. Based on the initial KEGG annotation that ROD_24811 was specific for D-ribose, the sugar was selected as the first candidate for testing against the reporter system. In addition, D-xylose and L-arabinose were also selected as sugars to be tested based on the collective Interpro search results for ROD_24811, 24831 and 24841 (**Figure 3-3**). Identified domains within the proteins predicted to be encoded by ROD_24831 and 2841 were the same (**Figure 3-3**). Notably, all three of these sugars fall within the same class and are aldopentose monosaccharides. Accordingly, activity from the LUX-promoter fusion construct for ROD_24811 was tested in the presence of the three sugars: D-ribose, L-arabinose, and D-xylose.



Position	Predicted domain
3-323	D-xylose-binding periplasmic protein (PTHR30036)
29-322	PBP1_LsrB_QS-like (cd20002)
10-326	YPHD-related domain (PTHR32196)
49-317	TM_PBP1_transp_ AraH-like (cd06579)
12-316	YPHD-related domain (PTHR32196)
50-313	TM_PBP1_transp_ AraH-like (cd06579)

Figure 3-3. Domain analysis of ROD_24811, 24831 and 24841 in *C. rodentium* ICC168. InterPro searches highlighted the presence and location of several domains within the predicted PBP (ROD_24811) and permeases (ROD_24831 and 24841) of a predicted ABC transporter. The position and predicted domain type are highlighted accordingly.

C. rodentium carrying pMK1*lux*-P_{24811} was grown in MEM-HEPES supplemented with 0.5 mg/mL of either L-arabinose, D-ribose, or D-xylose. Comparison to growth in MEM-HEPES alone revealed no differences in the RLU (raw luminescence/OD_{600}) for D-xylose and L-arabinose, while only a modest increase in promoter activity was observed for D-ribose (**Figure 3-4**).

To better infer what the ROD_24811-61 system might be specific for, attention was turned to the annotation and predicted function of metabolically associated genes, ROD_24851 and 24861. Both ROD_24851 and 24861 are predicted to be found in the cytoplasm and had no identifiable signal peptide to suggest translocation across the membrane. ROD_24851 was predicted to belong to the family of pentulose kinases (IPR006003), and specific for D-ribulose. Modelling of the protein using the Phyre2 engine provided structural support for the encoded protein being a ribulokinase. The model highlighted two domains of different sizes (I and II) commonly seen for other ribulokinases (**Figure 3-5A**). When overlaid with the X-ray crystal structure of the only characterised ribulokinase (AraB; 3QDK) in the PDB, conservation in the architecture could also be observed (**Figure 3-5B**). However, sequence alignment revealed the two proteins share only 28 % identity at the amino acid level. As the X-ray crystal structure of AraB used in the analysis had been determined with L-ribulose bound, residues important for ligand-binding were sought in ROD_24851. The residues W129, K208, D274 and E329 of AraB required for interaction with L-ribulose (via hydrogen bonds), were completely conserved in ROD_24851 when aligned (**Figure 3-5CD**). These residues were also identified to lie in proximity of the binding cleft between the two domains of the ROD_24851 model, as also seen for AraB. Though a final residue (A96) was not conserved in ROD_24851, T81 was identified via sequence alignment and may act as substitute for A96. However, due to the pronounced differences in the proposed location of T81 compared to A96, relative to the substrate binding pocket of both models, it seemed unlikely that T81 would be involved in substrate interaction. Instead, an alternative alanine residue elsewhere in ROD_24851, unable to be identified via sequence alignment, may localise to the substrate binding pocket once folded. The ATP-binding motif (GGLPQK) within the small domain (I) of AraB was only found to be fully conserved in 3/6 residues in ROD_24851 when aligned. However, a search of ROD_24851 in InterPro similarly recognised this region as a predicted ATP-

binding site. In parallel, a search of ROD_24861 in InterPro scan returned several predicted domains, including a KpsF-like SIS-domain (IPR035474). In *E. coli*, KpsF is a D-arabinose-5-phosphate isomerase that catalyses the conversion of ribulose-5-phosphate to D-arabinose-5-phosphate.

Since the associated kinase for the locus was predicted to be specific for D-ribulose, *C. rodentium* with pMK1*lux*-P_{24811} was grown in the presence of D-ribulose. Supplementation of 0.5 mg/mL D-ribulose led to significantly higher promoter activity compared to the media only control and supplementation with D-ribose, suggesting that this may be the physiologically relevant substrate for the system (**Figure 3-4**).

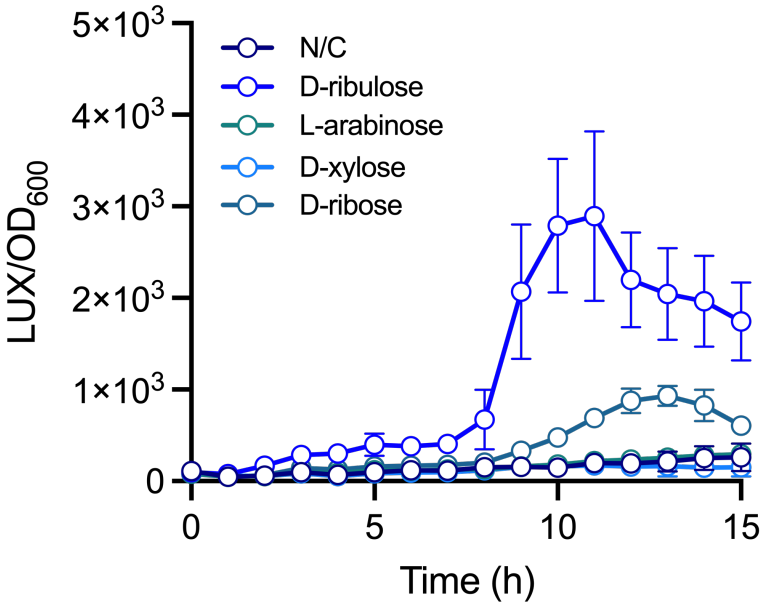


Figure 3-4. ROD24811-61 promoter activity is induced by D-ribulose. WT ICC168 was grown in MEM-HEPES supplemented with either 0.5 mg/mL D-ribulose, L-arabinose, D-xylose, or D-ribose. ROD_24811-61 promoter activity was measured over 15 h as an output of luminescence (LUX) and normalised against the OD_{600}. MEM-HEPES alone (containing 0.9-1.1 mg/mL D-glucose as the carbon source) was used as a control. $n = 3$ biological replicates and error bars indicate SD.

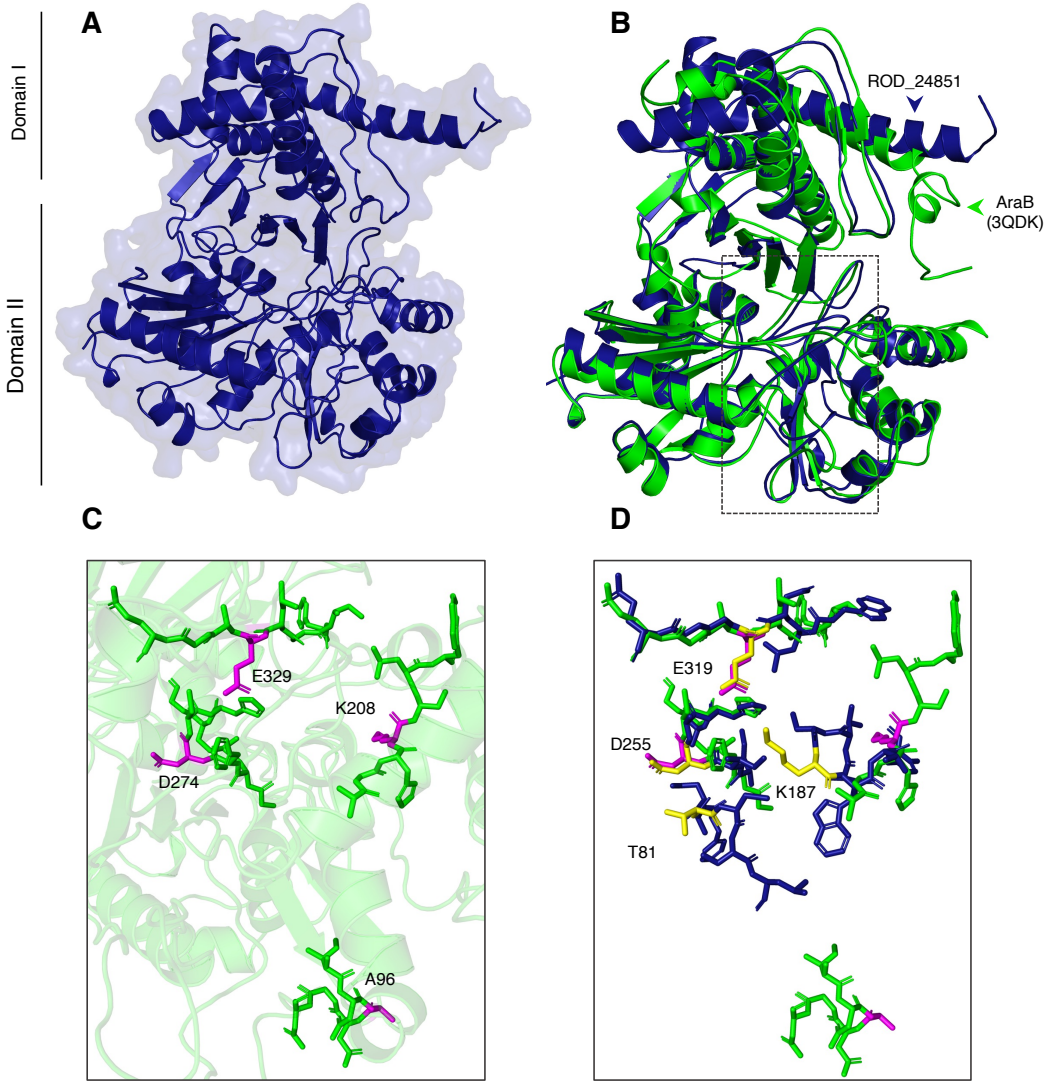


Figure 3-5. Homology model of ROD_24851 in *C. rodentium* ICC168. (**A**) Phyre2 generated homology model of ROD_24851, comprising two domains of different sizes. Model was generated using c3gg4B as a template; (**B**) Overlay of ROD_24851 model (Blue) with the crystal structure of L-ribulokinase (AraB) from *Bacillus halodurans* (Green), retrieved from the PDB (3QDK); (**C**) Residues known to interact with L-ribulose (Magenta) in AraB from *Bacillus halodurans*; (**D**) Overlay of residues involved in interaction with L-ribulose in AraB from *B. halodurans* (Magenta) with the residues of ROD_24851 (Yellow). All images were produced in PyMol.

3.4. Role of ROD_24811-51 in *C. rodentium* growth on D-ribulose

As D-ribulose was found to induce ROD_24811 promoter activity strongly, it was hypothesised that the encoded proteins would allow for uptake of the sugar and subsequent growth. The contribution of ROD_24811-61 to fitness when grown on D-ribulose as a sole carbon source was therefore assessed. A ΔROD_24811-51 mutant was generated using the Lambda Red approach (**Figure 3-6A**), and its growth profile in M9 minimal media with D-ribulose as a sole carbon source was compared to the WT (**Figure 3-6B**). Deletion of ROD_24811-51 was found to completely abolish the ability of *C. rodentium* to grow on D-ribulose (0.5 mg/mL). In addition, there was no significant growth differences between the WT and ΔROD_24811-51 when grown in media supplemented with D-ribose (**Figure 3-7**), despite modest pMK1*lux*-P_{24811} expression being observed previously (**Figure 3-4**). These findings suggested the locus to have an essential role in the utilisation of D-ribulose, and herein the locus will be referred to as '*rbl*'.

To next determine whether it was the complete absence of *rbl* that led to an inability to grow on D-ribulose, or whether the phenotype was attributable to the deletion of a single gene carried by the locus, independent deletions were made in each of the encoded components (**Figure 3-6A**). Deletion of the genes for the predicted ABC transporter (24811-41) and ribulokinase (24851) had the greatest effect on the ability of *C. rodentium* to grow on D-ribulose (**Figure 3-6B**), with no growth being observed for both mutants. Deletion of ROD_24861 did not contribute to any major growth defects on D-ribulose and displayed a growth profile essentially identical to the WT. These data therefore suggest the ABC transporter encoded by ROD_24811-41 is the only route of D-ribulose uptake in *C. rodentium*, making it indispensable for growth on the sugar. Furthermore, even when able to uptake D-ribulose, absence of the associated predicted kinase prevents growth, validating its role as an integral component of this operon.

The inability of Δ*rbl* to grow on D-ribulose was validated by complementation. The complete *rbl* locus was cloned into pSUPROM under the constitutive control of the TAT promoter (pSU-*rbl*) and transformed into Δ*rbl*. When grown on D-ribulose, the Δ*rbl* mutant with pSU-*rbl* was restored in its ability to grow on the sugar as a sole carbon source, similar to WT levels (**Figure 3-6B**).

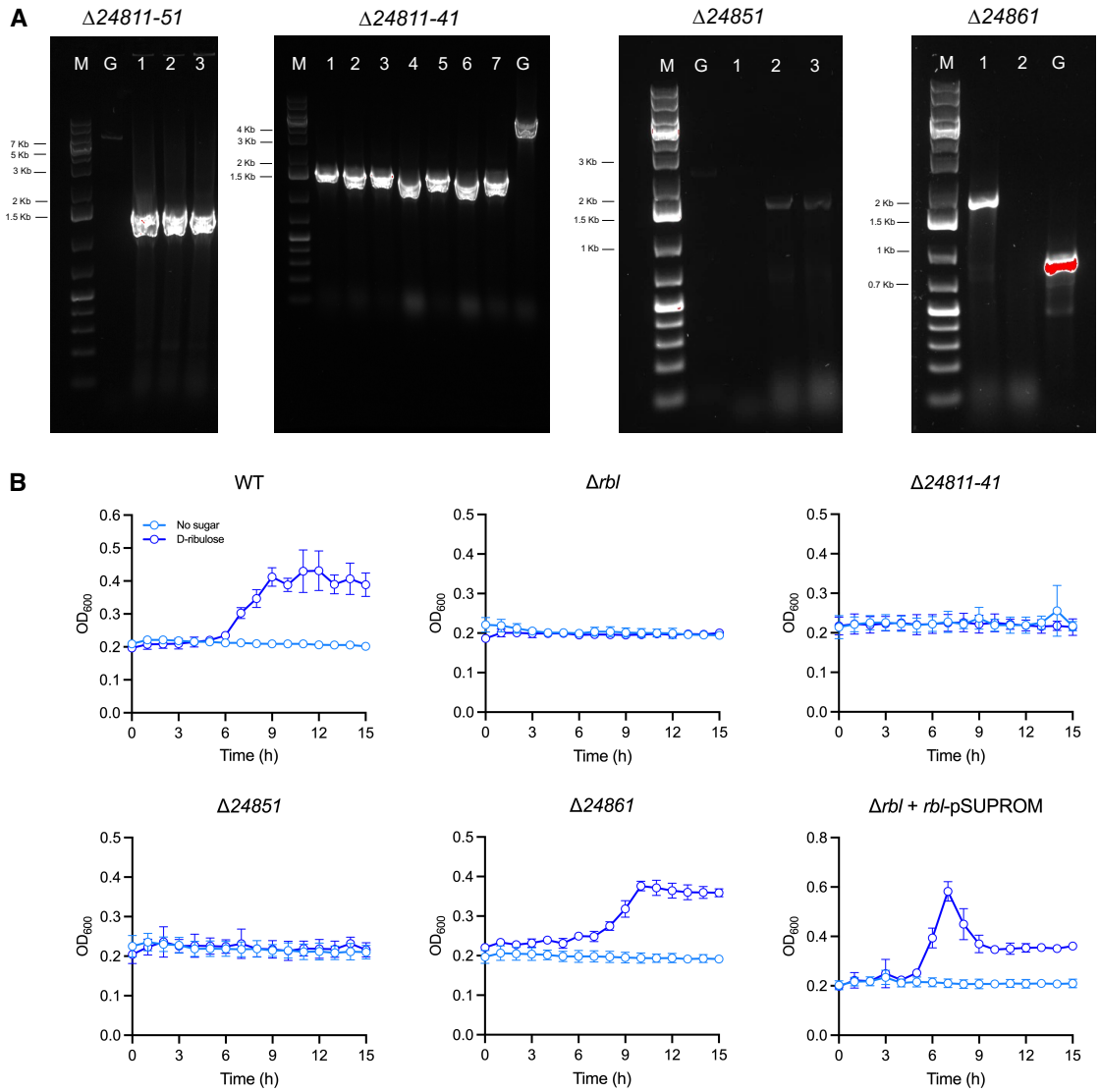


Figure 3-6. The Rbl system (ROD_24811-51) in *C. rodentium* ICC168 is responsible for growth on D-ribulose. (**A**) Gel electrophoresis confirmation for each Rbl system mutant. Lane M corresponds to a 1 Kb plus DNA ladder. Lane G corresponds to the WT gDNA control; (**B**) Growth profiles for WT and corresponding Rbl system mutants in ICC168, grown in M9 minimal media supplemented with or without 0.5 mg/mL D-ribulose over 15 h. *n* = 3 biological replicates and error bars indicate SD.

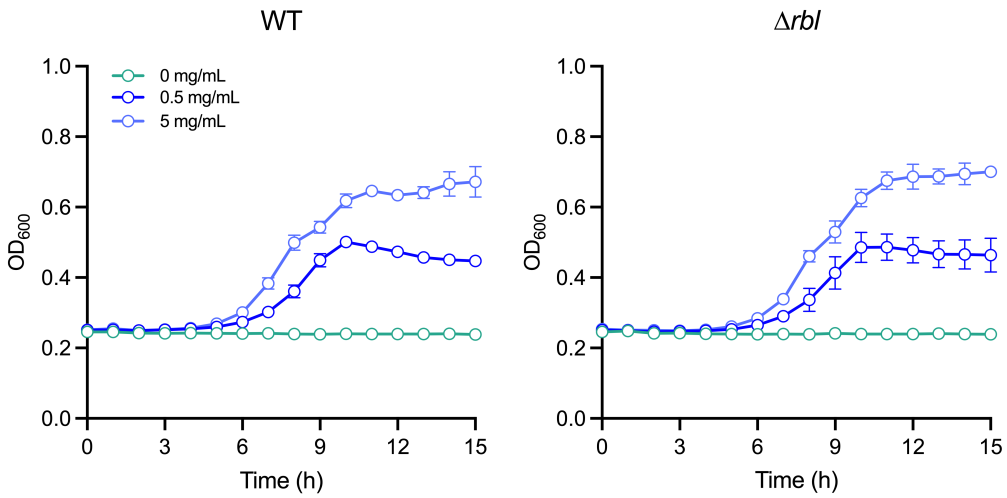


Figure 3-7. Growth profile of *C. rodentium* on D-ribose as a sole carbon source. Growth profiles for WT and Δ*rbl* when grown in M9 minimal media supplemented with 0, 0.5 and 5 mg/mL D-ribose over 15 h. *n* = 3 biological replicates and error bars indicate SD.

3.5. Analysis of *rbl* carriage across the *Citrobacter* spp.

To explore whether *rbl* was exclusive to *C. rodentium,* carriage of the locus across 18 candidate species within the *Citrobacter* genus was assessed (**Figure 3-8A**). ROD_24861 was not included as part of these analyses as the gene was already annotated as a gene conserved across *Citrobacter* spp. A BLASTN search of candidate *Citrobacter* spp. revealed the locus to be present across 14 species; with at least one strain displaying ≥ 80 % similarity at the nucleotide level for each gene. Hits for ROD_24811 and 24831 in *C. sedlakii* had > 90 % sequence similarity. In contrast, no hits for ROD_24811-51 were returned for *C. koseri*, *C. gillenii*, *C. diversus* and *C. murliniae*.

In *C. rodentium* and other species positive for ROD_24811-51 carriage, the locus was located between genes predicted to encode a putative export protein and stationary inducible protein (*csiE*) (**Figure 3-8B**), which were ROD_24791 and ROD_24871 in *C. rodentium*, respectively. Comparison with the same region in *C. koseri* (negative for the locus) confirmed absence of the locus within the region of the equivalent genes (CKO_00245, CKO_00246). The similar GC content of the locus and flanking genes (~58 %) suggested the system may have been lost, rather than gained, from a common ancestor.

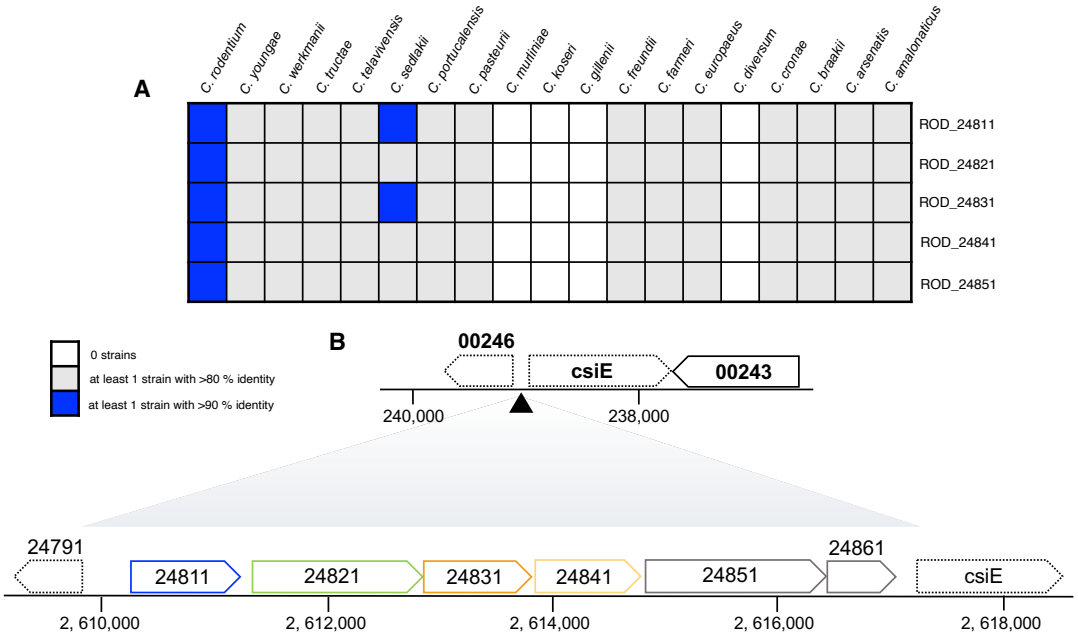


Figure 3-8. Schematic overview of ROD_24811-51 carriage across multiple *Citrobacter* spp. (**A**) Carriage analysis across 19 *Citrobacter* spp. (including *C. rodentium*). Carriage of a gene within a species was assumed when a homologue of the *C. rodentium* system was present in at least 1 strain with ≥ 80 % nucleotide sequence identity. (**B**) Location of ROD_24811-61 in strains carrying the locus compared to strains not carrying the locus.

3.6. Carriage of ROD_24851-61 outside of *Citrobacter*

As ROD_24811-61 appeared to be largely present across the *Citrobacter* genus, its carriage amongst the broader Enterobacteriaceae was investigated (**Table 3-2**). The entire locus was found to be carried by at least one representative strain of the following genera: *Escherichia*, *Yokenella*, *Cedecea*, *Kosakonia*, *Pluralibacter* and *Klebsiella*. Loci displaying the greatest percentage identity of > 85 % belonged to *Pseudescherichia*. The loci of *Escherichia, Yokenella*, *Cedecea*, *Kosakonia* and *Pluralibacter* displayed < 85 % (but more than > 80 %) sequence identity to that of the *C. rodentium* system. Although displaying lower percentage identity, the greatest number of strains found to be positive for carriage of the locus belonged to *Klebsiella* spp. (*K. variicola, K. pneuomoniae*).

Table 3-2. Overview of ROD_24811-61 carriage outside of the *Citrobacter* genus. The entire ROD_24811-61 locus was blasted against Enterobacteriaceae excluding *Citrobacter*. The top 100 returned hits were summarised and arbitrarily grouped into moderate (> 70 %), high (> 80 %) and very high (> 85 %) sequence identity with the *C. rodentium* system. The frequency of strains within a species included in the analysis were also calculated.

Genus	Species	Strains	Identity
Pseudescherichia	*vulneris*	2	>85%
Escherichia	*coli*	1	>80%
Yokenella	*regensburgei*	1	
Cedecea	*neteri*	4	
	lapagei	1	
Kosakonia	*oryzae*	2	
	radicincitans	5	
	cowanii	5	
	sacchari	3	
	pseudosacchari	2	
	arachidis	1	
	oryzendophytica	1	
Pluralibacter	*gergoviae*	2	
Klebsiella	*variicola*	12	>70%
	pneuomoniae	43	

3.7. Identification of a ROD_24811 homologue in EHEC str. EDL933

As previously described, *C. rodentium* is a murine pathogen used to model EHEC infections due to their overlapping pathogenic mechanisms. A search for homologues of ROD_24811 in EHEC O157:H7 str. EDL933 was conducted and led to the identification of Z0415. Z0415 shares > 60 % identity at the nucleotide level with ROD_24811 and is similarly predicted to encode a PBP of a simple sugar transport system (**Figure 3-9**). This was similarly reflected at the protein level, with ROD_24811 and Z0415 sharing 60 % sequence identity.

On the chromosome, Z0415 clustered with the genes annotated Z0416, Z0417, Z0418 and Z0419 (**Figure 3-9**). Searches in KEGG revealed the genes to encode a predicted ATP-binding protein component, ATP-binding protein, and permease proteins (**Figure 3-9A**). Whilst the size of genes predicted to encode the permease proteins (Z0418, Z0419) were approximately similar to those of the Rbl system in *C. rodentium*, the predicted ATPase was encoded across two genes of different sizes (261 bp and 1179 bp) (**Figure 3-9B**). Collectively, the size of these genes when summed were equivalent to the ATPase of the Rbl system. Alignment of the amino acid sequences in MUSCLE, revealed approximately 60 % sequence identity for each of the predicted proteins across EHEC and *C. rodentium*. Due to the absence of any associated genes predicted to encode a carbohydrate kinase and isomerase, the locus of EDL933 was found to be significantly smaller compared to the ICC168 locus.

A

Gene	Size (bp)
Z0415	987
Z0416	261
Z0417	1179
Z0418	972
Z0419	957

B

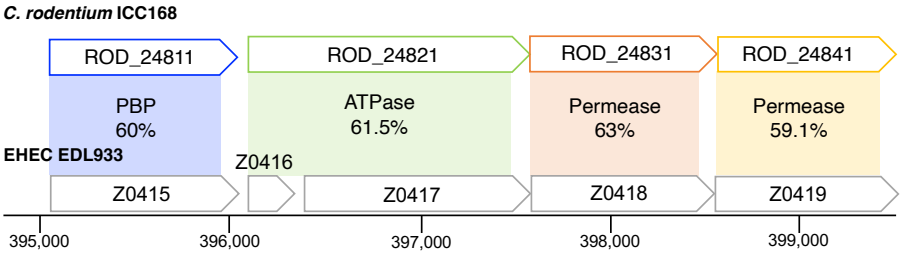


Figure 3-9. Z0415 is a homologue of ROD_24811 in EHEC O157:H7 str. EDL933. (**A**) ROD_24811 shares > 60 % sequence identity with Z0415 at the nucleotide level. (**B**) Chromosomal context of Z0415 in EDL933. Z0415 clusters with four additional genes: Z0416, Z0417, Z0418 and Z0419, spanning a total of 4.4 Kb.

A more in-depth analysis of the proteins predicted to be encoded by Z0415-19 revealed several details that supported their probable role in forming an ABC transporter. This included the identification of a signal peptide (Sec/SPI) between residues 27 and 28 of Z0415, supporting its localisation in the periplasm. Similarly, the AlphaFold2 model of Z0415 displayed the same characteristic bilobal appearance as ROD_24811 (**Figure 3-910AB**). Additionally, a topology search of Z0418 and Z0419 in DeepTMHMM revealed the two predicted permease proteins to both comprise of 10 (α) helices, characteristic of Type II ABC transporters. When sought, characteristic domains of the ATPase were also identifiable, including a RecA-type core, ABC-specific 3-stranded ß sheet (ABCß) and α-helical subdomain (ABCα) (**Figure 3-10C**). The overall model of Z0415-19 displayed similarity to that of ROD_24811-41 and had a distinctive ABC transporter-like appearance (**Figure 3-10D**).

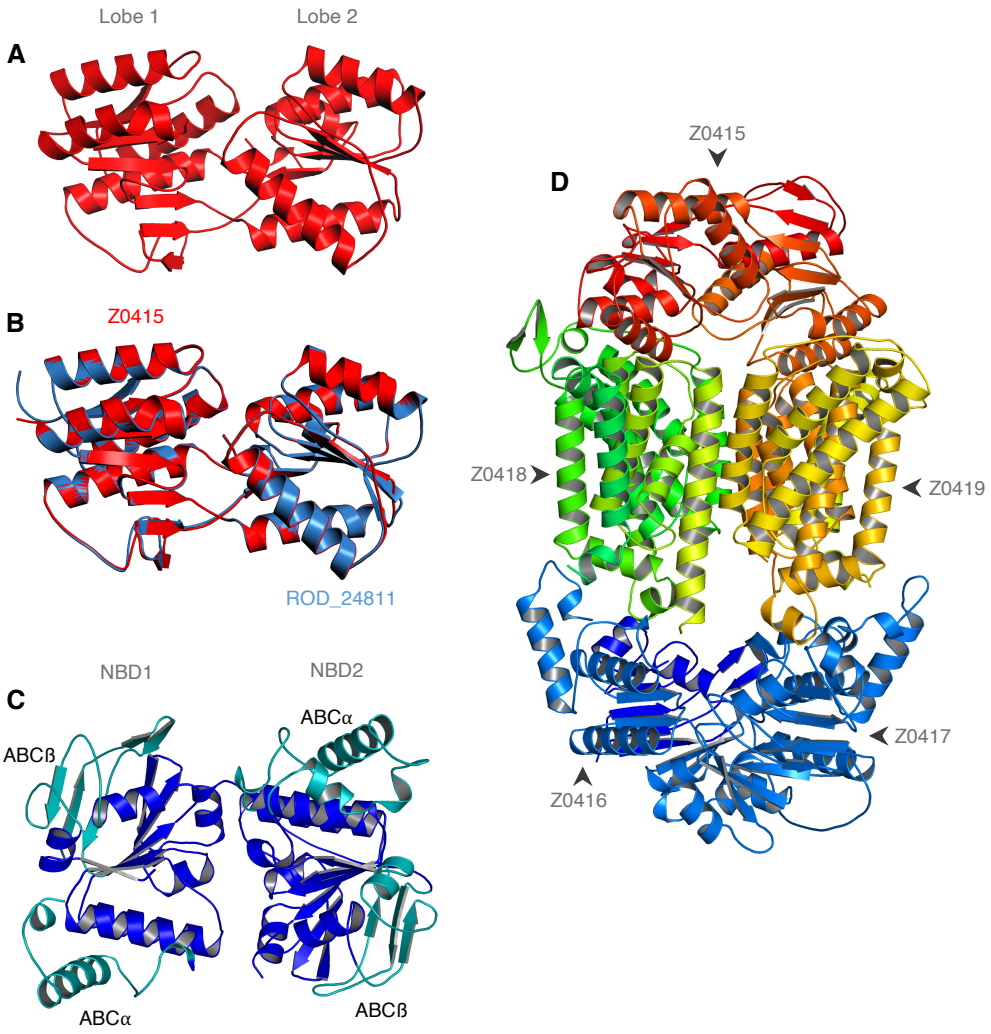


Figure 3-10. Homology model of Z0415-19 in EDL933. (**A**) Representative model of Z0415; (**B**) Overlay of Z0415 (red) and ROD_24811 (blue) in *C. rodentium* ICC168; (**C**) Representative model of the ATPase encoded by Z0416 and Z0417 in EDL933, comprising two NBDs. Each domain consists of a single RecA-like portion (blue) surrounded by an ABC-specific 3-stranded ß sheet and α-helical subdomain (cyan), annotated as ABCß and ABCα respectively; (**D**) Representative model of the Z0415-19 ABC transporter. All models were generated using AlphaFold2 and images produced in PyMol.

3.8. Genomic context of Z0415-19 in EHEC str. EDL933

In the context of the EDL933 genome, the Z0415-19 locus spanned 4.4 Kb and was located to a unique region of the chromosome not found in K-12 strains, termed an O-island. Only one other gene was identified to be carried on OI-17 with Z0415-19: Z0414. Z0414 was found to lie 251 bp upstream of Z0415 and predicted to encode a hypothetical protein belonging to the DUF984 family, with only a single ASCH domain identifiable. ASCH domains are speculated to have roles in RNA processing during translation, and not transport, therefore Z0414 was hypothesised to be independent of Z0415-19. Computational predictions supported these claims by revealing the gene to lie outside of the Z0415-19 locus. Comparisons between the EDL933 and MG1655 genome localised the OI to between the genes *yahJ* and *yahK* (**Figure 3-11**), which are predicted to encode a putative deaminase and oxidoreductase, respectively. In MG1655, analysis of the intergenic region of *yahJ* and *yahK* revealed a 72 bp fragment homologous to the very 3′ end of Z0419 (including the STOP codon) (**Figure 3-11**), potentially highlighting a previous site of homologous recombination.

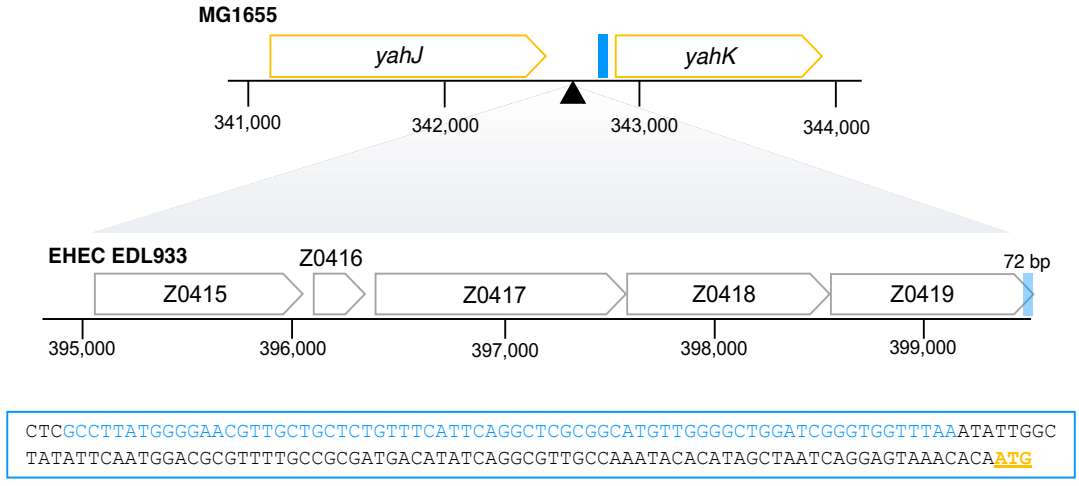


Figure 3-11. Z0415-19 is located to an EDL933 specific OI. Z0415-19 is located on OI-17 between *yahJ* and *yahK* in EDL933 but absent *in E. coli* MG1655. A 72 bp nucleotide stretch (blue) located within the intergenic region of *yahJ* and *yahK* is homologous to the very 3′ end of Z0419 in EDL933. The start codon of *yahK* is highlighted in yellow.

3.9. Phylogenomic analysis of Z0415-19 carriage across the *E. coli* species

A phylogenomic analysis of 948 *E. coli* (*n* = 933) and *Shigella* (*n* = 16) strains was carried out to determine the carriage of Z0415-19 across the *E. coli* phylogeny (**Figure 3-12**). The strains used for the analysis were well-representative of the generally accepted *E. coli* and *Shigella* clade structure making up phylogroups as follows: B2 (266/949, 28.03 %), B1 (227/949, 23.9 %), E (221/949, 23.3 %), A (148/949, 15.6 %), D (62/949, 6.5 %), and F (11/949, 1.2 %). The presence or absence of Z0415-19 across the sample set was investigated by reconstruction of a maximum likelihood phylogeny (**Figure 3-12**). Despite some carriage across the six phylogroups, the locus was found not to be completely conserved across all strains used.

Phylogroup E was predominantly positive for the carriage of Z0415-19 (Z0415-19^{+VE}), except for three strains: str. 2845650, str. KTE196 and KTE117 (**Figure 3-13**). Similarly, the majority of phylogroup B1 and D strains were found to be largely Z0415-19^{+VE} (**Figure 3-13**). Phylogroup B1 however appeared to split into two separate clades where the larger lineage of the two (based on these data) was Z0415-19^{+VE}, and the smaller Z0415-19^{-VE} (**Figure 3-12**). In stark contrast, phylogroup B2 was > 35 % negative for carriage of the locus, whilst phylogroup A was entirely Z0415-19^{-VE} all bar a single strain (str. HS) (**Figure 3-13**). None of the *Shigella* strains included in the analysis were found to be Z0415-19^{+VE}.

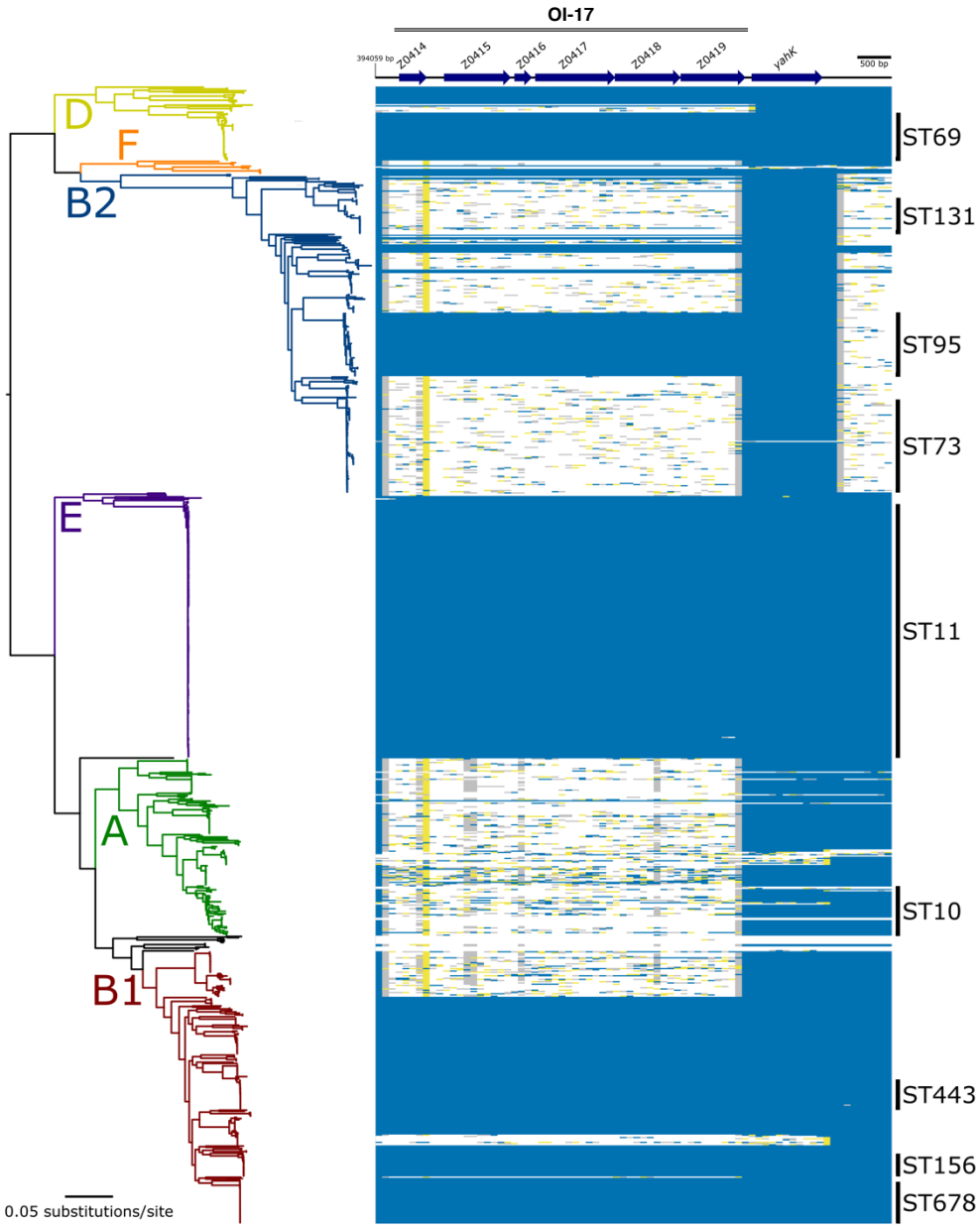


Figure 3-12. Carriage of Z0415-19 across the *E. coli* phylogeny. Maximum likelihood analysis built from 245,518 core-genome SNPs relative to the reference chromosome of EDL933. Phylogeny is rooted according to the actual root by *Escherichia fergusonii* ATCC 35469 (Omitted for visualisation). Branch colours indicate the six main phylogenetic groups. Branch lengths and scale bar represent number of nucleotide substitutions per site. The presence/absence analysis of loci is based on the uniform coverage at each 100 bp window size in SPANDx. Coverage is shown as a heat map where ≥ 80 % identity is highlighted in blue, ≥ 50 % identity is highlighted in yellow, and ≥ 1 % is highlighted in grey. White plots indicate regions that are absent. The Z0415-19 locus is indicated above the phylogeny.

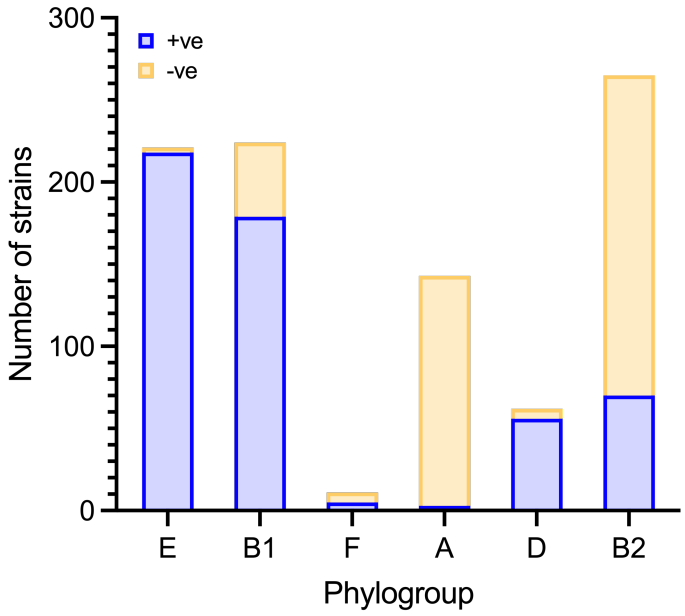


Figure 3-13. Overview of Z0415-19 carriage across the *E. coli* phylogeny. The presence (+ve) and absence (-ve) of Z0415-19 within each phylogroup was determined.

Phylogenetic clustering allows for the grouping of distinct *E. coli* strains, and therefore encapsulation of 'pathotypes'. As Z0415-19 carriage was not completely isolated to an individual phylogroup, analyses were conducted to address the question of whether this locus might be more greatly associated with pathogenic *E. coli* strains, and if so, whether there is an affinity for carriage of the locus by a defined pathotype (**Figure 3-14**). For both Z0415-19^{+VE} and Z0415-19^{-VE} strains, the InPEC pathotypes were sought, allowing for the identification of (a) EHEC/STEC/VTEC; (b) EPEC; (c) ETEC; (d) EAEC; (e) AIEC strains present in the dataset. Searches also allowed for the identification of commensal *E. coli* strains. Extra-intestinal pathotypes such as UPEC and MNEC, as well as non-human associated APEC strains, were more broadly grouped and fell into the 'other' category. For a total of 704 strains, accounting for 76.1 % (704/924) of the sample set, a pathotype was unable to be assigned and were therefore included in these analyses as 'unassigned'. However, for those which could be assigned a pathotype, Z0415-19^{+VE} strains largely associated with EHEC/STEC/VTEC, accounting for 51 % of the assignable Z0415-10^{+VE} strains. ETEC (25 %; 68/271 assignable Z0415-19^{+VE} strains) and EAEC (12.9 %; 35/271) strains had the next greatest association with Z0415-19, whilst a total of 5 (1.8 %) assignable strains were EPEC. Only a single strain assigned to the AIEC pathotype was Z0415-19^{+VE}.

Together, this suggested that Z0415-19 is more likely to be carried by enteric pathotypes.

In contrast, assignable Z0415-19^{-VE} strains were largely associated with the ETEC pathotype (104/393; 26.5 %). Fewer EAEC strains (2) were found to be Z0415-19^{-VE} than previously identified for Z0415-19^{+VE} strains, whilst a similar number of EPEC (5) and AIEC (1) strains were Z0415-19^{-VE} as they were Z0415-19^{+VE}. The occurrence of commensal *E. coli* strains being Z0415-19^{+VE} (8) or Z0415-19^{-VE} (10) were also similar. Only a single *E. coli* strain identified to be Z0415-19^{-VE} as part our analyses belonged to the EHEC/STEC/VTEC pathotype.

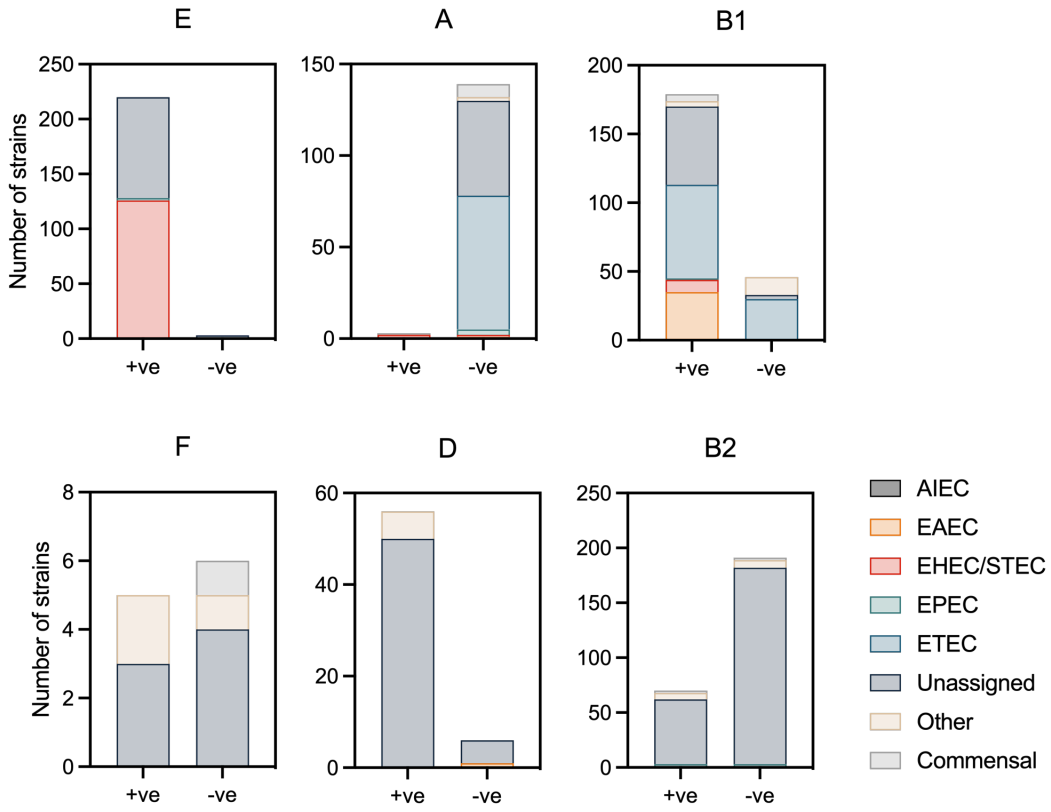


Figure 3-14. Pathotype association with Z0415-19 carriage across *E. coli*. Strains that were both positive and negative for the carriage of Z0415-19 across six *E. coli* phylogroups (E, A, B1, F, D, B2) were categorised by intra-intestinal pathotype. Extra-intestinal strains were defined as 'other', whilst those strains unable to be allocated to a pathotype were defined as 'unassigned'.

The occurrence of Z0415-19 outside of the genus *Escherichia* was next assessed to decipher whether it was carried by other members of the Enterobacteriaceae, specifically those that are pathogenic. A BLASTN search of the entire Z0415-19 nucleotide sequence was conducted to exclude *Escherichia* and is summarised by **Figure 3-15**. Although not identified as being positive for carriage of the locus in our previous phylogenomic analysis, Z0415-19 was identified to be present in 12 *Shigella* strains, spanning the species: *S. dysenteriae*, *S. sonnei* and *S. flexneri*. These strains displayed an extremely high nucleotide sequence identity of > 99 %, with that of the EDL933 locus. An additional *Shigella* strain, with species unknown was identified to share between 98 to 99 % similarity with Z0415-19. Although displaying lower (but still high) sequence identity, the locus was also present in three strains of *K. quasipneuomoniae*, and a single strain of *S. enterica*. Notably, in *K. quasipneuomoniae* the locus was shown to be carried on the pKPC plasmid, carried by carbapenem-resistant isolates. Furthermore, a BLASTN search for each individual genes of Z0415-19 returned a greater number of strains from various species and genera. Therefore, whilst these strains might carry homologues to the genes of Z0415-19, they do not necessarily always carry the entire locus. The identified homologues may instead belong to different systems, and possess alternative functions, to those of the EDL933 system.

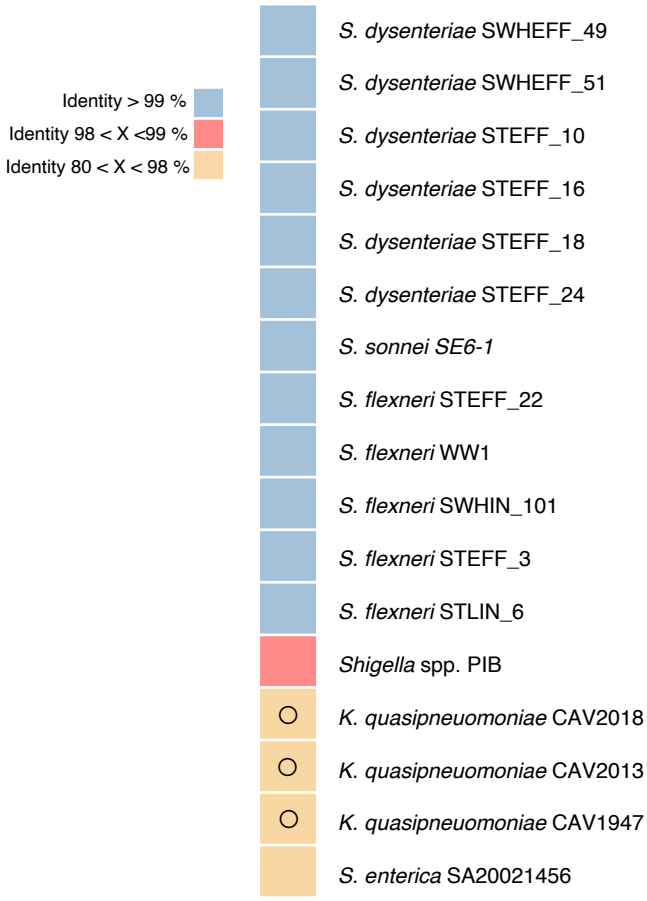


Figure 3-15. Carriage of the entire Z0415-19 locus outside of *Escherichia*. BLASTN search results for the entire Z0415-19 nucleotide sequence, excluding *Escherichia*. Strains were categorised based on having very high (> 99 %), high (> 98 %) and moderately high (> 80 %) percentage identity scores. Cells containing (O) correspond to those strains identified to carry the locus on extrachromosomal elements.

3.10. Analysis of ABC transporters in EHEC str. EDL933

A phylogenetic analysis was conducted to identify ABC transporters closely related to Z0415-19 and therefore allow further inferences on the transporter's substrate specificities. As the NBDs are a defining feature of ABC transporters, they are commonly used in the classification of such proteins. Rather than using the entire Z0416-17 sequences, the ABC_tran pfam domain (pfam00005) was used to restrict the analysis to a highly conserved core region of the NBDs of ABC transporters in EDL933. Specifically, Pfam00005 belongs to the P-loop NTPases common to ABC transporters. Protein sequences were retrieved from the Integrated Microbial Genomes (IMG) system using pfam00005 prior to phylogenetic analysis.

A total of 88 proteins with pfam_00005 were identified in EDL933 (**Figure 3-16**). A fewer number of proteins were returned with pfam_00005 for MG1655 (**Figure 3-16A**). Based on the clusters of orthologous groups of proteins (COGs), and the substrates that they are predicted to transport, proteins in EDL933 were grouped into 7 distinct categories: sugar, peptide, amino acid, multi-drug, metal/ions, anti-microbial and vitamin (**Figure 3-16B**). Additionally, proteins identified to have broader substrate specificities such as the transport of cholesterol, sulphate, taurine and microcins were grouped as 'other'. Those proteins with no COG allocated or designated as uncharacterised were grouped as 'unassigned' and 'uncharacterised' respectively. Proteins predicted to transport sugar, peptides, and metal/ions were found to be the most common in EDL933. Of the 12 proteins predicted to be involved in the transport of sugar, only 2 had unknown substrates, excluding Z0416/17.

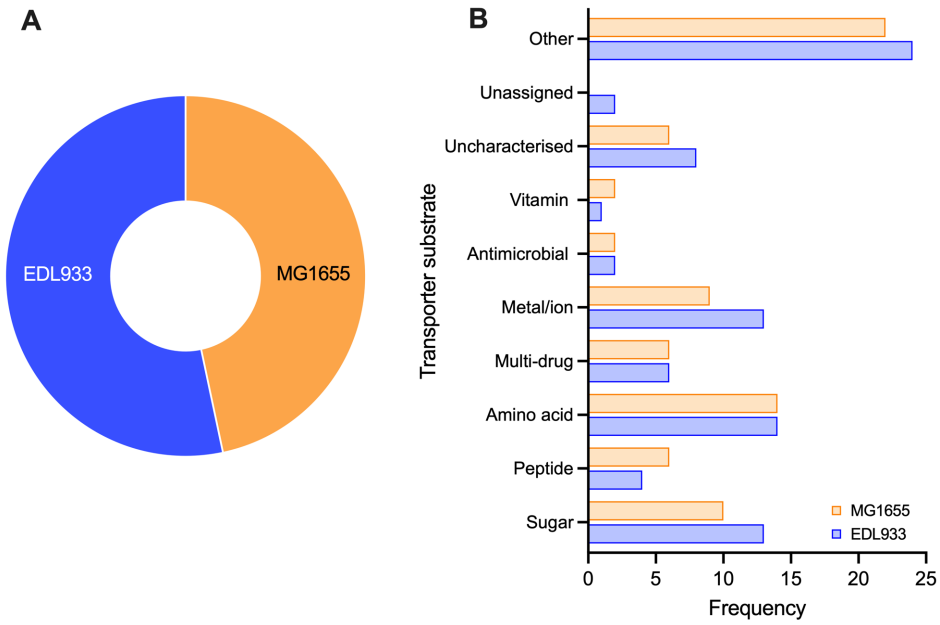
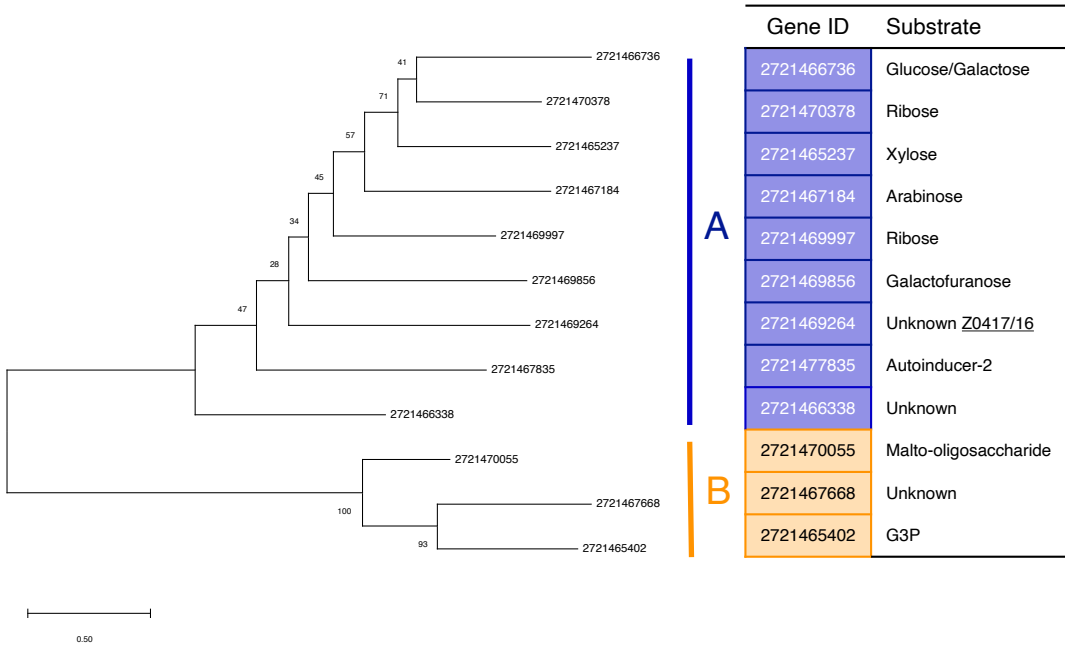


Figure 3-16. Comparative analysis of ABC transporters in EDL933 and MG1655. (**A**) The ABC_tran pfam domain (pfam00005) was searched in EDL933 and MG1655 to identify the number of ABC transporters present in their genomes. (**B**) Transporters were arbitrarily categorised based on their predicted substrates, using their assigned COG, for comparison between the two strains. Those with putative or unknown function were defined as 'uncharacterised', whilst those with no assigned COG were defined as 'unassigned'. Those transporters with known substrate specificity but unable to be allocated to one of the seven substrate groups, were defined as 'Other'.

The amino acid sequences of the 12 identified proteins predicted to transport sugars were aligned by MUSCLE and used to generate a phylogenetic tree (**Figure 3-17**). Proteins were found to split into two distinct clades (A & B) based on the sugar they were predicted to transport. Clade A was comprised of 9 proteins predicted to transport simple sugars, such as the monosaccharides D-ribose, L-arabinose, D-xylose, and D-galactose. Clade B was considerably smaller than clade A and comprised only 3 proteins. The proteins of clade B appeared to transport more complex substrates such as a malto-oligosaccharide and modified alcohol, glycerol-3-phosphate. Z0416/17 was found to belong to Clade A, supporting previous bioinformatic inferences that Z0415-19 encodes a simple sugar transport system.



Gene ID	Substrate
2721466736	Glucose/Galactose
2721470378	Ribose
2721465237	Xylose
2721467184	Arabinose
2721469997	Ribose
2721469856	Galactofuranose
2721469264	Unknown Z0417/16
2721477835	Autoinducer-2
2721466338	Unknown
2721470055	Malto-oligosaccharide
2721467668	Unknown
2721465402	G3P

Figure 3-17. Phylogenetic analysis of sugar-specific ABC transporters in EDL933. Phylogeny was inferred using the maximum-likelihood method and Le Gascuel model, with a Gamma distribution in MegaX. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The scale bar represents 0.5 substitutions per site. Bootstrap values are indicated on the respective branches. Clades A and B are coloured blue and orange respectively. The predicted substrate for each transporter are outlined in the table.

3.11. EHEC is unable to use D-ribulose as a sole carbon source

Based on the homology shared between Z0415 and ROD_24811, it was hypothesised that Z0415-19 would also allow for growth on D-ribulose in EHEC. However, when grown in M9 minimal media supplemented with D-ribulose EHEC was unable to grow (**Figure 3-18**), suggesting Z0415-19 not to transport the sugar. When transformed into EHEC, pSU-*rbl* enabled growth in M9 minimal media supplemented with 0.5 mg/mL D-ribulose, additionally providing support for *rbl* as a transport system specific for D-ribulose in *C. rodentium*. Therefore, despite displaying similarity, the identified systems in *C. rodentium* and EHEC were found to possess different substrate specificities that probably reflect the differential nutrients specific to their natural hosts.

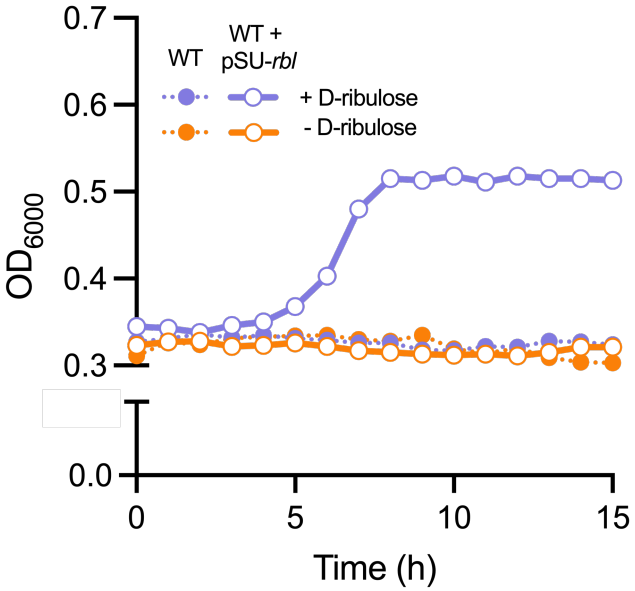


Figure 3-18. Growth profile of EHEC on D-ribulose as a sole carbon source. WT EHEC with and without pSU-*rbl* was grown in M9 minimal media supplemented with or without 0.5 mg/mL D-ribulose for 15 h.

3.12. Discussion

This study is the first detailed report of an ABC transporter specific for D-ribulose in *C. rodentium*, and Enterobacteriaceae for that matter. Using bioinformatic, genetic and physiological approaches, the annotation and predicted function of ROD_24811 and associated genes (ROD_24821-61) was validated. Specifically, the essentiality of ROD_24811-41 for growth on D-ribulose was demonstrated, with *C. rodentium* no longer being able to grow on the sugar as a sole carbon source when deleted for this locus. Moreover, at the transcriptional level, the locus was found to be expressed in direct response to D-ribulose. A search for homologues of ROD_24811 in EHEC str. EDL933, led to the identification of a novel PBP (Z0415). In trend with the *C. rodentium* system, Z0415 appears to be part of an ABC transporter specific-locus (Z0415-19), with specificity for a simple monosaccharide substrate. Carriage analysis of this locus revealed a strong association with pathogenic *E. coli*, particularly those of phylogroup E that consist predominantly of EHEC strains. However, growth analyses demonstrated that D-ribulose does not support growth of EHEC and can only be achieved upon trans-complementation of EHEC with the *C. rodentium rbl* system. This suggests EHEC and *C. rodentium* not to have identical nutrient preferences and raises the possibility that Z0415 has specificity for another nutrient source.

It is well recognised that enteric pathogens employ sophisticated strategies to overcome CR and colonise the highly competitive environment of the human gut (Khan et al., 2021). This is largely achieved through the avoidance of competition with the native gut microbiota, whereby pathogens such as *C. rodentium* and EHEC instead prioritise the utilisation of uncontested nutrients (Khan et al., 2021). Subsequently, studies have shown a divergence in nutrient preferences between commensal and pathogenic isolates of *E. coli* (Chang et al., 2004; Fabich et al., 2008). In the current study, it is proposed that *C. rodentium* possess a transport system specific for D-ribulose, within which the PBP (ROD_24811) has previously been shown to be upregulated during infection (Connolly et al., 2018).

D-ribulose is a metabolically associated sugar with roles as an intermediate metabolite in the PPP and glucuronate interconversions of both eukaryotes (e.g., fungi and plants) and prokaryotes. In the gut, the concentration of free D-ribulose is currently unknown. Though, in an unrelated study of enteric pathogens, ribulose was found to be detectable in the faeces of C57BL/6 mice by mass spectrometry (Zhang et al., 2021). The caveat of these findings was that the exact enantiomer was unknown (i.e., L- or D-ribulose). However, in a second study investigating the faecal metabolome of Winnie mice, D-ribulose specifically was detected (Robinson et al., 2016). These published data, taken together with the *in vitro* reporter data presented in this work, provide reasonable evidence that D-ribulose is present in the mouse gut, especially when considering the significant upregulation of ROD_24811 *in vivo* (Connolly et al., 2018). Interestingly, minimal activity was observed from the ROD_24811 promoter when grown in the presence of D-ribose, despite having no apparent growth defect in the Δ*rbl* mutant. It is likely that as both sugars share overlapping metabolic pathways and converge in the PPP, small amounts of imported ribose are converted to D-ribulose, allowing for activation of the ROD_24811 promoter. This would explain both the lag and reduction in promoter activity seen for D-ribose. The mechanism by which D-ribulose drives transcriptional changes in the expression of ROD_24811-61 remains to be determined. It is likely, as it is with most ABC transporters, including that of D-ribose, that the sugar is bound by a regulatory protein responsible for controlling expression of the locus (Kaplan et al., 2008). Due to no regulator being encoded at the *rbl* locus, it is suspected that the protein of question is encoded elsewhere in the genome. Prospective RNA-sequencing experiments looking to focus on transcriptomic changes

in *C. rodentium* in response to D-ribulose would likely help in identifying such regulatory factors. It may be that D-ribulose does not have its own cognate regulator and is instead governed by a regulator that is cognate to another sugar transport system. This phenomenon has been exemplified by the repression of the D-xylose transport system by the L-arabinose-specific regulator, AraC (Ammar et al., 2018; Groff et al., 2012). Subsequently, the $P_{ROD24811}$ activity observed in the presence of D-ribose could be instead conferred by a shared regulatory protein with a lower affinity for D-ribose, rather than its conversion to D-ribulose. Moreover, ROD_24811 was seen to be significantly upregulated *in vivo* in tandem with *rbsD*; a component of the canonical D-ribose utilisation system, providing strong evidence that high relative concentrations of D-ribose may also be present in the mouse gut (Shimada et al., 2013).

Whilst the identification of a locus specific for D-ribulose transport was surprising, it represents a unique opportunity for the pathogen to gain a foothold on the highly competitive gut environment. Exploitation of rare substrates of the mouse gut would prove extremely advantageous and allow *C. rodentium* to establish a novel nutrient niche, necessary for its colonisation. Moreover, this work identifies the carriage of two genes predicted to encode a D-ribulokinase and isomerase, which would allow for the intracellular processing of D-ribulose and entry into central metabolism (**Figure 3-19**).

Though carriage analysis showed the locus, including ROD_24851, to be present across several other genera, D-ribulokinases are only conserved in few bacterial lineages (Singh et al., 2017). Unlike the better-known L-ribulokinase, AraB, bacterial D-ribulokinases are highly specific for D-ribulose (Singh et al., 2017). What defines the high specificity of D-ribulokinases is perplexing as the residues known to be important for binding of L-ribulose in AraB are almost entirely conserved in the D-ribulokinase investigated in this study. In contrast, AraB has been described as ‘promiscuous’ and to also phosphorylate D-ribulose, albeit to a lower catalytic efficiency than for L-ribulose (Singh et al., 2017). In *E. coli*, AraB has been shown to display selectivity based on the configuration of -OH groups of pentulose sugars (Lee et al., 2001). Furthermore, the same study demonstrated AraB to have greater affinity (low K_m) for the erythro than the threo counterpart (Lee et al., 2001). Therefore, irrespective of the conservation in the residues required for ligand binding, a greater selectivity exists based on the finer characteristics of the given sugar. Subsequently, it could be that D-ribulokinase

possess greater selectivity making it not possible to accommodate both forms of ribulose. These differences may have been the outcome of changes over evolutionary time with D-ribulokinases being shown to have evolved from AraB (Zhang et al., 2011).

Present in other enterobacteria such as *K. pneumoniae*, D-ribulokinase are typically encoded within the ribitol utilisation operon, of which D-ribulose acts as an inducer of the operon (Heuel et al., 1998). In contrast to the *C. rodentium* system, a NAD dehydrogenase is associated with the ribitol operon of *K. pneumoniae*, responsible for the initial conversion of ribitol to D-ribulose (Heuel et al., 1998). Thus, it is unlikely that the Rbl system of *C. rodentium* is specific for ribitol, particularly as its transport is reliant on an ion symporter encoded by a single gene.

Deletion of ROD_24851 in *C. rodentium* led to a complete inability to grow on the sugar over 15 h suggesting D-ribulose cannot be used for growth without prior phosphorylation. Furthermore, it does not appear that AraB, specific for the L-form of ribulose, is capable of acting as surrogate under these isolated conditions. That being said, phosphorylation of D-ribulose by AraB may occur when grown in the presence of L-arabinose, which is required to activate its expression (Johnson & Schleif, 1995). Unlike the deletion of ROD_24851, the ΔROD_24861 strain was able to grow relatively similar to the WT. Processing by ROD_24861, or at least is expression, therefore appears not to be an absolute requirement for D-ribulose utilisation so long as ROD_24811-51 are present. This is likely due to the conversion of D-ribulose-5-phosphate to D-xylulose-5-phosphate and D-ribose-5-phosphate, thereafter, entering alternative metabolic pathways (Sprenger, 1995). Moreover, paralogous genes (*gutQ*, *kdsD*) encoding isomerases specific for D-ribulose exist in *E. coli* and can substitute for each other (Sommaruga et al., 2009). In *C. rodentium*, *kdsD* and *srlQ* (a *gutQ* orthologue) were able to be identified. Isomerisation of D-ribulose-5-phosphate by a KspF-like isomerase such as that predicted to be encoded by ROD_24861 perhaps suggests the metabolic fate of D-ribulose to terminate with D-arabinose-5-phosphate in *C. rodentium* (**Figure 3-19***)*. This metabolite is known to act as a precursor of 2-keto-3-deoxy-octanate biosynthesis in *E. coli*, which is a major constituent of surface glycolipids such as lipopolysaccharide (LPS) (Meredith & Woodard, 2006; Sommaruga et al., 2009). Interestingly, recent work has demonstrated LPS to have roles in masking

C. rodentium from host antibody recognition as part of the adaptive immune response, as well as being necessary for virulence *in vivo* (Chen et al., 2023).

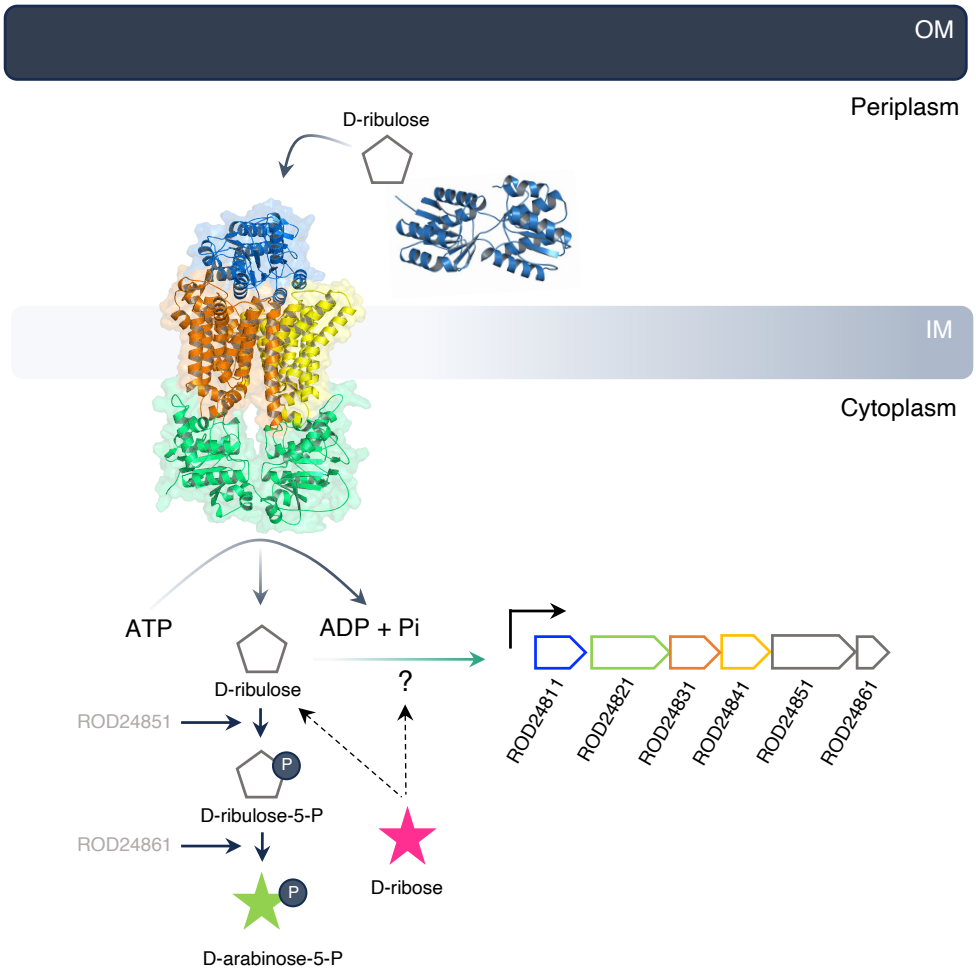


Figure 3-19. Proposed model of D-ribulose utilisation via the Rbl system in *C. rodentium* ICC168. D-ribulose is bound by the PBP (ROD_24811) in the periplasm and delivered to the TMDs (ROD_24831 and 24841) in the IM. D-ribulose is then transported across the IM via the hydrolysis of ATP in the NBDs of the cytoplasmic ATPase (ROD_24821). In the cytoplasm, free D-ribulose is able to activate the expression of the *rbl* locus via an unknown mechanism. D-ribulose is then phosphorylated by a D-ribulokinase (ROD_24851) prior to its isomerisation by ROD_24861. Additional isomerases, KdsD and SrlQ, may also convert the sugar to D-arabinose-5-phosphate. In addition, D-ribose activates the expression of the *rbl* locus either through its conversion to D-ribulose or binding to a transcriptional regulator common to D-ribulose.

Typically, studies of EHEC pathogenesis begin with *in vitro* experiments to first gain a mechanistic understanding of the processes that are occurring at the molecular level. Homologous systems are then sought in *C. rodentium* to test hypotheses *in vivo* (Wiles et al., 2006). In this work a reverse workflow was applied and instead a homologous system to that of Rbl in *C. rodentium* was identified in EHEC. This led to the identification of Z0415 in EHEC, which was also predicted to encode a PBP. Further analysis revealed Z0415 to form an operon with other genes predicted to encode the remaining components of an ABC transporter. This operon lay on an OI specific to EDL933, which are typically associated with virulence traits. For example, the LEE is carried on OI-148, whilst numerous NLE effectors are carried on several different islands, including OI-36, 50, 51 and 71 (Jiang et al., 2021).

Though similar to the Rbl system, key differences were identified in the EHEC system such as the absence of a kinase and isomerase. A BLASTP search of the ROD_24851 amino acid sequence against EDL933 returned no hits with significant percent identity. As ROD_24851 was demonstrated to be important for growth on D-ribulose, it was therefore unsurprising that EHEC was unable to grow on the sugar as a sole carbon source. As such, it is apparent that these systems have different substrate specificities despite sequence similarity in their PBPs and overall predicted structure for the transporter. Notably, whilst the PBPs displayed similarity to class B PBPs, the topology of the TMDs suggested transport of conflicting substrates. That being, class B PBPs encompass those specific for simple monosaccharides such L-arabinose (AraF), D-ribose (RbsB), D-xylose (XylF) and D-galactose (MglB) (Maqbool et al., 2015), whilst importers with TMDs comprising 10 helices belong to the Type II transporter group (Srikant, 2020; Ter Beek et al., 2014). In total there are said to be three types of importers (Type I-III) in prokaryotes, with Type I substrates being those required in bulk (i.e., sugars, amino acids) and Type II being those required in smaller quantities (i.e., vitamins, ions) (Ter Beek et al., 2014). Representative examples for each of the two types include $MalFGK_2$ (Maltose transporter) (Oldham et al., 2007) and BtuCD (Vitamin B_{12}) (Locher et al., 2002), respectively. Hence, it would be expected that a transporter with class B PBP would likely associate with a Type I importer. Furthermore, a phylogenetic analysis of NBDs with pfam00005 specific for sugars supported the hypothesis that Z0415-19 was likely a Type I importer and clustered with those characterised to transport simple monosaccharides. However, though completely

speculative, Z0415 could interact and deliver substrate to a Type I importer encoded elsewhere in the genome.

Comparatively, there appeared to be less conservation in the carriage of Z0415-19 outside of *Escherichia* than that seen for *rbl* outside of *Citrobacter*. These findings are suggestive that the system is likely more specific to *Escherichia* spp. Interestingly, whilst the *Shigella* spp., included as part of the large phylogenomic analyses were found to be negative for Z0415-19 carriage, a BLASTP search for the entire locus revealed a number of representatives positive for its carriage. Identification of Z0415-19 positive *Shigella* spp. was however unsurprising due to their close genetic relatedness to *E. coli*. In *Klebsiella* on the other hand, the locus was predicted to be carried on the KPC plasmid found amongst carbapenem resistant *K. pneumoniae* strains (Yang et al., 2021). KPC plasmids have been found to be able to conjugate between *K. pneumoniae* and *E. coli in vitro* (Yang et al., 2021). This therefore highlights a potential route of transfer of the locus between species, likely by HGT.

Finally, an in-depth phylogenomic analysis of *E. coli* was conducted to determine associations between Z0415-19 carriage and phylogroup or pathotype. Most strikingly, carriage of the locus was almost entirely absent from the strains of phylogroup A which largely comprises commensal *E. coli* (Clermont et al., 2000). Since Z0415 is carried on an OI which are generally absent in commensal strains, this finding was logical. That being said, instances could be found whereby commensal strains were found to be positive for carriage of the locus (e.g., in phylogroup B1). Additionally, although OI-17 designation is extremely specific to EHEC O157:H7 strain EDL933, carriage was observed for almost entirely all of phylogroup E. Locus carriage was also heavily observed across phylogroups D and B1. However, due to the smaller number of strains that comprise phylogroup D, the number of strains positive for Z0415-19 carriage was in fact similar to that of phylogroup B2. Therefore, it is likely that the association between locus carriage and strains of phylogroup D is perhaps not as strong as that seen for E and B1. Notably, both phylogroups E and B1 encompassed mainly EHEC strains, supporting what is known in the literature about phylogroup E largely comprising the O157:H7 lineage (Clermont et al., 2021). A strong association therefore exists between Z0415-19 carriage and a strain being EHEC, particularly of the

O157:H7 lineage. Nonetheless, exceptions could again be observed, and a number of non-EHEC strains, mainly EAEC and ETEC, were identified in phylogroup B1.

Although not as extreme as phylogroup A, reduced carriage of the locus was observed across phylogroup B2, known to be enriched with ExPEC strains (up to 50%), which typically cause infection outside the gut (Geurtsen et al., 2022; Lagerstrom & Hadly, 2023). This represents an interesting phenomenon whereby having these systems may be disadvantageous or advantageous dependent on the environment. Subsequently, it could be that the system plays a role in niche restriction and without it certain strains are unable to colonise or be virulent within a particular niche. This is best exemplified by the *dsdCXA* locus found in extraintestinal UPEC strains which allows the pathogen to detoxify and survive the inhibitory concentrations of D-serine in the urinary tract (Connolly et al., 2014). In EHEC, the same locus is truncated but D-serine is found at sub-inhibitory concentrations in the intestinal environment, to which EHEC is restricted (Connolly et al., 2016). The difficulty with drawing clear conclusions on locus carriage and pathotype is that a large number of strains were unable to be assigned a defined pathotype, so likely skew the associations. Moreover, carriage amongst various pathotypes across the phylogroups further reduced the clarity of these associations but is likely explainable by the high rates of recombination seen for *E. coli* (Denamur et al., 2020). It should also be noted that carriage does not necessarily suggest functionality and experimental evidence is needed to validate the systems role. Nevertheless, phylogroup E evidently represents a hotspot for Z0415-19 carriage due to its almost exclusive presence, and chapter 4 will discuss the roles identified for this locus in EHEC fitness.

3.13. Conclusions

Taken together, the Rbl system (ROD_24811-61) represents a route for D-ribulose uptake in the mouse gut, potentially allowing for the establishment of a novel niche. Furthermore, the transport of ribulose and downstream metabolic processing may have roles in LPS biosynthesis, contributing to virulence in the host. The identified system in EHEC (Z0415-19) is distinct from the Rbl system and lacks associated enzymes that would enable D-ribulose utilisation. However, Z0415-19 potentially has a role in EHEC pathogenesis based on its location on an OI and strong association with InPEC strains.

4. Investigating the regulation and physiological role of a novel ABC transporter in EHEC str. EDL933

4.1. Introduction

Bacterial sugar transport systems need only be expressed when their cognate sugar is present in the surrounding environment. The expression of these dedicated transport systems must therefore be temporally and spatially controlled to prevent wasted energy expenditure (Choudhury & Saini, 2019). This process is overseen by regulatory proteins that are capable of activating and/or repressing transcription of the genes encoding the corresponding transporter and metabolic enzymes (Choudhury & Saini, 2019; Kaplan et al., 2008). Control is typically achieved by the regulator binding to the sugar, driving a change in how the protein interacts with the regulatory sites within the promoter region of target genes (Kaplan et al., 2008). Mainly this occurs via two mechanisms, with sugar-regulator binding (a) encouraging the regulator to bind to target sites that activate gene expression (e.g., an activator) (Johnson & Schleif, 1995) or (b) alleviating the pre-bound regulator such that it can no longer suppress gene expression (e.g., a repressor) (Beckwith, 1967). Additionally, some of these regulators exert control over their own expression in an attempt to maintain homeostasis (Kaplan et al., 2008).

A well-studied example is the system specific for the aldopentose sugar L-arabinose. L-arabinose is imported into the cytoplasm from the periplasm via the high affinity ABC transporter, AraFGH, and the low affinity symporter, AraE (**Figure 4-1A**) (Johnson & Schleif, 1995). The expression of these systems is controlled by the L-arabinose-specific regulator, AraC (Johnson & Schleif, 1995). Co-expressed with these transporters are the enzymes responsible for downstream processing of L-arabinose (Miyada et al., 1984). These genes comprise a catabolic operon (*araBAD*) that is similarly governed by AraC activity, encoding an L-ribulokinase (araB), L-arabinose isomerase (araA) and L-ribulose-5-P 4-epimerase (**Figure 4-1B)** (Lee et al., 1986; Miyada et al., 1984). The terminal product of this pathway, D-xylulose-5-phosphate then enters the PPP.

AraC belongs to the AraC/XylS family of regulators and functions as a homodimer (Schleif, 2000). The monomeric form of AraC comprises a DNA binding domain and

dimerisation domain containing the arabinose binding pocket, which are joined by a flexible linker (**Figure 4-1C**) (Schleif, 2000). In the absence of L-arabinose, AraC is bound to the DNA at two regulatory half-sites denoted as O_2 (Operator site 2) and I_1 (Inducer site 1), located upstream of the promoter for L-arabinose utilisation genes (AUGs) (**Figure 4-1D**) (Lobell & Schleif, 1990). Binding at these locations results in the looping of DNA and therefore blocking the access of RNAP meaning that transcription is prevented (Schleif, 2000). In the presence of L-arabinose AraC binds the sugar and undergoes an allosteric change such that one of the DNA binding domains instead binds a secondary I site (I_2), opening the DNA loop (**Figure 4-1**) (Lobell & Schleif, 1990). When bound at these sites, RNAP access is no longer blocked and so transcription can occur (Lobell & Schleif, 1990). Whether the same regulatory mode occurs at every promoter regulated by AraC is not necessarily known.

An additional layer of complexity is introduced to the regulation of AUGs through the involvement of the master transcriptional regulator, cyclic AMP (cAMP) receptor protein (CRP), in a phenomenon known as carbon catabolite repression (Kaplan et al., 2008). Catabolite repression occurs by preventing the expression of genes required to use secondary carbon sources when glucose is present (Görke & Stülke, 2008). Following the complete use of glucose, the adenylate cyclase required for ATP to be converted into cAMP is no longer repressed, and intracellular concentrations of the signalling molecule increase (Shimada et al., 2011). cAMP is then bound by CRP and in turn activates the regulator such that it is able to then activate the expression of alternative carbon utilisation promoters (Aidelberg et al., 2014; Shimada et al., 2011). Typically, this shift from growth on the preferred carbon source of glucose to a secondary carbon source results in growth that is diauxic, best exemplified by the glucose-lactose shift in *E. coli* (Aidelberg et al., 2014). Subsequently, a hierarchy of expression exists amongst sugar utilisation systems, and the order of non-glucose sugar system expression in *E. coli* is as follows: lactose > arabinose > xylose > sorbitol > rhamnose > ribose (Aidelberg et al., 2014). Interacting with the alpha subunit of RNAP (Dhiman & Schleif, 2000), the binding of CRP-cAMP to the promoter region is required for the full activation of AUGs (Johnson & Schleif, 1995).

The presence of novel and additional sugar transport systems highlights a mechanism by which EHEC could potentially scavenge additional nutrients from the surrounding

environment and gain a competitive advantage. However, it is important to understand how these systems are regulated and the signals that feed into their regulation. Moreover, underpinning the substrates that regulate these novel uncharacterised systems in EHEC will likely provide information on their transport specificity and physiological significance. This chapter therefore expands upon the work surrounding the EHEC-specific locus identified in chapter 3, addressing the mechanism by which the locus is transcriptionally regulated and its contribution to fitness when grown on L-arabinose. Additionally, competition for L-arabinose between EHEC and the prominent member of the gut microbiota, *Bt*, is explored.

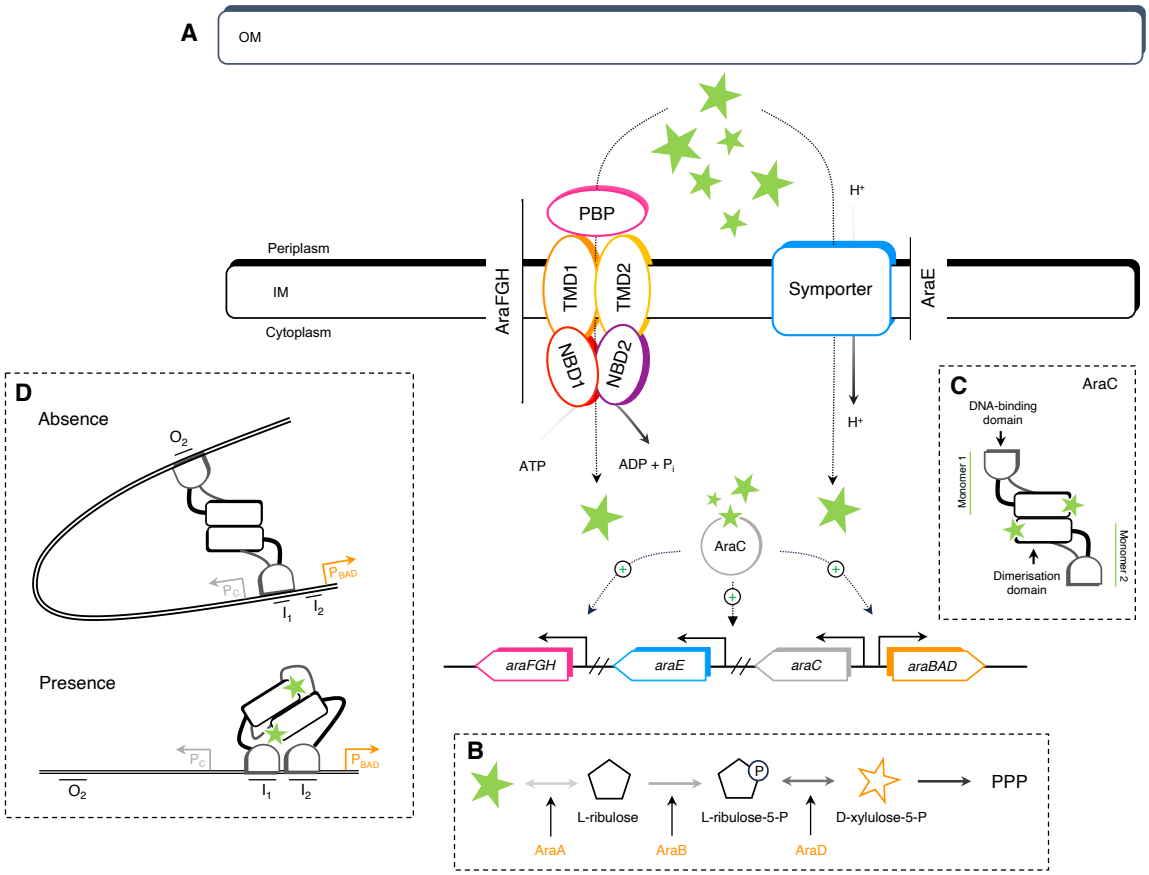


Figure 4-1. Overview of L-arabinose regulation, transport, and metabolism in *E. coli*. (**A**) L-arabinose is transported into the cytoplasm via the AraFGH ABC transporter and AraE symporter. In the cytoplasm L-arabinose is bound by AraC activating expression of the L-arabinose utilisation system; (**B**) Expression of the system allows for the downstream metabolism of L-arabinose by the enzymes encoded by *araBAD*. L-arabinose is sequentially converted to D-xylulose-5-P for entry into central metabolism via the PPP; (**C**) AraC comprises a homodimer with each monomer consisting of a DNA-binding domain and dimerisation domain; (**D**) In the absence of L-arabinose the system is not expressed. AraC without L-arabinose binds to the regulatory regions O_2 and I_1 upstream of AUGs. Binding at these two sites causes the DNA to loop and block the access of RNAP, required for transcription. In the presence of L-arabinose, AraC binds the sugar and undergoes a conformational change such that the DNA loop opens. AraC instead binds both I_1 and I_2 sites, allowing RNAP access. L-arabinose is denoted as a green star.

4.2. Transcriptional responsiveness of Z0415-19 to sugar substrates

Clear similarities could be observed between the *C. rodentium* and EDL933 systems investigated. An InterPro search of the PBP (Z0415) and two TMDs (Z0418, Z0419) in EDL933 revealed identical predicted domains to those identified for the respective components of the Rbl system (**Table 4-1**). However, due to the absence of any associated catabolic enzymes and inability to grow on D-ribulose as a sole carbon source, it was hypothesised that the substrate specificity of Z0415-19 was likely different. Based on the initial annotation in KEGG, InterPro scan results and previous phylogenetic analysis of ABC transporters in EDL933, the system's specificity to D-ribose, D-xylose and L-arabinose was explored. As with ROD_24811, a 300 bp promoter region upstream of Z0415 was cloned into the pKM1*lux* vector for testing its transcriptional responsiveness to the chosen sugars.

Table 4-1. Domain analysis of Z0415-19 components in EDL933. Predicted domains and locations identified in Z0415, Z0418 and Z0419 proteins using the InterPro Scan tool.

Protein	Position (Aa)	Predicted domain
Z0415 (PBP)	12-325	D-xylose-binding periplasmic protein (PTHR30036)
	32-325	PBP1_LsrB_QS-like (cd20002)
Z0418 (TMD1)	5-320	YPHD-related domain (PTHR32196)
	44-312	TM_PBP1_transp_AraH-like (cd06579)
Z0419 (TMD2)	3-307	YPHD-related domain (PTHR32196)
	45-306	TM_PBP1_transp_AraH-like (cd06579)

When grown in the presence of L-arabinose, D-ribose and D-xylose, activity from pKM1*lux*-P_{Z0415} was measured in WT EDL933 (**Figure 4-2**). Promoter activity was only observed across all three media types (DMEM, LB and M9) when supplemented with L-arabinose. The addition of L-arabinose to M9 minimal media had the greatest effect on promoter activity, despite displaying poorer growth when compared to growth in LB and DMEM (**Figure 4-2**), likely due to there being less to feedback and repress P_{Z0415} expression. A much smaller signal of < 0.5 x 10^5 RLU at the peak of activity was observed when EHEC was grown in LB supplemented with L-arabinose (**Figure 4-2**), whilst moderate promoter activity (~2 x 10^5) was seen when grown in DMEM (**Figure 4-2**). The time at which peak promoter activity was observed also differed across media types when supplemented with L-arabinose. Though displaying the lowest activity of the three media types, promoter activity was found to peak after 2 h when grown in LB supplemented with L-arabinose, as opposed to 8 and 9 h for M9 minimal media and DMEM respectively (**Figure 4-2**). Experiments were also conducted in MEM-HEPES, a derivative of DMEM containing low glucose, which similarly demonstrated P_{Z0415} promoter activity to be highest when supplemented with L-arabinose (**Figure 4-3**). This activity was also found to increase with concentration of the sugar.

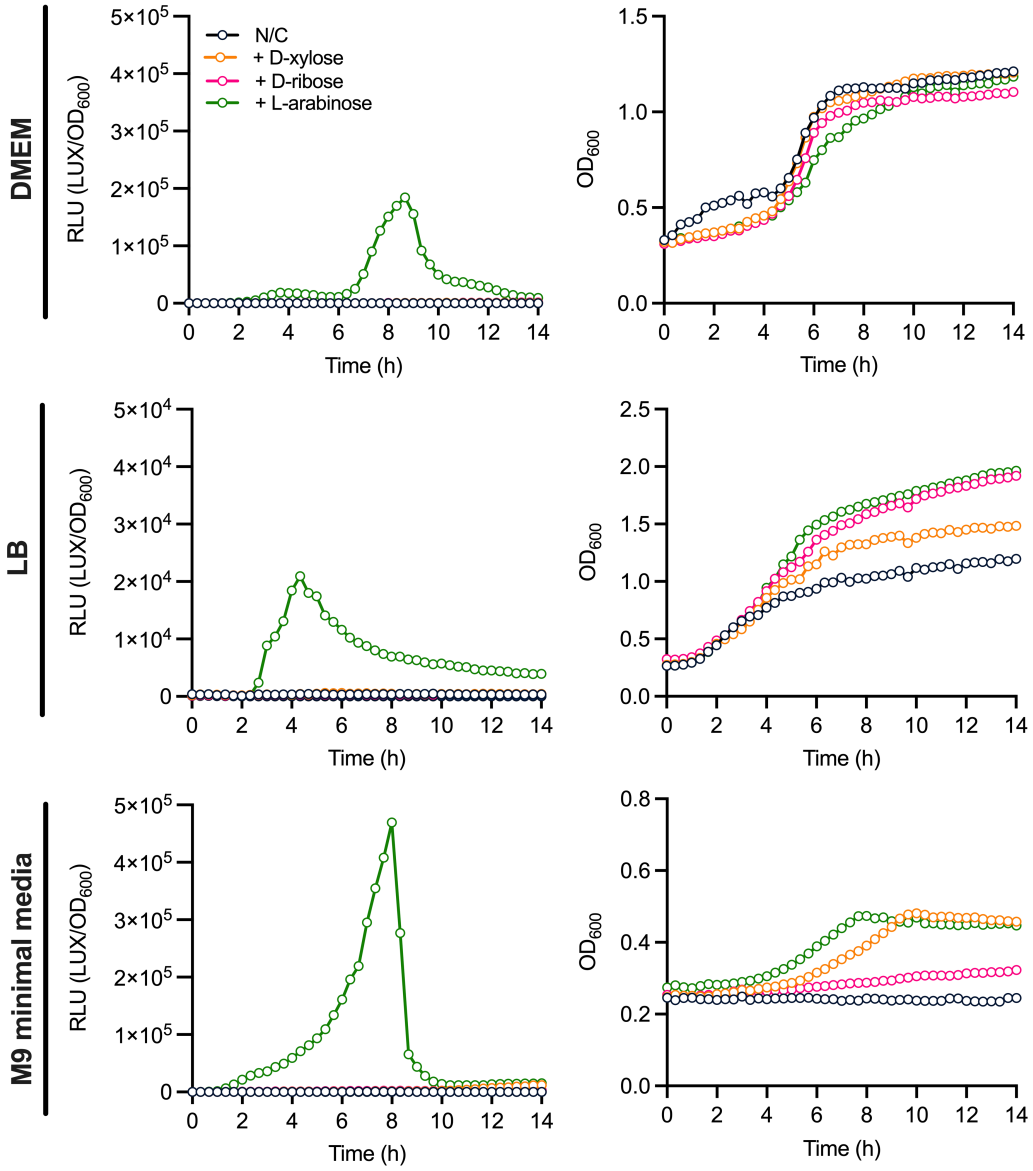


Figure 4-2. Effect of different aldopentose sugars on P_{Z0415}-LUX activity across different media types. WT TUV93-0 with pKM1*lux*-P_{Z0415} was grown in DMEM, LB and M9 minimal media supplemented with either 0.5 mg/ml D-xylose, D-ribose, or L-arabinose for 14 h. A no carbon (N/C) condition was used as a control. Panels on the left and right depict the luminescence outputs normalised against OD_{600} (RLU) and OD_{600} values across media, respectively.

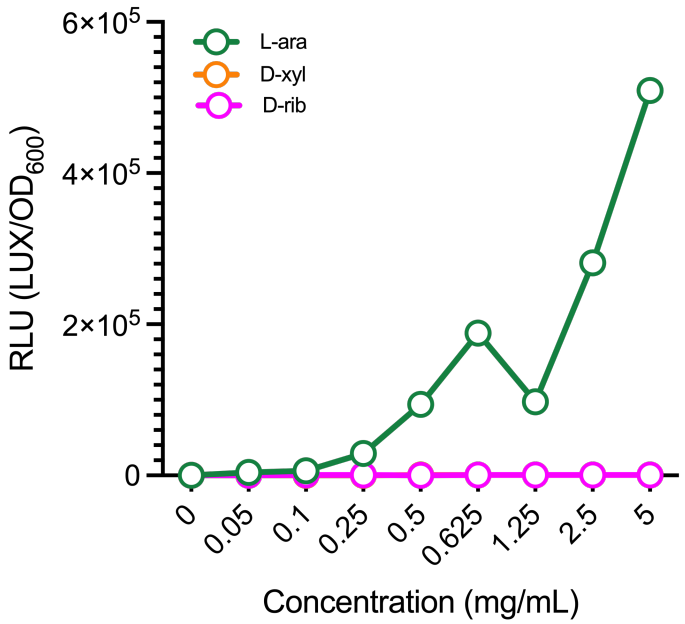


Figure 4-3. Concentration-dependent effect of aldopentose sugars on P_{Z0415} expression. WT TUV93-0 with pKM1*lux*-P_{Z0415} was grown in MEM-HEPES supplemented with either D-xylose, D-ribose, or L-arabinose, across a concentration range. Data is plotted as raw luminescence normalised against OD_{600} (RLU). Readings were taken at the mid-exponential phase of growth (~0.6; 4 h).

To validate the reporter analysis of L-arabinose on Z0415-19 transcription, RNA was isolated from WT TUV93-0 grown on MEM-HEPES supplemented with or without L-arabinose (5 mg/mL). RT-qPCR was then used to quantify the relative transcript abundance for Z0415 in TUV93-0. In the presence of L-arabinose, the relative expression of Z0415 was found to have significantly increased by > 60-fold when compared to in the absence of the sugar (**Figure 4-4**). Similarly, a significant increase of 50-fold and 30-fold was observed for Z0417 and Z0418 respectively (**Figure 4-4**). These data supported the conclusion that L-arabinose is the inducer of P_{Z0415} expression.

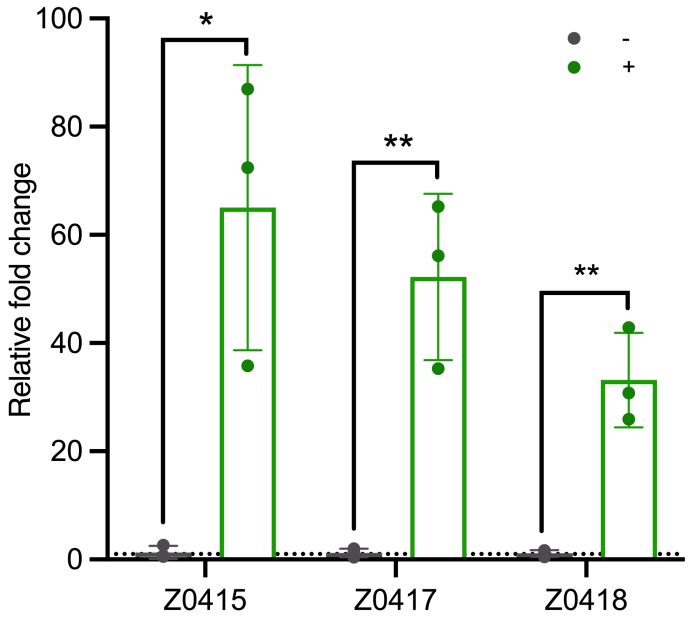


Figure 4-4. RT-qPCR analysis of relative expression across Z0415-19 in the presence of L-arabinose. The mean relative expression of Z0415, Z0417 and Z0418 was derived using RNA extracted from WT TUV93-0 grown in MEM-HEPES supplemented with 5 mg/mL L-arabinose, relative to the untreated control. All cultures were grown to the late-exponential phase of growth at 37 °C. The black dotted line indicates baseline expression of the control. $n = 3$ biological replicates and error bars indicate SD. (*) indicates $P < 0.05$; (**) $P < 0.01$ - calculated using a student's T-test.

4.3. Bioinformatic analysis of Z0415-19 regulation by L-arabinose

To gain a better understanding of how L-arabinose might transcriptionally drive Z0415-19 expression, attention turned to the transcriptional regulator, AraC. It was hypothesised that AraC binds regulatory sites upstream of Z0415-19 and activates its expression, as with the canonical L-arabinose utilisation system in *E. coli*. The I site to which AraC binds in *E. coli* is responsible for the activation of AUGs and has been extensively footprinted. Using an *in-silico* based approach, the I site of three AUGs from *E. coli* str. MG1655 (P_{araB}, P_{araE}, P_{araJ}) were aligned with the previously cloned 300 bp of sequence upstream of Z0415 (**Figure 4-5A**). A motif consisting of 17 residues was located 132 bp upstream of the Z0415 start codon and displayed similarity to the consensus sequence of the I site. Whilst sequences were largely variable, the residues at positions 4, 9, 12, 15, 16 and 17 were completely conserved. A sequence logo for the aligned I sites aided the generation of a consensus sequence to which AraC may recognise and bind (**Figure 4-5B**). Comparisons of I sites between

AUGs and Z0415 revealed that no two sites shared a percent identity of > 65.4 % (**Figure 4-5CD**). The I site of Z0415 was found to be most similar with that of P_{araE} (59.3% identity score) (**Figure 4-5CD**). The I site of P_{araJ} was the most dissimilar to all other I sites.

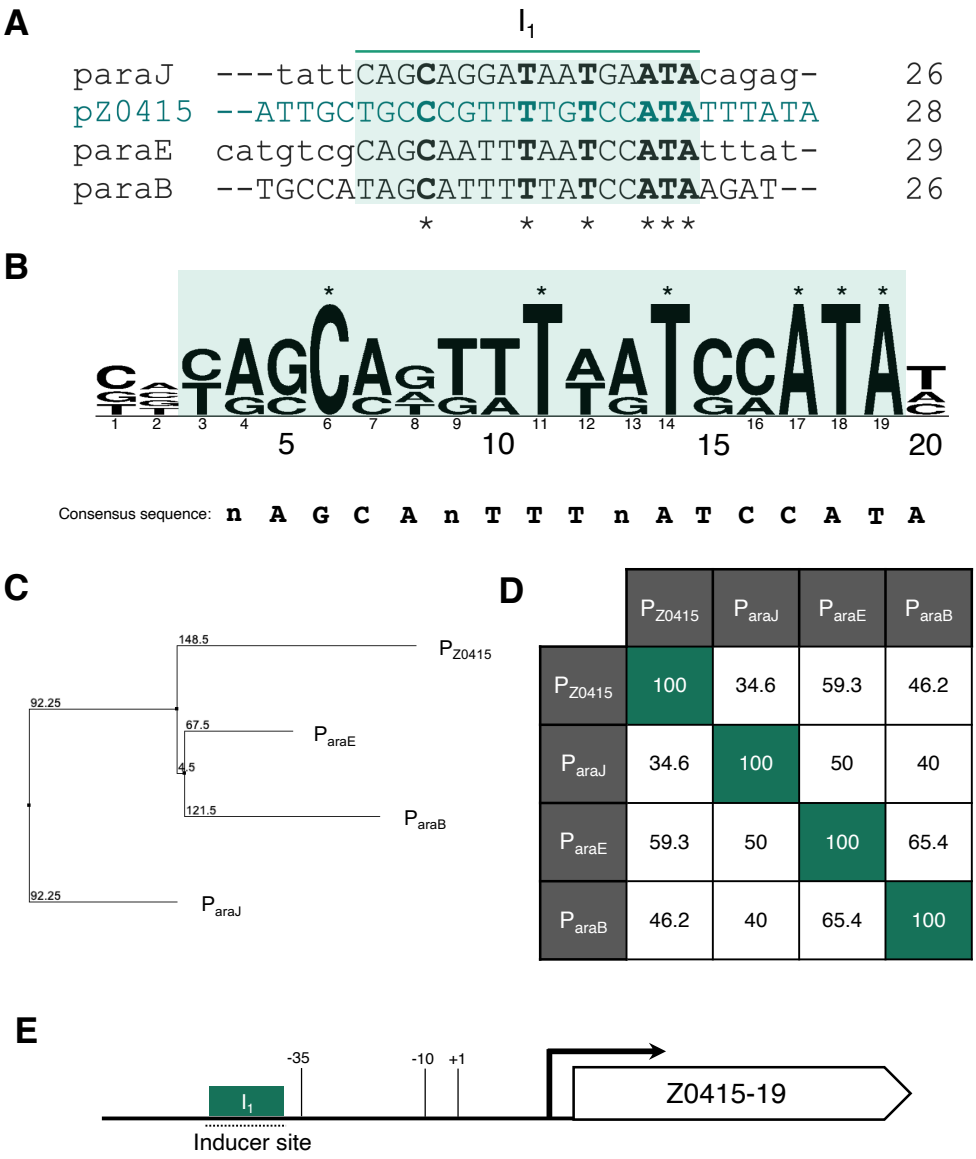


Figure 4-5. *In silico* searches for inducible AraC regulatory sites in P_{Z0415}. (**A**) Multiple sequence alignment of known inducer (I_1) sites in the promoter regions of three AUGs with sequence upstream of the Z0415 start codon. I_1 site sequences were taken from RegulonDB for *E. coli* MG1655. (*) Asterisks denote residues that are completely conserved across sequences. (**B**) Sequence logo of the I_1 site consensus sequence derived from the multiple sequence alignment of I_1 sites. (**C**) Neighbour joining tree of aligned I_1 site sequences generated in Jalview. (**D**) Percent identity matrix of I_1 site sequences generated in ClustalOmega. Green boxes represent an alignment where all residues in the sequence are completely conserved. (**E**) Schematic of predicted I_1 site location in relation to P_{Z0415}. Highlighted are the predicted -35 and -10 sites inferred by the BRPOM server.

For the induction of AUG expression, AraC is required to bind a second inducer site (I_2) adjacently located downstream of the I_1 site described. When bound to both I_1 and I_2 sites DNA looping is alleviated and allows RNAP access to the promoter of AUGs. Using similar approaches, a search for a second I site downstream of the I_1 site of Z0415 was carried out, but this time excluding P_{araJ} due to its divergent sequence. Predicted to be located 13 bp downstream of I_1, the secondary I site of Z0415 was again variable, however, was conserved with that of P_{araB} and P_{araE} at positions 1, 10 and 13-17 (**Figure 4-6**). Despite this conservation, sequence similarity shared by the I_2 site of Z0415 with either P_{araB} or P_{araE} was lower than the similarity shared by the I_2 sites of P_{araB} and P_{araE}.

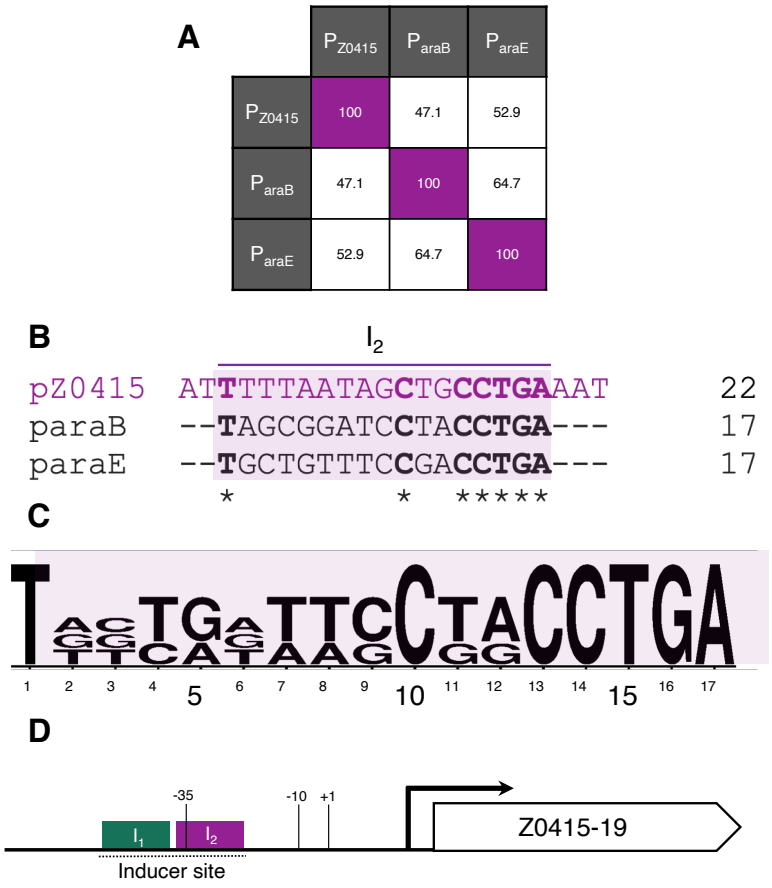


Figure 4-6. *In silico* searches for inducible AraC regulatory sites in P_{Z0415}. (**A**) Percent identity matrix of I_2 site sequences generated in ClustalOmega. Purple boxes represent an alignment where all residues in the sequence are completely conserved. (**B**) Multiple sequence alignment of known inducer (I_2) sites in the promoter regions of two AUGs with sequence upstream of the Z0415 start codon. I_2 site sequences were taken from RegulonDB for *E. coli* MG1655. (*) Asterisks denote residues that are completely conserved across sequences. (**C**) Sequence logo of the I_1 site consensus sequence derived from the multiple sequence alignment of I_1 sites. (**D**) Schematic of the predicted I_2 site location in relation to P_{Z0415}. Highlighted are the predicted -35 and -10 sites inferred by the BRPOM server.

In addition to the requirement for AraC binding both I sites (I_1 & I_2), full activation of AUG expression requires cAMP-CRP binding. Taking the known CRP binding sequence of the P_{araB} promoter, searches to identify a similar site within the promoter region of Z0415 were undertaken. Unlike the I_2 site, the CRP binding site was located upstream of I_1, hence the rationale to explore this region of sequence. Sharing 53 % identity, a stretch of sequence 8 bp upstream of I_1 was identified as a predicted CRP binding site within the promoter region of Z0415 (**Figure 4-7A**).

The predicted presence of both I (I_1 and I_2) and CRP binding sites support a mechanism of how L-arabinose could regulate Z0415-19 expression under inducing conditions. However, under non-inducing conditions, AraC is bound to the I_1 site and an additional secondary site (O_2) to allow for DNA looping and restricting the access of RNAP. As part of these analyses, a potential O site was unable to be identified implying a different mechanism of Z0415 repression when compared to canonical AUG expression. Additionally, the predicted -35 site was identified to overlap with the predicted I_2 site of Z0415 (**Figure 4-7B**), complicating the basis for RNAP binding in the presence of AraC. That being said, the ability to accommodate the -35 site overlap is dependent upon the specific sigma factor that is responsible for recognising the Z0415 promoter. The stress response sigma factor, RpoS, for example displays reduced dependency upon -35 site binding compared to the housekeeping sigma factor, RpoD (Battesti et al., 2011). Subsequently, if this regulation occurs under carbon-limited or early stationary phase conditions, of which RpoS is the designated sigma factor, it could be that the presence of AraC in the vicinity of the -35 site can be tolerated. Regardless, these findings provide strong evidence for the role of AraC in regulating Z0415-19 expression.

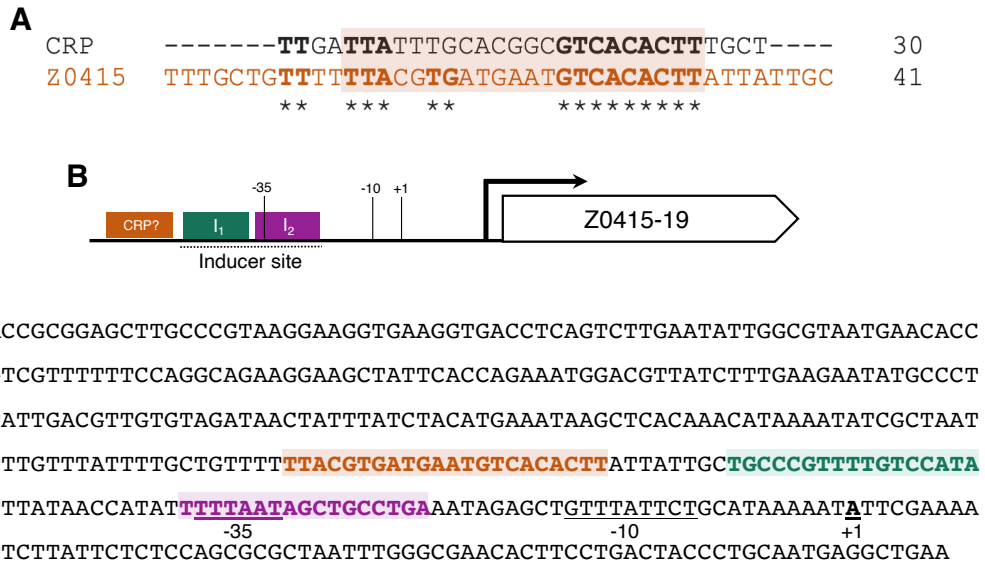


Figure 4-7. *In silico* searches for CRP binding sites in P_{Z0415}. (**A**) Multiple sequence alignment of a single known CRP site upstream of P_{araB} with sequence upstream of the Z0415 start codon. The known CRP site sequence was taken from RegulonDB for *E. coli* MG1655. (*) Asterisks denote residues that are completely conserved. (**B**) Schematic of the predicted CRP site in relation to P_{Z0415}. The P_{Z0415} sequence is also shown with the predicted I_1, I_2 and CRP sites highlighted.

4.4. Dissecting the role of AraC in the regulation of Z0415-19

Based on the hypothesis that L-arabinose drives Z0415-19 expression via AraC, its contribution to the observed phenotype was experimentally tested by deletion of *araC* in TUV93-0 (**Figure 4-8A**). As expected, the Δ*araC* strain was no longer able to grow on L-arabinose as the sole carbon source when compared with the WT (**Figure 4-8B**). Next, the pKM1*lux*-P_{Z0415} construct was transformed into the Δ*araC* mutant and grown under the same conditions. Deletion of the regulator led to a complete absence of expression, with levels being similar to the WT in the absence of L-arabinose (**Figure 4-8A**). As previous, the WT only displayed expression in the presence of L-arabinose (**Figure 4-8A**).

Trans-complementation of the mutant with pACYC184-*araC* under the control of its native promoter (P_{araC}) recovered P_{Z0415} expression when grown in the presence of L-arabinose, although being delayed (**Figure 4-9B**). When grown at higher concentrations of L-arabinose (5 mg/mL), the delay in P_{Z0415} expression was found to be reduced in the complemented strain (**Figure 4-9C**). Despite higher expression in the WT and complemented strain when supplemented with 5 mg/mL L-arabinose, the onset of P_{Z0415} expression was unaffected. At its highest (8 h), P_{Z0415} expression in the complemented strain was 3-fold lower than that for the WT when grown with 5 mg/mL L-arabinose (**Figure 4-9C**). Nevertheless, P_{Z0415} expression was found to be consistently higher in the complemented strain when supplemented with L-arabinose, reflecting the WT phenotype (**Figure 4-9BC**). To validate these findings, the induction and regulation of P_{Z0415} was tested in a second EHEC strain, Sakai. Again, P_{Z0415} expression was recovered to a level almost identical to the WT (**Figure 4-9D**), further confirming the regulatory role of AraC. However, P_{Z0415} expression in WT Sakai was found to be lower than that for TUV93-0 to begin with.

To next decipher whether AraC needs to be bound with L-arabinose to induce P_{Z0415} expression, the *araC* gene was cloned into pSUPROM such that it would be constitutively expressed. When grown in the absence of L-arabinose, no P_{Z0415} expression was observed for TUV93-0 with pSU-*araC* (**Figure 4-9E**). However, as previously seen for the WT and pACYC184-*araC* complemented Δ*araC* mutant, supplementation with 0.5 mg/mL L-arabinose led to P_{Z0415} expression, suggesting AraC must be bound with L-arabinose for induction.

Finally, to address whether the expression of P_{Z0415} was specific to L-arabinose, WT TUV93-0 with pKM1*lux*-P_{Z0415} was grown in MEM-HEPES supplemented with the D-isomer of arabinose. *E. coli* is typically unable to use D-arabinose for growth and, accordingly, was found to have no effect on P_{Z0415} (**Figure 4-9F**). Taken together, these data provide strong evidence that OI-17 encoded Z0415-19 is regulated by and responsive to L-arabinose exclusively.

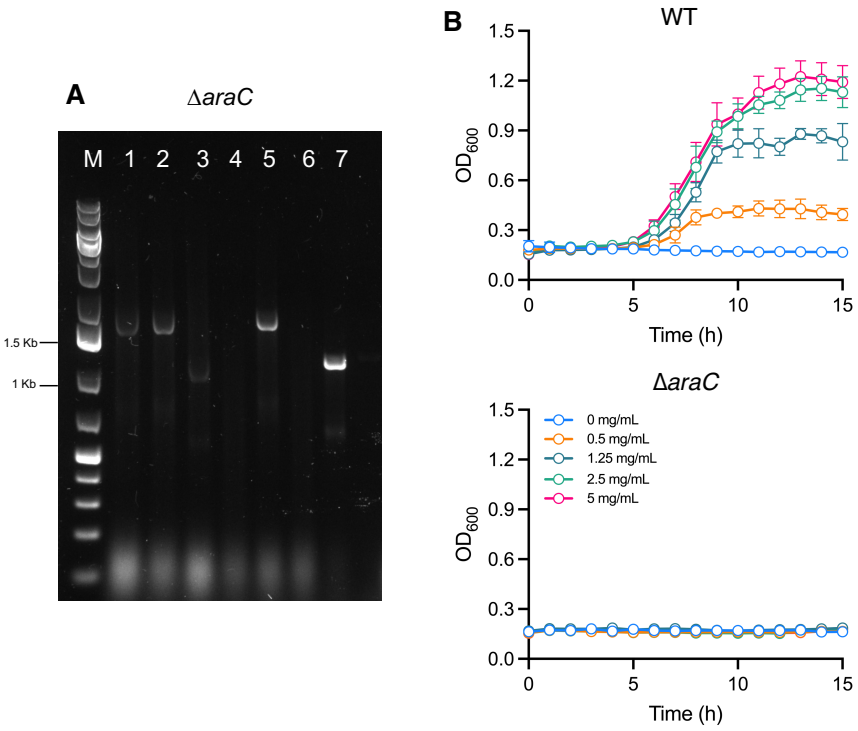


Figure 4-8. An Δ*araC* mutant in TUV93-0 is unable to grow on L-arabinose as a sole carbon source. (**A**) Agarose gel confirmation Δ*araC* in TUV93-0. Lanes 1-6 correspond to colonies tested for replacement of *araC* with the Kan^R cassette (1.6 Kb) if successfully deleted. Lane 7 corresponds to the WT gDNA control. Lane M corresponds to the DNA ladder. (**B**) Growth of WT and Δ*araC* TUV93-0 in M9 minimal media supplemented with L-arabinose over 15-h. $n = 3$ biological replicates and error bars indicate SD.

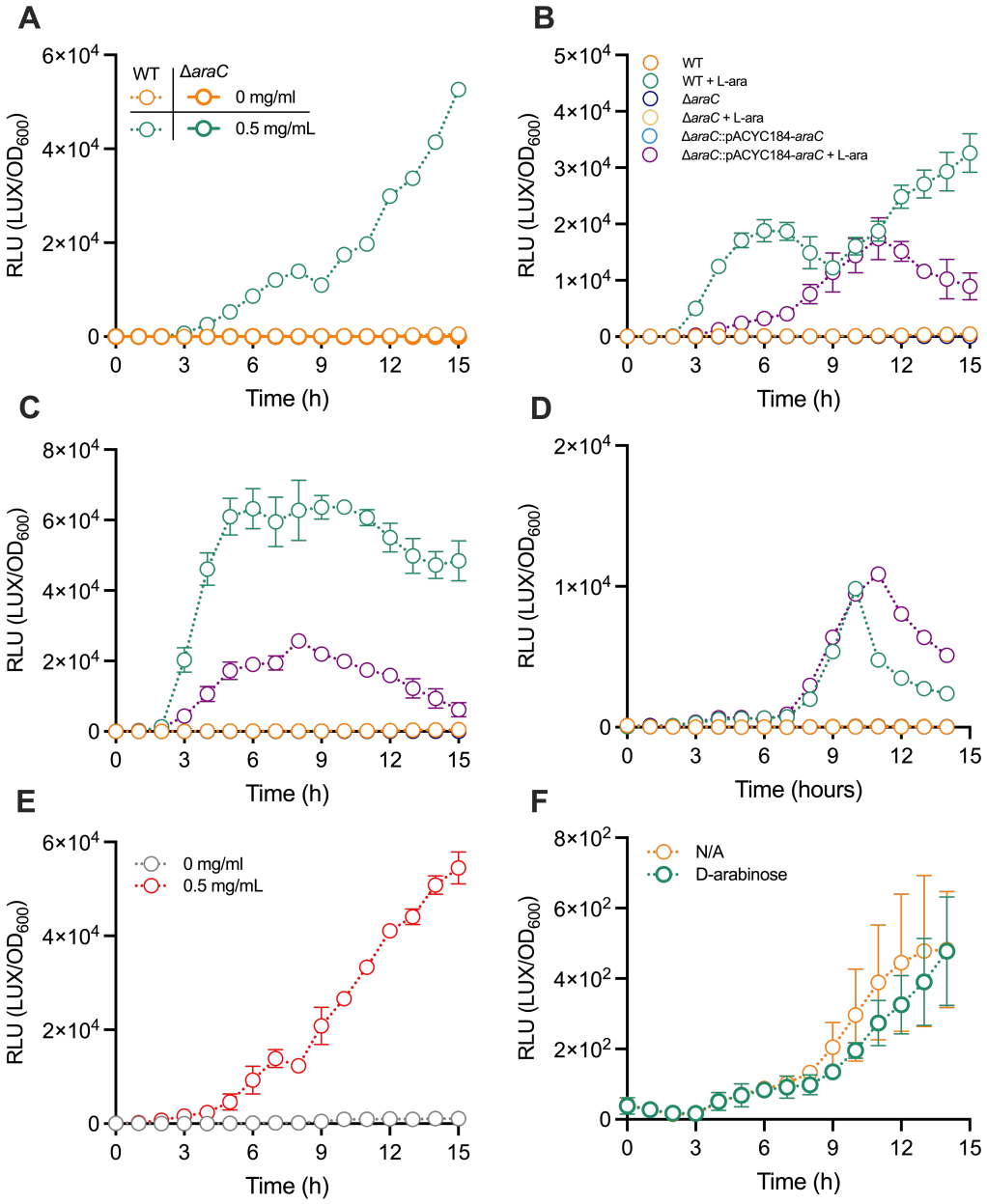


Figure 4-9. AraC bound with L-arabinose is essential for P_{Z0415} expression. RLU profiles of (**A**) WT and Δ*araC* TUV93-0 grown with +/- L-arabinose; (**B**) WT TUV93-0, Δ*araC* TUV93-0 and pACYC184-*araC* complemented Δ*araC* strains grown with +/- 0.5 mg/mL L-arabinose; (**C**) WT TUV93-0, Δ*araC* TUV93-0 and pACYC184-*araC* complemented Δ*araC* strains grown with +/- 0.5 mg/mL L-arabinose; (**D**) WT Sakai, Δ*araC* Sakai and pACYC184-*araC* complemented Δ*araC* strains grown with +/- 0.5 mg/mL L-arabinose; (**E**) pSU-*araC* complemented Δ*araC* TUV93-0 grown with +/- 0.5 mg/mL L-arabinose; (**F**) WT TUV93-0 grown with +/- 0.5 mg/mL D-arabinose. All assays were conducted in MEM-HEPEs over a period of 15 h. Strains used all carried the pKM1*lux*-P_{Z0415} reporter. When shown, error bars are indicative of SD. *n* = 3 biological replicates.

4.5. Determining the contribution of Z0415-19 to growth on L-arabinose

Following the determination that AraC is essential for Z0415-19 expression, it was tested whether the locus could contribute to the fitness of EHEC when grown on L-arabinose as a sole carbon source. To address its role, Z0415-19 was deleted in TUV93-0 and grown on M9 minimal media supplemented with L-arabinose across a range of concentrations: 0, 0.25, 0.5, 1, 1.25, 2.5, 5 and 10 mg/mL (**Figure 4-10**). Upon comparison with the WT, no significant differences could be observed in the growth profile of Z0415-19 at any concentration (**Figure 4-10B**).

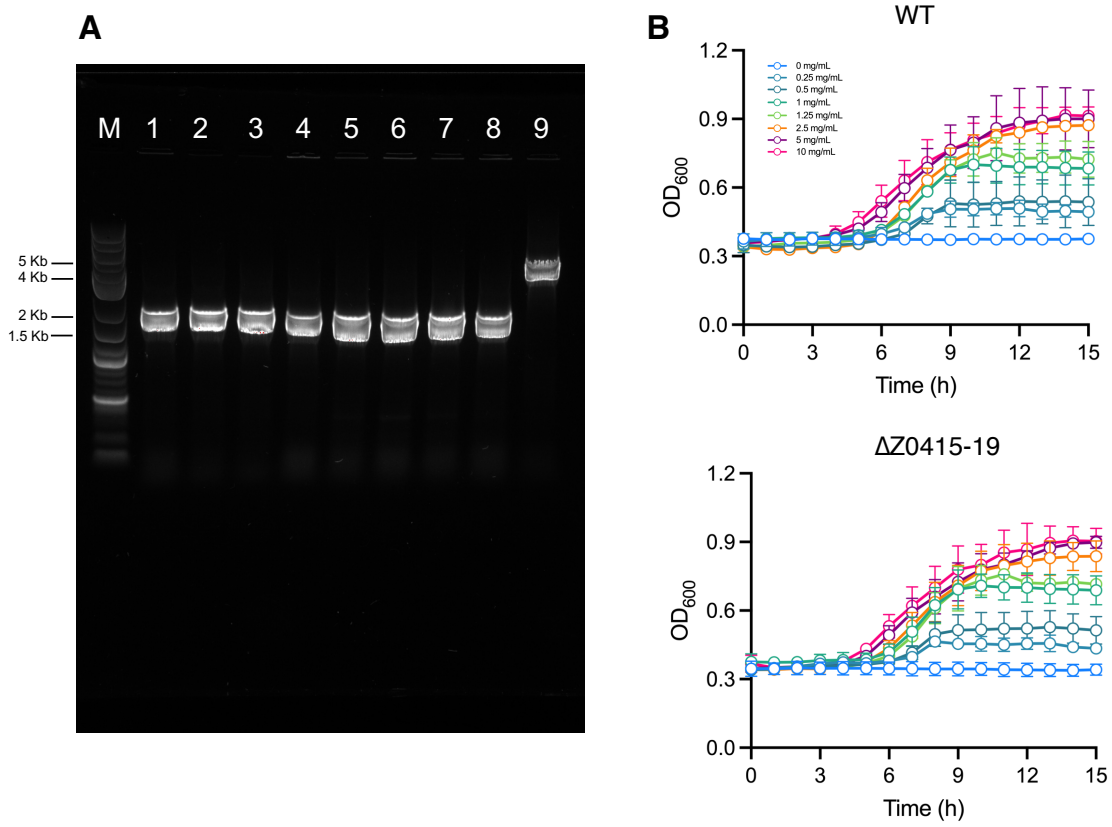


Figure 4-10. Deletion of Z0415-19 in TUV93-0 does not impact growth in minimal media supplemented with L-arabinose. (**A**) Gel electrophoresis confirmation of ΔZ0415-19. Lanes 1-8 correspond to colonies tested for replacement of Z0415-19 with the Kan^R cassette (1.6 Kb). Lane 9 corresponds to the WT gDNA control. Lane M corresponds to the 1 Kb plus DNA ladder. (**B**) Profiles for WT and Z0415-19 when grown in M9 minimal media supplemented with L-arabinose over 15 h. $n = 3$ biological replicates and error bars indicate SD.

4.6. AraE is the major route of L-arabinose uptake in EDL933

It was next hypothesised that a lack of phenotype for ΔZ0415-19 was likely due to at least two additional routes of L-arabinose uptake (AraE and AraFGH) remaining in the mutant background, therefore masking the true effects of its deletion. To better understand the role of L-arabinose transport in EHEC, *araE* and *araFGH* were independently deleted from the chromosome of TUV93-0 (**Figure 4-11AB**).

Despite being considered the main route of L-arabinose uptake, growth on L-arabinose as the sole carbon source was unaffected in the Δ*araFGH* mutant (**Figure 4-11C**). In contrast, growth of the Δ*araE* mutant on L-arabinose was heavily impaired to such an extent that no growth was observed at concentrations of L-arabinose lower than 5 mg/mL (**Figure 4-11D**). Even when grown at concentrations higher than 5 mg/mL the Δ*araE* mutant displayed an extensive lag phase and was unable to reach the same OD_{600} as the WT over the entire 15 h. The independent deletion of these transporters was therefore insufficient in completely preventing the growth of TUV93-0 on L-arabinose as sole carbon source.

Whether AraFGH was compensating for the deletion of *araE* and allowing for growth at concentrations of L-arabinose higher than 5 mg/mL was next explored. Deleting *araFGH* in the Δ*araE* mutant was found to be insufficient in preventing growth on L-arabinose at these higher concentrations (**Figure 4-11E**). These data therefore supported the presence of additional routes for L-arabinose transport in TUV93-0. For this reason, it was thought that the ability of TUV93-0 to still grow at these concentrations in the absence of *araE* and *araFGH* might be attributable to Z0415-19. However, following deletion of Z0415-19 in the Δ*araE*/Δ*araFGH* background, growth was still observed at concentrations of L-arabinose higher than 5 mg/mL (**Figure 4-11F**).

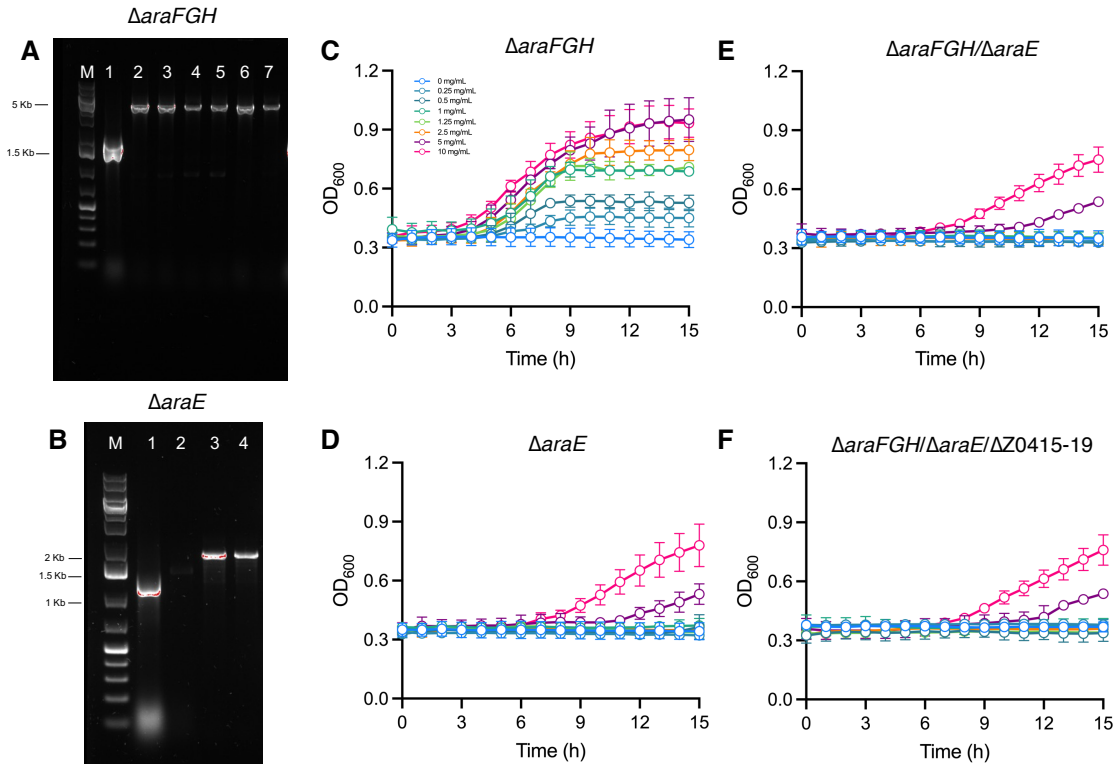


Figure 4-11. Deletion of Δ*araE* in TUV93-0 has the greatest effect on fitness when grown on L-arabinose. (**A**) Gel electrophoresis confirmation of Δ*araFGH*. Lanes 1-6 correspond to colonies tested for replacement of *araFGH* with the Kan^R cassette (1.6 Kb). Lane 7 corresponds to the WT gDNA control. (**B**) Gel electrophoresis confirmation of Δ*araE*. Lanes 1-3 correspond to colonies tested for replacement of *araE* locus the Cm^R cassette (1.1 Kb). Lane 4 corresponds to the WT gDNA control. Lane M for both gels correspond with the 1 Kb plus DNA ladder. Growth profiles for (**C**) Δ*araFGH*, (**D**) Δ*araE*, (**E**) Δ*araE*/Δ*araFGH* and (**F**) Δ*araE*/Δ*araFGH*/ΔZ0415-19 mutants grown in M9 minimal media supplemented with L-arabinose over 15-h. *n* = 3 biological replicates and error bars indicate SD.

4.7. An SNP is present in the ATPase of the Z0415-19 EDL933 transporter

Given the lack of an observable phenotype in TUV93-0 and to explore the functionality of this system more widely, a small subset of candidate *E. coli* strains were selected as representatives for a given pathotype (e.g., UPEC, MNEC and EPEC). The locus was found to be present in the MNEC str. CE10, but not in the UPEC str. CFT073 and EPEC str. E2348/69 selected. However, despite CE10 carrying the locus, differences could be observed in its architecture upon comparison with EDL933. Instead of carrying two genes predicted to encode ATPase-related proteins, CE10 displayed a tripartite arrangement commonly observed for ABC transporters, whereby the ATPase is encoded by a single gene.

Intra-strain analyses also revealed differences in the carriage of the ATPase gene across 17 additional EHEC strains, retrievable from BioCyc. Of these 17 strains, 5 (29%) were identified to encode the ATPase across two asymmetrically sized genes, displaying the same architecture as EDL933 (**Figure 4-12**). However, most strains exhibited the tripartite arrangement identified in CE10. Regardless of the variation in the ATPase genes, the remaining genes of the Z0415-19 transporter were conserved in their arrangement.

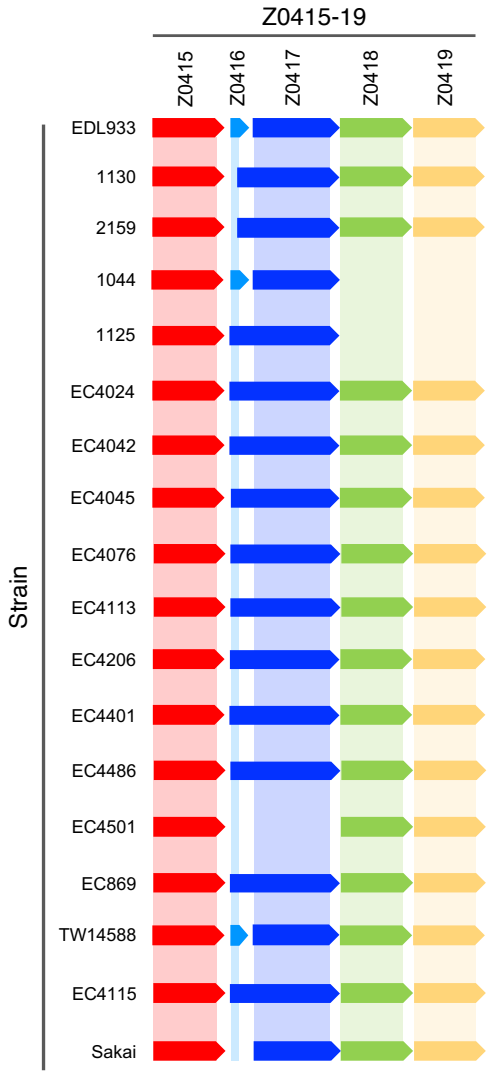


Figure 4-12. The ATPase of Z0415-19 is not consistently encoded by a single gene in O157:H7 strains. The architecture of Z0415-19, and occurrence of a single, or double set of ATPase encoding genes, was assessed across 18 O157:H7 strains (including EDL933). Loci used were taken from all strains available in the BioCyc DB. Absence of an arrow does not strictly suggest that a gene of the locus is not present, but rather indicates no ORF was annotated.

4.8. Exploring the significance of the ATPase SNP

The nucleotide sequence of Z0416/17 was explored to identify any differences that might account for the architectural variation observed. Multiple sequence alignments of the ATPase nucleotide sequences revealed a single nucleotide difference at position 259 between EHEC strains carrying two separate genes for the ATPase, and those carrying a single gene (**Figure 4-13**). The substitution could also be observed when the Z0416/17 sequence was aligned with the sequence of the CE10 ATPase gene (CE10_RS01535) (not shown). The SNP identified was a cytosine to thymine (259C>T) substitution, thus resulting in a STOP codon (TAG), as opposed to a codon for the amino acid glutamine (CAG). Notably, thymine, and therefore a STOP codon, was found at position 261 in all strains annotated to have their ATPase encoded by two genes.

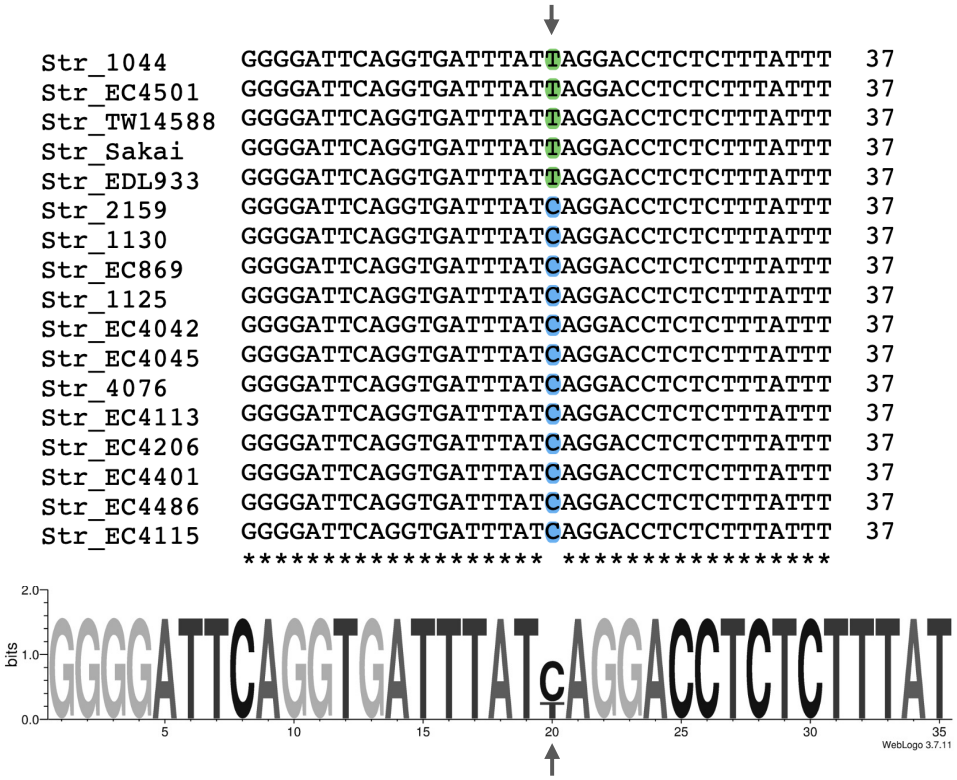


Figure 4-13. An SNP is present in the ATPase of Z0415-19 across O157:H7 strains. Part multiple sequence alignment of the ATPase across the 18 available O157:H7 strains in the BioCyc DB. Asterisks (*) denote residues that are completely conserved across sequences. The arrow highlights the position in the sequence the nucleotide substitution has occurred.

Genome sequencing has massively advanced since the original sequencing of the EDL933 genome, and searches in the NCBI database revealed amendments have been made to its annotation. Similar circumstances have also been observed for *E.*

coli str. Sakai following identification of > 50 single base errors, 11 sites of insertion and 10 sites of deletion after re-sequencing by Illumina MiSeq and PacBio in 2018. Although unlikely, due to the occurrence of the SNP across > 25 % of the strains analysed, sequencing of this allele was used to confirm the substitution event, and that nucleotide differences were not due to historical sequencing errors. A substitution event (nonsense mutation) of this kind would call for a premature STOP codon and truncate the translated protein. However, Z0417 was identified to have an alternative start codon (GTG) (**Figure 4-14A**), and it was therefore hypothesised that both Z0416 and Z0417 would be expressed and potentially interact. This was based on the provision that the region lost between the stop codon of Z0416 (conferred by the premature STOP codon) and alternative start codon of Z0417, did not encode for any functionally important residues.

Searches for the loss of highly conserved residues known to be important for ATPase activity were conducted through annotation and comparison of the Z0416/17 amino acid sequence. The CE10 ATPase (RS01535) amino acid sequence was also included in the analyses as a reference (**Figure14A-C**). Found in all ABC NBDs, the Walker A (GxxGxGK), Walker B ($ø_4$D) and Signature motifs (LSGGQ) were conserved, and unaffected by the SNP (**Figure 4-14C**). Additionally, both D- and H-loops required for ATP-hydrolysis were unaffected. However, the glutamine (Q) of the Q-loop; important for nucleotide binding and interaction with the TMDs, was identified to be absent (**Figure 4-14C**). Usually encoded by CAG, the SNP in TUV93-0 (C→T substitution) meant that this glutamine was lost.

Using AlphaFold, the Z0416 and Z0417 proteins were modelled to show the predicted location of the described motifs above, as well as highlight any predicted structural discrepancies when compared with RS01535. Modelling of RS01535 revealed regions of the protein designated Z0416 and Z0417 in EDL933 to be directly joined by a short stretch of 16 amino acids (**Figure 4-15**). This region also included the conserved Q87 of the Q-loop, that was identified to be absent via sequence alignment (**Figure 4-14C**). When overlaid, the Q-loop and linking region of RS01535 looked to lie in proximity with the TMD helices of Z0418, forming a potential interface between the two domains (**Figure 4-15B**).

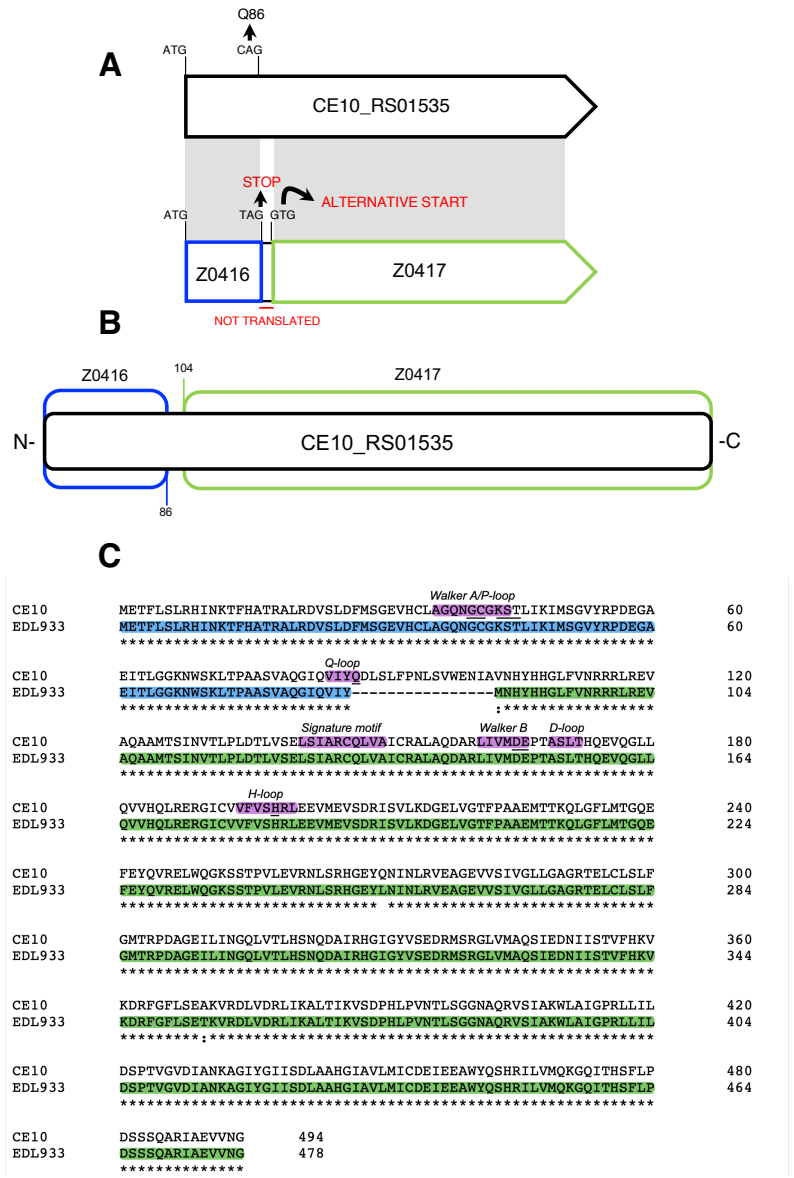


Figure 4-14. Functional residues are lost in the Z0415-19 ATPase. (**A**) In the full-length ATPase of CE10 there is no 259C>T substitution and the codon for glutamine (CAG) is unaffected. In EDL933, 259C>T and downstream alternative start codon (GTG) results in two asymmetrical ATPase genes. Regions conserved across ATPase genes are indicated by the grey box. (**B**) Schematic of the CE10 ATPase with the regions corresponding to Z0416 and Z0417 in EDL933 highlighted. The region between residues 86 and 104 in RS01535 are absent in EDL933. (**C**) Multiple sequence alignment of ATPase amino acid sequences for both EDL933 and CE10. Regions that correspond to Z0416 are highlighted in blue, and regions that correspond to Z0417 are highlighted in green. Functionally important motifs are highlighted in purple, with key residues underlined in black. Asterisks (*) denote residues that are completely conserved across sequences, whilst dashes (-) indicate the absence of aligned residues.

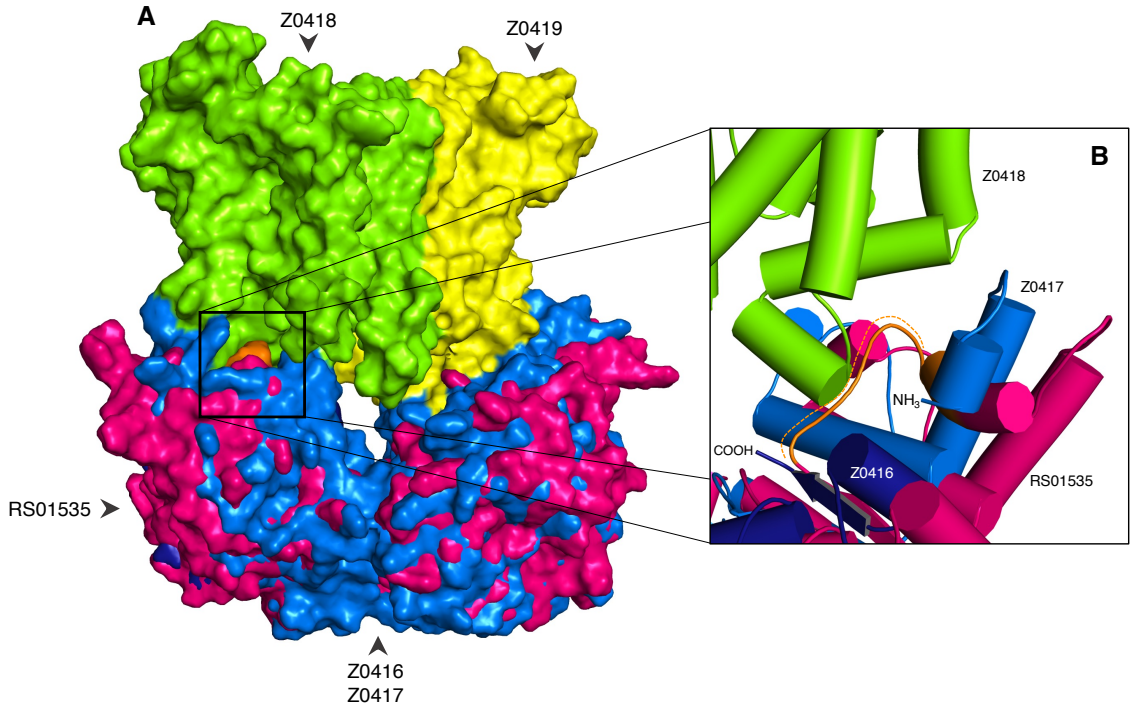


Figure 4-15. Overlay of the EDL933 Z0415-19 ATPase with the ATPase of CE10. (**A**) Surface view of the EDL933 transporter TMDs Z0418 (Green) and Z0419 (Yellow) with ATPase comprised of Z0416 (Purple) and Z0417 (Blue). The equivalent ATPase (RS01535) of CE10 is overlayed with the EDL933 ATPase shown in pink. Region of the CE10 ATPase found to be absent in the EDL933 ATPase as result of the SNP is highlighted in orange. The PBP (Z0415) is not shown; (**B**) Detailed view of the missing 16 amino acids in the EDL933 ATPase present in RS01535 (Orange). Proteins are coloured and labelled respectively with panel A. Both the C-terminus of Z0416 and N-terminus of Z0417 are labelled. Helices are represented by cylinders for simplicity. Models were generated using AlphaFold2 and images generated in PyMOL.

4.9. Analysis of SNP carriage across the *E. coli* phylogeny

To explore SNP occurrence across the *E. coli* phylogeny, a search was conducted to determine those strains carrying the 259C>T substitution, using the sample of *E. coli* strains previously used in the Z0415-19 carriage analyses (See *section 3.9.*). Across the six phylogroups (A, B1, B2, D, E, F), the SNP event did not appear widespread, and was restricted to Z0415-19 positive strains of phylogroup E only (**Figure 4-16**). Specifically, 86/216 (39.8 %) Z0415-19 positive strains of phylogroup E carried the SNP, suggesting it to likely be of clonal origin.

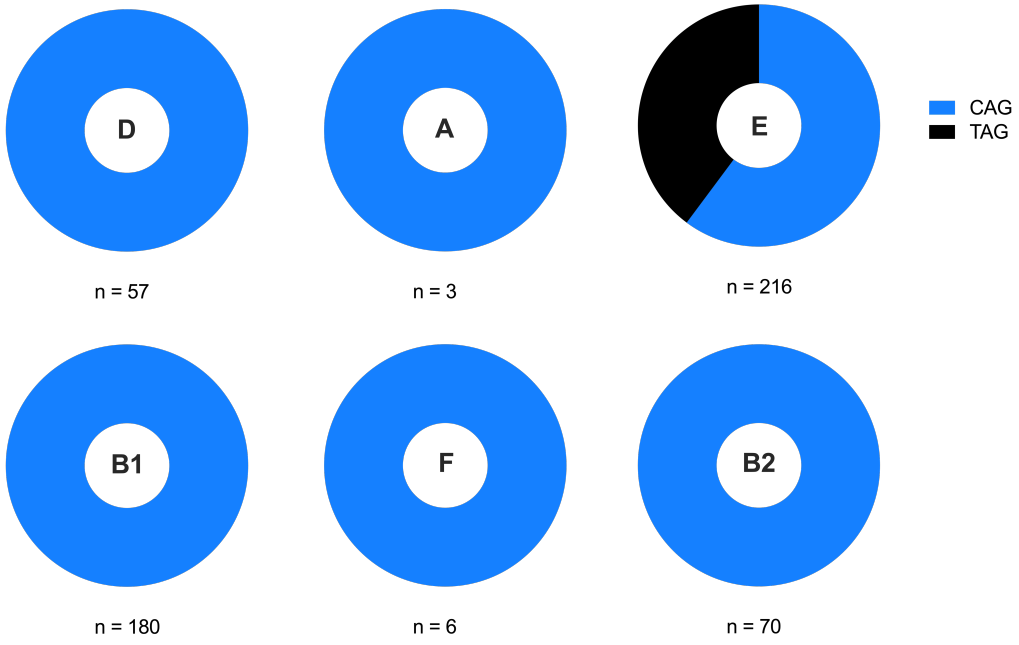


Figure 4-16. The SNP of the Z0415-19 ATPase in EDL933 is isolated to phylogroup E. Overview of strains positive for Z0415-19 carriage across the *E. coli* phylogeny carrying the SNP. 'n' denotes the total number of strains belonging to each phylogroup.

4.10. The Z0415-19 ATPase SNP has functional implications

Although a nonsense mutation had been identified in EDL933, instances of STOP codon readthrough have been reported in *E. coli* (Zhang et al., 2020). To rule out the possible redundancy of the premature STOP codon, work sought to determine whether the ATPase was truncated or not. A 3xFLAG-tag (DYKDDDDK) was added to the C-terminus of Z0417, to detect its expression (**Figure 4-17A**). The encoded ATPase, with its C-terminal FLAG-tag, was then probed for by immunoblotting of cell free extracts following growth in the presence and absence of 5 mg/mL L-arabinose. When probing,

the ATPase was only detected in the presence of L-arabinose, supporting the work in this thesis so far that Z0415-19 is expressed only in the presence of L-arabinose. However, expression of the protein appeared to be low. Notably, the corresponding band was ~54 kDa (**Figure 4-17B**), despite the theoretical molecular weight (MW) of Z0417 (+ 3xFLAG) being ~45 kDa. The combined MW of Z0416 and Z0417 was calculated to be 55 kDa, with Z0416 having a theoretical MW of 9 kDa, suggesting the ATPase to be expressed in its full length irrespective of the premature STOP codon.

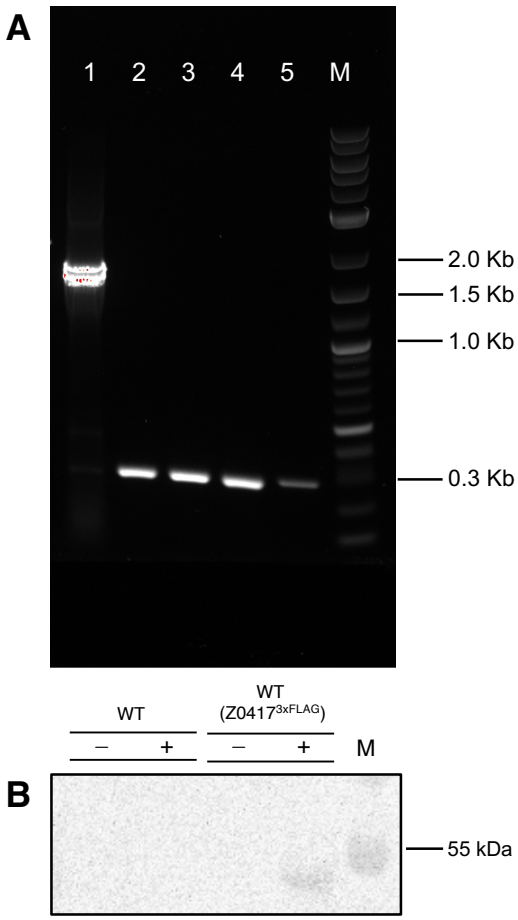


Figure 4-17. Z0417 is chromosomally expressed only in the presence of L-arabinose in TUV93-0. (**A**) Agarose gel confirmation of 3xFLAG-tagged Z0417 in TUV93-0. Lanes 1-4 correspond to colonies tested for correctly 3xFLAG-tagged Z0417. Lane 5 corresponds to the WT control. Lane M corresponds to the 1 Kb plus DNA ladder. (**B**) Immunoblot for TUV93-0 Z0417-3xFLAG expression when grown to an OD_{600} in MEM-HEPES supplemented with 5 mg/mL L-arabinose (+) compared to the untreated control (-). WT TUV93-0 was included as a control.

The ATPase size in non-SNP harbouring strains and the low chromosomal expression of Z0417 in TUV93-0 was next investigated. The ATPase of (a) CE10 (RS01535), (b) Z0417 of TUV93-0 and (c) Z0416-17 of TUV93-0 was cloned into pSUPROM to include

a C-terminal 3xFLAG-tag. This allowed for the constitutive expression of these genes in the absence of L-arabinose. Immunoblotting revealed an identical band of ~54 kDa for the Z0416-17 construct to that of the band for the chromosomally tagged Z0417 (**Figure 4-18A**). However, a band of ~54 kDa was also obtained for Z0417 being expressed from pSUPROM by itself. This ruled out the possibility of STOP codon readthrough and that the band seen for the chromosomally tagged Z0417 was not the expressed full-length ATPase (both Z0416 and Z0417).

Upon comparison with the construct for the ATPase of CE10, further differences could be observed. The ATPase was again larger (~65 kDa) than its calculated theoretical MW of ~54 kDa (**Figure 4-18A**). Chromosomal tagging of the CE10 ATPase returned the same result with a band of ~65 kDa, confirming the differences in MW were not the outcome of the ATPase being expressed from pSUPROM (**Figure 4-18B**). Nonetheless, comparison of the EDL933 and CE10 ATPase demonstrated there to be clear differences in the size of the full-length and SNP-harbouring ATPase (**Figure 4-18**), supporting the hypothesis that the SNP would truncate the ATPase of EDL933.

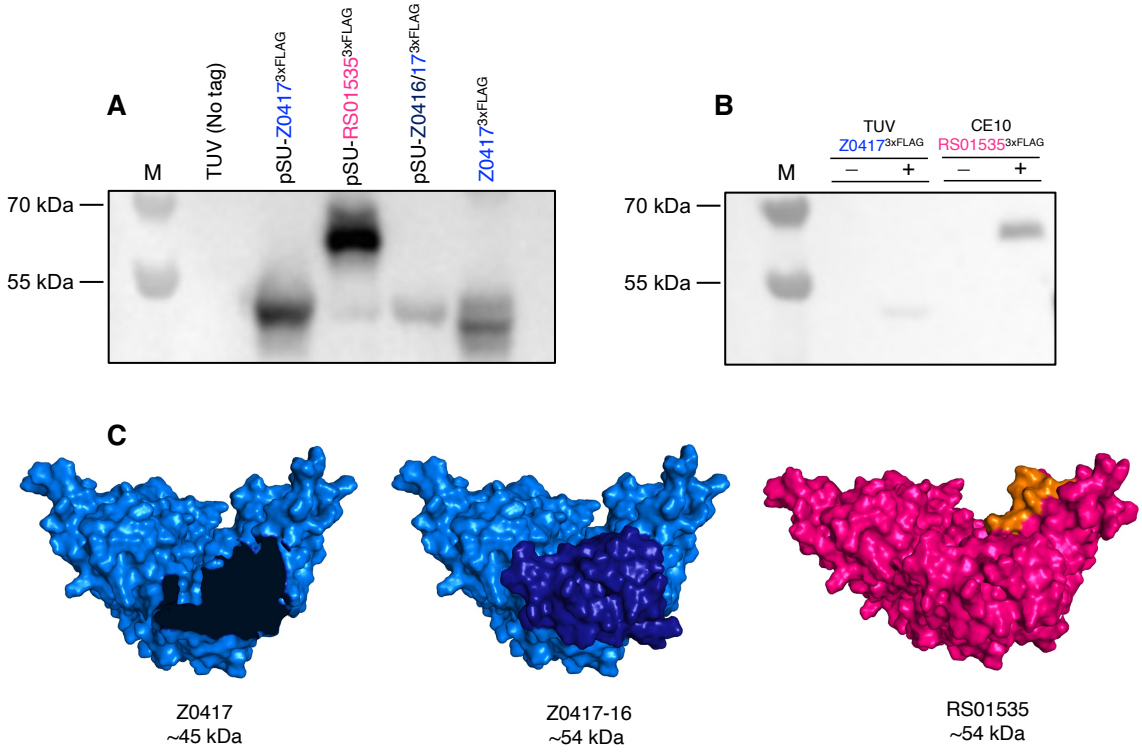


Figure 4-18. Differences can be observed in the size of the expressed ATPase between SNP and non-SNP harbouring *E. coli* strains. (**A**) Immunoblot of 3xFLAG-tagged ATPase expression from pSUPROM in O157:H7 str. TUV93-0 and O7:K1 str. CE10 when grown to an OD_{600} of 0.7 in MEM-HEPES. WT TUV93-0 (lane 1) and chromosomally 3xFLAG-tagged Z0417 (lane 5) whole cell lysates were also included as control. Lane M corresponds to the protein ladder. (**B**) Immunoblot of the chromosomally 3xFLAG-tagged ATPase in TUV93-0 and CE10. Strains were grown in MEM-HEPES to an OD_{600} of 0.7, supplemented with 5 mg/mL L-arabinose (+), compared to the untreated control (-). (**C**) Overview of the predicted ATPase proteins expressed from each of the constructs used with (approximate) theoretical MW labelled). The region absent in EDL933 but present in CE10 and responsible for adjoining Z0416 and Z0417 is highlighted in orange. Models were generated in PyMOL.

4.11. Assessing the role of Z0415-19 in *E. coli* strains with an intact ATPase

The inability to hydrolyse ATP due to a non-functional ATPase would render the rest of the ABC transporter defective and might explain the lack of a growth phenotype in TUV93-0 at transporting target substrates. To address this, the equivalent locus to Z0415-19 was deleted in the EHEC str. ZAP193 (Δ0432-35) (**Figure 4-19A**). Additionally, the locus was deleted in the MNEC O7:K1 str. CE10 (Δ01530-45), previously used in the identification of the SNP, as non-EHEC comparator (**Figure 4-19B**). When grown on L-arabinose as the sole carbon source across a range of concentrations, neither ΔZ0415-19 from ZAP193 or CE10 displayed any growth defects upon comparison with the corresponding parent strain (**Figure 4-19C**). Therefore, regardless of whether the SNP is present or not, deletion of Z0415-19 had no apparent effect on the transport of L-arabinose under these conditions.

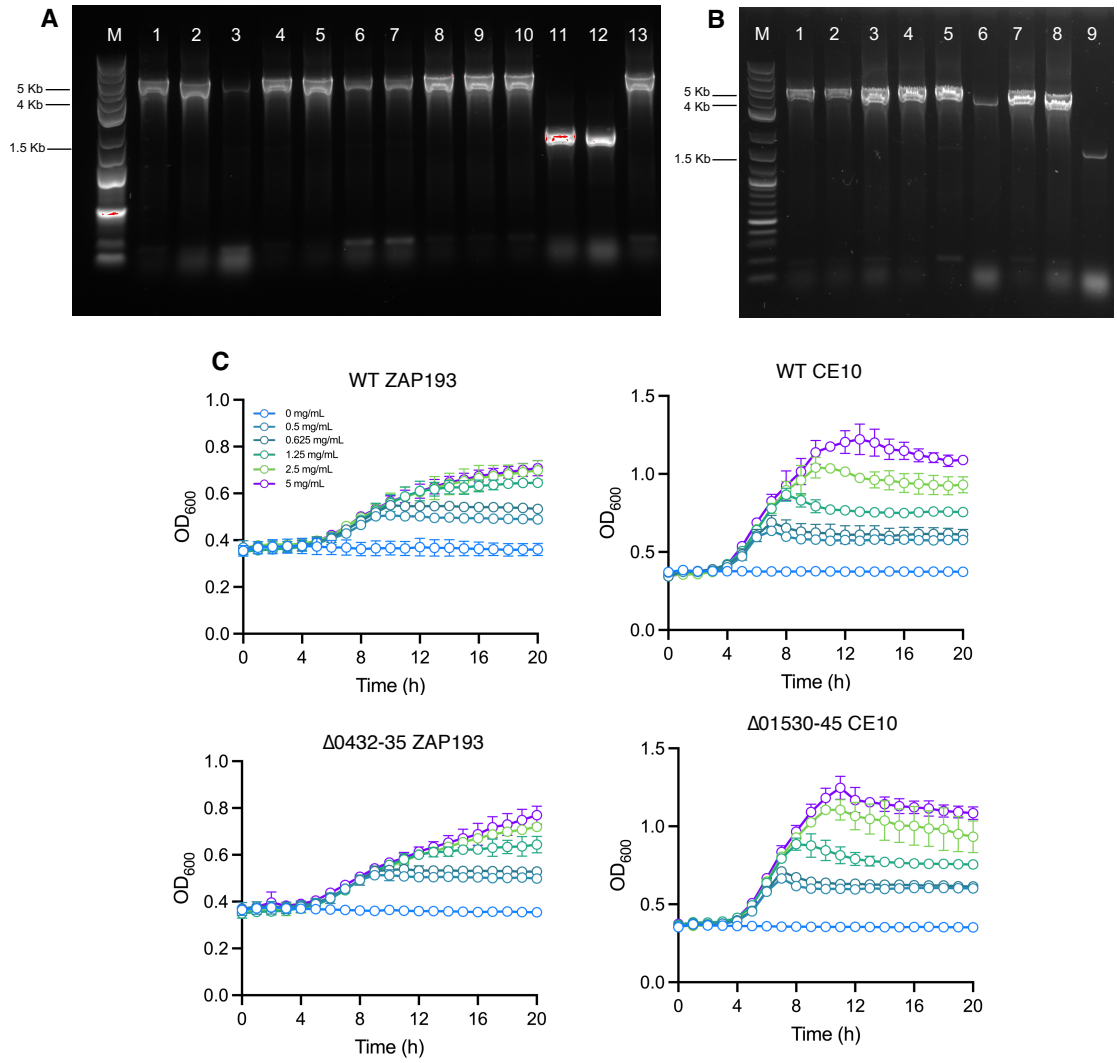


Figure 4-19. Deletion of Z0415-19 in ZAP193 and CE10 has no effect on growth in minimal media with L-arabinose as the sole carbon source. (**A**) Agarose gel confirmation of ΔZ0415-19 in ZAP193 (Δ0432-0435). Lanes 1-12 correspond to colonies tested for replacement of Z0415-19 with the Kan^R cassette (1.6 Kb). Lane 13 corresponds to the WT gDNA control. (**B**) Agarose gel confirmation of the ΔZ0415-19 deletion in CE10 (Δ01530-45). Lane 1 corresponds to the WT gDNA control. Lanes 2-9 correspond to colonies tested for replacement of Z0415-19 with the Kan^R cassette. Lane M for both gels corresponds with the 1 Kb plus DNA ladder. (**C**) Growth profiles for WT and Z0415-19 ZAP193 and CE10 strains when grown in M9 minimal media supplemented with L-arabinose over 20 h. *n* = 3 biological replicates and error bars indicate SD.

4.12. Growth of TUV93-0 on L-arabinose under anaerobiosis

Despite significant expression in the presence of L-arabinose, transport via Z0415-19 had not been observed under the conditions tested so far. Additionally, whilst there was no evidence of STOP codon readthrough, it was hypothesised that the proteins may still interact post-translationally and confer activity. Since the environment may be an important factor in governing Z0415-19 expression, conditions more closely mirroring the ecological context of EHEC were next investigated.

As the colonic environment is largely anaerobic, the ability of TUV93-0 to grow on L-arabinose in the absence of oxygen was first tested. Under minimal conditions, when supplemented with 5 mg/mL L-arabinose, TUV93-0 was able to grow anaerobically. Of note, supplementation of sodium nitrate ($NaNO_3$) was necessary to allow for growth and act as terminal electron acceptor. Growth on L-arabinose was found to improve significantly with increasing concentrations of $NaNO_3$ (**Figure 4-20**). 20 mM $NaNO_3$ was therefore selected as the optimal concentration for use in future experiments.

To determine if Z0415-19 was active under anaerobiosis, WT and ΔZ0415-19 was grown in M9 minimal media supplemented with L-arabinose (**Figure 4-21**). No significant differences were observed in the growth between strains for any concentration tested. Little differences were also seen in the overall growth across concentrations, with only the length of the lag phase varying. For example, the onset of exponential growth occurred ~2 h earlier at 5 mg/mL compared to at 1.25 mg/mL. It was therefore concluded that the absence of oxygen was not functionally important for transport of L-arabinose by Z0415-19.

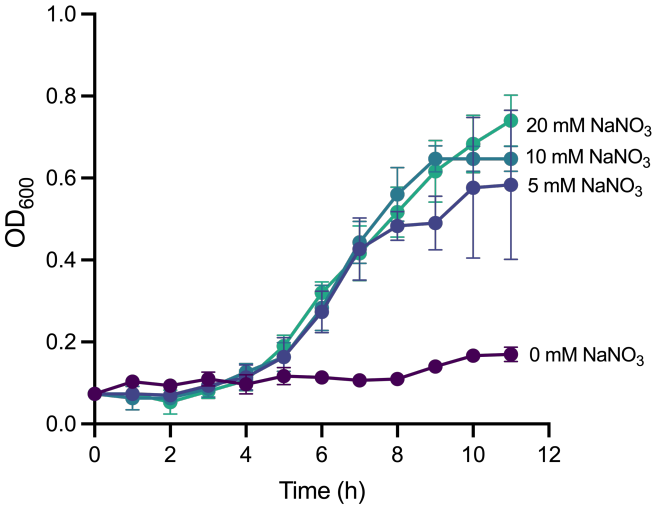


Figure 4-20. $NaNO_3$-dependent anaerobic growth of TUV93-0. WT TUV93-0 was grown in M9 minimal media supplemented with 5 mg/mL L-arabinose and $NaNO_3$ at various concentrations. OD_{600} was measured manually hourly for 11 h. All growth was conducted under anaerobic conditions at 37 °C. *n* = 3 biological replicates and error bars indicate SD.

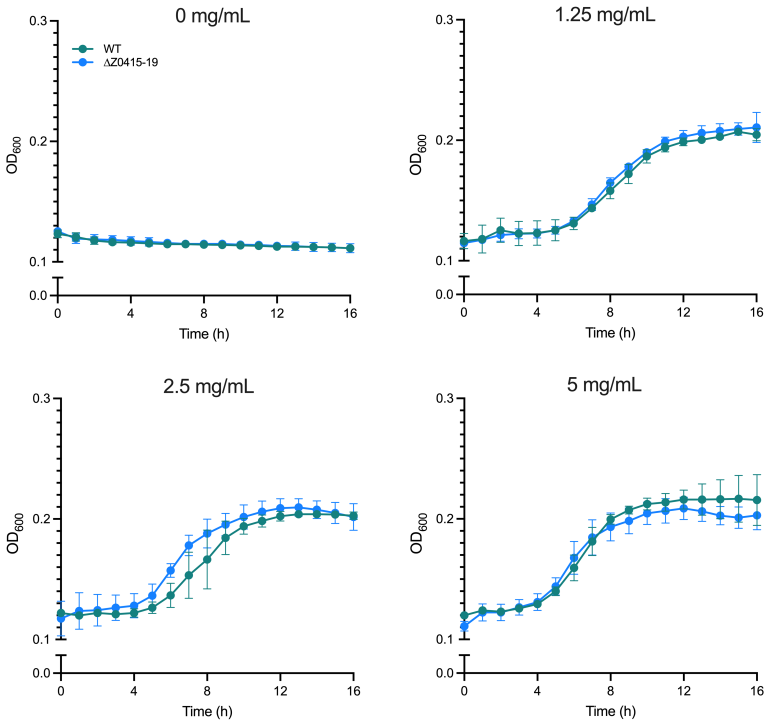


Figure 4-21. Anaerobic growth of WT and ΔZ0415-19 TUV93-0 strains on L-arabinose. WT and ΔZ0415-19 were grown in M9 minimal media (+ 20 mM $NaNO_3$) supplemented with 0, 1.25, 2.5 and 5 mg/mL L-arabinose. All growth was conducted under anaerobic conditions at 37 °C. *n* = 3 biological replicates and error bars indicate SD.

4.13. Competition for L-arabinose between EHEC and *Bt*

Due to the highly competitive environment of the human gut, pathogens must outcompete the native gut microbiota if they are to benefit from a given nutrient and establish a novel niche. To better understand potential competition with the gut microbiota and the role of L-arabinose transport during anaerobic growth on L-arabinose, EHEC was co-grown *Bt*. The minimal requirements necessary to support the growth of *Bt* differs to those for EHEC. Subsequently, TUV93-0 growth in BMM when supplemented with L-arabinose (+ 20 mM $NaNO_3$) was confirmed prior to co-growth with *Bt* (**Figure 4-22**). Despite the aerobic growth of EHEC on glucose being much faster than on L-arabinose, growth on the two sugars under anaerobic conditions were found not to be as dissimilar.

Under minimal conditions, both TUV93-0 and *Bt* were able to grow on L-arabinose as a sole carbon source in monoculture (**Figure 4-23**). However, comparisons between the growth of *Bt* and TUV93-0 revealed the former to have a higher mean specific growth rate and maximum OD_{600} reading of 0.21 h^{-1} and 1.93 respectively (**Figure 4-23A**). The specific growth rate and maximum OD_{600} reached by TUV93-0 was 0.15 h^{-1} and 1.5 respectively (**Figure 4-23A**). These differences were similarly reflected in the CFU/ml counts for both monocultures (**Figure 4-23B**).

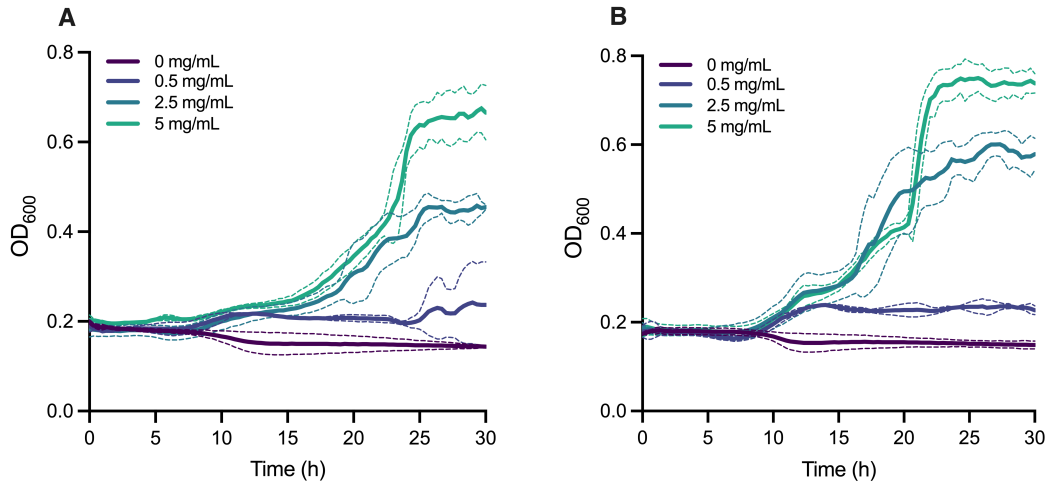


Figure 4-22. Anaerobic growth of TUV93-0 in BMM. WT TUV93-0 was grown in BMM supplemented with (**A**) glucose and (**B**) L-arabinose under anaerobic conditions over 30 h. *n* = 3 biological replicates and dashed lines indicate SD.

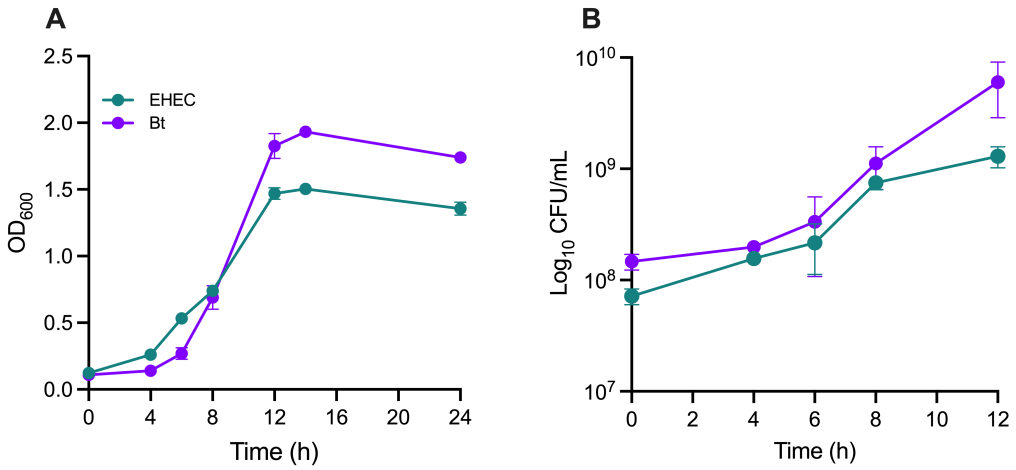


Figure 4-23. *Bt* displays faster growth on L-arabinose than TUV93-0. *Bt* and TUV93-0 were anaerobically grown in BMM supplemented with 20 mM $NaNO_3$ and 5 mg/mL L-arabinose. Growth was determined by measuring (**A**) the OD_{600} and (**B**) CFU/mL counts. $n = 3$ biological replicates and error bars indicate SD.

In co-culture, TUV93-0 displayed a similar growth rate to that seen in monoculture (**Figure 4-24A**). The growth of *Bt* on the other hand was greatly impaired and dramatically fell post 4 h of co-growth with TUV93-0 (**Figure 4-24A**). Subsequently, between 4 h and 12 h of co-growth, significant differences in the CFU/mL of each strain were observed ($P < 0.01$), with EHEC continuously outgrowing *Bt*. The calculated competitive index confirmed EHEC to have a significant competitive advantage and outcompete *Bt* across all timepoints, including 4 h, despite no significant difference in the CFU/mL being observed at this timepoint (**Figure 4-24B**). A comparison between the final CFU/mL counts (12 h) of *Bt* across mono- and co-cultures revealed a > 300-fold decrease, suggesting co-growth with EHEC to negatively impact the growth of *Bt*.

A qualitative assessment of L-arabinose utilisation over time by TLC (**Figure 4-24C**), across both mono- and co-cultures, affirmed the count data. In monoculture, TUV93-0 had completely used all the L-arabinose supplemented into the media approximately 4 h earlier than *Bt*. Thus, L-arabinose was no longer detectable by TLC post 8 h of TUV93-0 growth, whilst its presence was still observed at 12 h for *Bt*. This was also reflected in the spent media from co-cultures, whereby L-arabinose was no longer detectable post 8 h of co-growth. Visually, there was less L-arabinose detectable from 6 h onwards in co-culture than in the monoculture of either strain (**Figure 4-24C**).

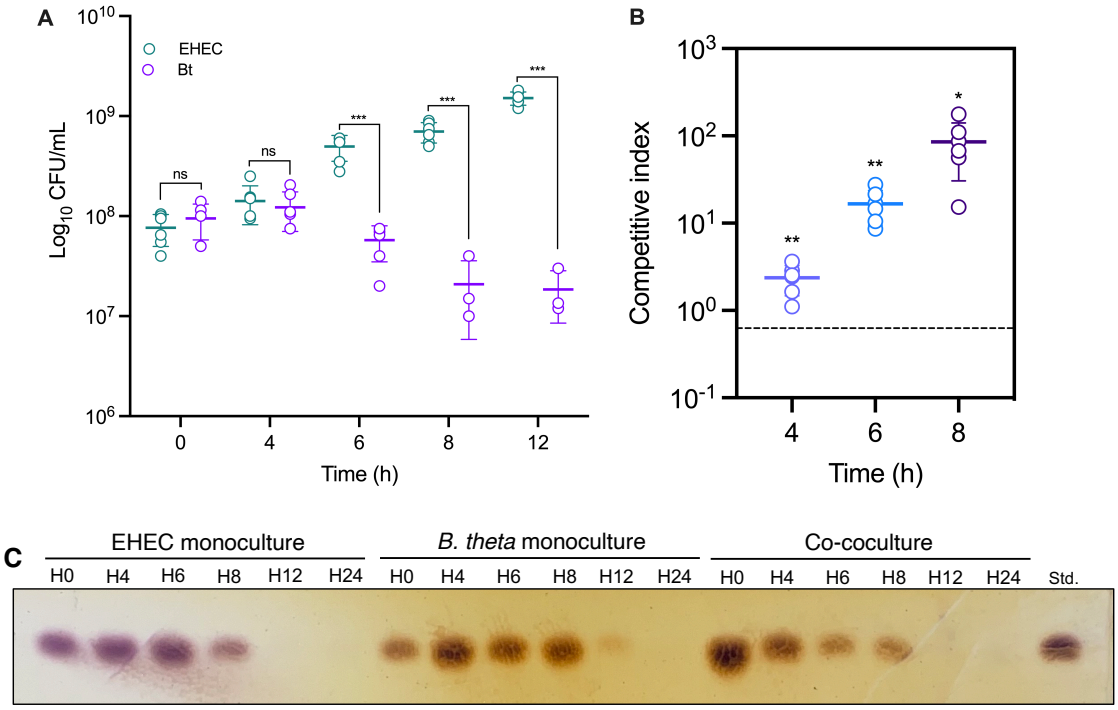


Figure 4-24. TUV93-0 outcompetes *Bt* in co-culture when grown on L-arabinose as the sole carbon source. *Bt* and TUV93-0 were anaerobically grown together 1:1 in BMM supplemented with 20 mM $NaNO_3$ and 5 mg/mL L-arabinose. (**A**) CFU/mL counts for TUV93-0 and *Bt* over 12 h. (**B**) Calculated competitive indices for *Bt* and TUV93-0. The dashed line denotes the point of which there is no difference in strain fitness. Points above the dashed line indicate TUV93-0 to have a competitive advantage in co-culture. n = 6 biological replicates and error bars indicate SD. (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$ - calculated using a student's *t*-test. (**C**) Qualitative assessment of arabinose uptake across both mono- and co-growth of TUV93-0 and *Bt* over 24 h. Spent media from each of the corresponding conditions was spotted onto a TLC silica plate and run alongside a 5 mM L-arabinose standard.

To further explore the phenomenon by which *Bt* growth was being so negatively affected when co-grown with TUV93-0 on L-arabinose, it was hypothesised that the impaired uptake of the sugar in TUV93-0 would limit its advantage. Previous growth experiments of TUV93-0 mutants in the transporters specific for L-arabinose uptake had shown AraE to be most influential, and its deletion to negatively impact growth on the sugar. Furthermore, deletion of Z0415-19 under anaerobic conditions had no effect on growth, so the observed growth advantage was not likely conferred by the transporter. With the Δ*araE* mutant retaining some capacity to aerobically grow on L-arabinose at higher concentrations, just at a slower rate, Δ*araE* TUV93-0 was grown in co-culture with *Bt* and will herein be the main transporter discussed.

In monoculture, Δ*araE* TUV93-0 displayed an extensive lag in growth (**Figure 4-25AB**), as seen for growth under aerobic conditions (**Figure 4-11D**). When co-grown, the dramatic reduction in CFU/mL previously seen for *Bt* was no longer observed, with there being no significant difference in the counts between the two strains, apart from at 6 h ($P < 0.05$) (**Figure 4-25D**). The overall trend in *Bt* growth was shown to increase over time but remained lower than the growth seen in monoculture (**Figure 4-25CD**). The incapacity of EHEC via the deletion of *araE* limited its previously observed advantage allowing *Bt* growth. This was reflected also in the calculated competitive indices where there was a no longer a difference between the growth of strains (**Figure 4-25E**). The juxtaposition of *Bt* growth when co-grown with WT or Δ*araE* TUV93-0, the impact on counts could be visibly noted (**Figure 4-25F**). These findings indicate the crucial role of L-arabinose in enabling EHEC to outcompete *Bt*. Moreover, the importance of multiple transporters for improved nutrient scavenging in the human gut is recognised, likely facilitating niche establishment and subsequent colonisation.

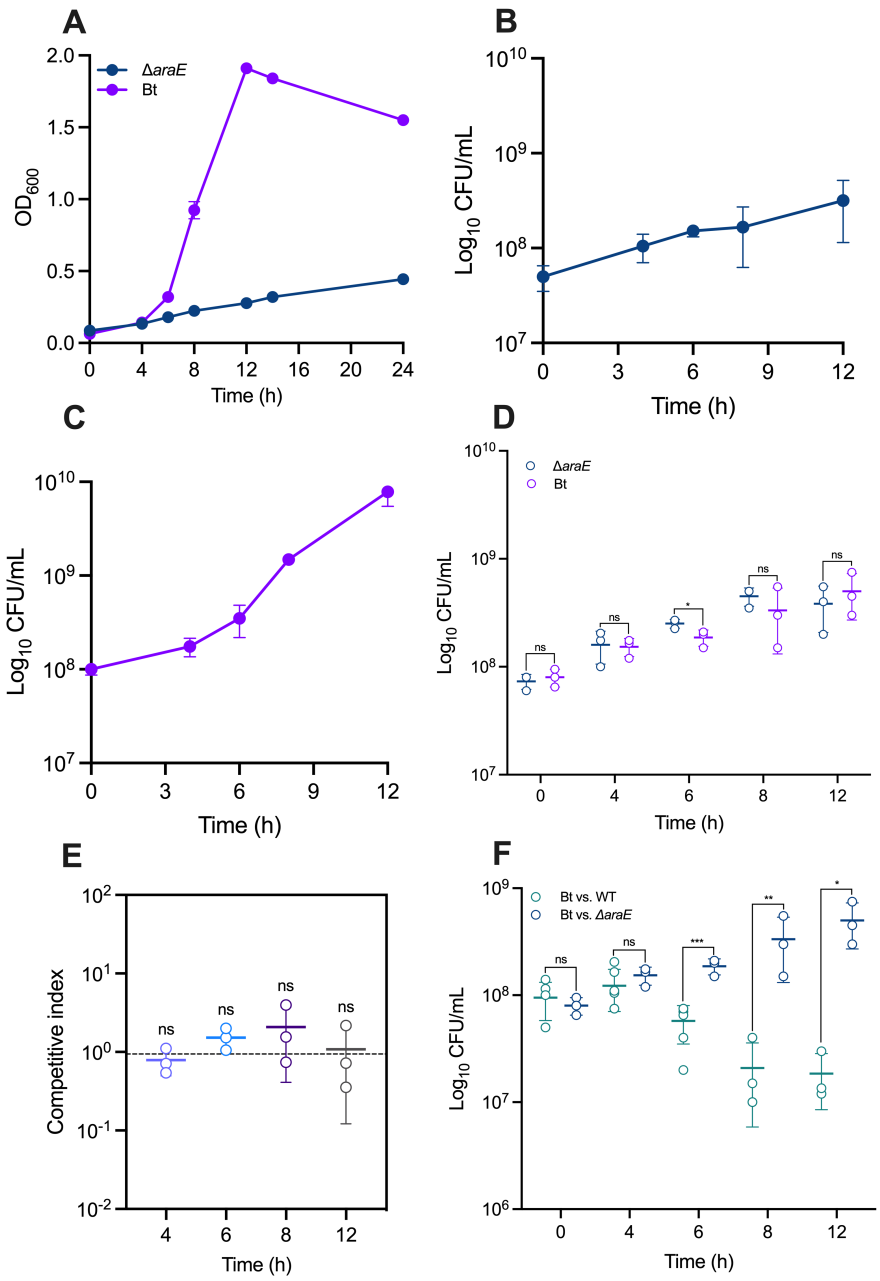


Figure 4-25. Deletion of *araE* prevents the ability of TUV93-0 to outcompete *Bt* for L-arabinose. *Bt* and Δ*araE* TUV93-0 were anaerobically grown in BMM supplemented with 20 mM $NaNO_3$ and 5 mg/mL L-arabinose. For monocultures, growth was determined by measuring (**A**) the OD_{600} and (**B,C**) CFU/mL counts. (**D**) CFU/mL counts for each strain when grown in co-culture after being inoculated 1:1. (**E**) Calculated competitive indices for *Bt* and TUV93-0. The dashed line denotes the point of which there is no difference in strain fitness. (**F**) Comparison between the growth of *Bt* in co-culture with the WT TUV93-0 and Δ*araE* TUV93-0 strains. $n = 3$ biological replicates and error bars indicate SD. (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$ - calculated using a student's *t*-test.

4.14. Discussion

The aim of this chapter was to explore the role of the EHEC Z0415-19 system identified in chapter 3. Using a combination of bacterial genetics and *in silico* analyses Z0415-19 was shown to be expressed exclusively in the presence of L-arabinose, in a manner that is dependent on the transcriptional regulator, AraC. It is likely that Z0415-19 is regulated via a mechanism reflective of those used to control the expression of the canonical L-arabinose transport systems by AraC in *E. coli*. Whether the Z0415-19 transporter is responsible for transporting L-arabinose remains unclear, due to no obvious phenotype being observed following its deletion and growth in minimal media with L-arabinose as a sole carbon source. Evaluation of the canonical L-arabinose transport system revealed AraE to act as the key player, with its deletion entirely disrupting growth on the sugar. In competition with gut *Bacteroides* for L-arabinose, EHEC possesses a competitive advantage which is driven by the presence of AraE.

In line with earlier discussions of this thesis, novel, and additional sugar uptake systems likely provide pathogens with the competitive advantage they require to colonise the gut. Whilst system identification is highly informative on the nutrients pathogens exploit for growth and niche establishment, it is important to understand how their expression is regulated to understand their function.

With transporters typically only being expressed in response to their cognate sugars being sensed, reporter-based assays are useful tools for investigating the dynamics of expression in response to the presence of a specified substrate (Aidelberg et al., 2014; Kaplan et al., 2008). Therefore, using the LUX reporter system previously described, Z0415-19 was identified to respond exclusively to L-arabinose in a manner that was concentration dependent. Media-dependent effects could also be observed, with both the induction and overall expression profile of Z0415-19 differing across media types. These differences were likely the outcome of catabolite repression, whereby the presence of glucose prevents L-arabinose utilisation by suppressing expression of the corresponding genes (Ammar et al., 2018).

The regulation of the canonical L-arabinose transport system has been extensively studied in *E. coli* K-12. On that basis, details of how L-arabinose regulates these transporters was applied to the transporter predicted to be encoded by Z0415-19 with

a focus on AraC. The regulatory sites of AUGs known for AraC-binding were identified upstream of the Z0415-19 start site and displayed encouraging sequence identity scores (35-60 %). Notably, variation has been identified in both the sequences of O and I sites of the canonical AUGs previously, with AraC being described as a regulator capable of binding degenerate sequence motifs (Stringer et al., 2014). Subsequently, AraC must be able to accommodate, recognise, and still bind regions irrespective of subtle differences in the sequence. It is likely that complete conservation in only specific residues of identifiable AraC-binding motifs is required, particularly considering the variability observed in medial residues. Furthermore, the DNA-binding domains of AraC are highly conserved across Enterobacteriaceae, suggesting similar DNA sequences are important (Stringer et al., 2014). In support of AraC being able to directly regulate genes not part of the canonical system, previous work has demonstrated the regulator to form a larger regulon in *E. coli* MG1655, with some genes of the regulon even having totally unrelated roles to L-arabinose utilisation (Stringer et al., 2014). Whilst it is easy to presume based on the presence of identifiable I_1 and I_2 sites that AraC regulates Z0415-19 expression through a looping mechanism similar to canonical AUGs (*araC* and *araBAD*), no O_2 site was found as part of these analyses. This can be explained by one of two ways, the first being that the O_2 site is present but poor conservation relative to MG1655 makes it difficult to identify via sequence analysis. The second being that the O_2 site is simply absent and that the regulation of Z0415-19 by AraC is conferred by a different mechanism to that of DNA looping, which is thought to be the case for the AraE transporter (Johnson & Schleif, 1995). Interestingly, the normal activation of AUG expression by AraC does not require the O_2 site and it can be deleted (Lobell & Schleif, 1990). In addition, the identification of a potential CRP binding site supports its role in regulating Z0415-19 as well as the probable effects of catabolite repression seen in the reporter assays in glucose containing media.

The major caveat to this model is based on the predicted location of the -10 and -35 sites required for RNAP recognition. As the identified I_2 site entirely overlaps with the -35 site, RNAP recognition could potentially be interfered with or completely blocked. That being said, these sites are based purely on sequence predictions, and have not been experimentally validated here. However, if these sites are correct then the overlap appears to cause no disruption to RNAP activity as expression of the locus

was observed via reporter assays and RT-qPCR. Regardless, the deletion of *araC* confirmed the absolute requirement for the regulator to be present for Z0415-19 expression. It is also apparent that the regulator must be bound with L-arabinose to induce expression. Without L-arabinose, it is unlikely AraC undergoes the conformational changes required to allow for binding at I sites (Lobell & Schleif, 1990). Whilst the effect of D-arabinose on Z0415-19 was tested, it is unlikely that the sugar has any role in the expression of the system, particularly as the WT AraC is unable to bind D-arabinose (Tang et al., 2008).

Ultimately to address the unknowns regarding AraC regulation of Z0415-19 further experimental work is required. This could include a more traditional approach seen for much of the original AraC work, whereby the region of interest (Z0415-19 promoter) is sequentially truncated in reporter assays to determine the location of important regulatory sites, inferred from a gain or loss of function (Dunn & Schleif, 1984). Moreover, electrophoretic mobility shift assays could be used to validate the binding of AraC to P_{Z0415}, with the exact binding regions later being determined by DNAse I footprinting. It may also be useful to explore the expression profile of Z0415-19 under more complex environmental conditions, particularly as the PBP has previously been shown to be upregulated during growth on lettuce lysates (Kyle et al., 2010). Additionally, Z0419 was identified to be an *in vivo* induced protein expressed during human infection (John et al., 2005). Therefore, activity of the system (i.e., transport) may be one that is dependent on factors present only during *in vivo* infection.

To date, L-arabinose transport is accepted to be governed by AraFGH and AraE in *E. coli.* In the human gut, both commensal and pathogenic *E. coli* are able to use L-arabinose as a carbon source (Fabich et al., 2008). Therefore, having an additional L-arabinose-specific transporter would presumably provide EHEC with a scavenging advantage. Notably, low affinity H^+-sugar symporters (e.g., AraE) have been previously claimed to be less important than ABC transporters (AraFGH) (Groff et al., 2012). However, in this investigation deletion of *araE* in TUV93-0 almost completely abolished the ability to grow on L-arabinose as a sole carbon source. In contrast, EHEC was still able grow as normal at high concentrations of the sugar independent of *araFGH* and Z0415-19, which was confirmed by their deletion. A lack of phenotype for ΔZ0415-19 was initially surprising given its strong induction by L-arabinose. However, this became

less surprising when considering the same phenotype was observed following the independent deletion of the canonical AraFGH ABC transporter. The capacity of EHEC to grow at high concentrations even when *araE*, *araFGH* and Z0415-19 were deleted (Δ*araE*Δ*araFGH*ΔZ0415-19) suggested that additional transporters for L-arabinose are found in EHEC. Indeed, examples of additional transporters have been reported to have potential implications in the utilisation of L-arabinose in *E. coli* M1655, including YtfQ and AraJ. YtfQ, is the PBP component of a galactose-specific ABC transporter known to bind L-arabinose but is repressed by AraC independent of L-arabinose (Horler et al., 2009; Stringer et al., 2014). AraJ on the other hand is part of the known L-arabinose regulon and is speculated to be involved in transport of the sugar (Reeder' & Schleif, 1991). The precise function of AraJ remains elusive with little research adding to its role since its initial identification by Reeder and Schleif (1991). Their initial work found *araJ* expression to be both AraC and CRP-dependent, however, its deletion and insertion had no detectable effects on the growth of *E. coli* on L-arabinose (Reeder' & Schleif, 1991). Due to the high sequence similarity of *araJ* to drug efflux proteins, more recent work has speculated the gene to encode an L-arabinose efflux pump involved in the homeostatic control of the L-arabinose system but concluded this not to be the case (Fritz et al., 2014). The potential presence of several L-arabinose specific transport routes suggests the sugar to be important in *E. coli*, even more so in EDL933 which carries Z0415-19 (**Figure 4-26**).

The ATPase can be likened to a motor that drives ABC transporter activity, and disturbances in its capacity to hydrolyse ATP are likely to hinder substrate transport. The identification of a SNP in the predicted ATPase sequence of Z0415-19 therefore indicated a possible loss of function in the encoded transporter, explaining the absence of a phenotype for growth on L-arabinose by ΔZ0415-19. Aside from the Q-loop, important motifs within the NBDs encoded across Z0416 and Z0417 were unaffected. The Q-loop is a conserved site that allows for interdomain communication (i.e., between NBDs and between NBD and TMD) and contributes to ATP-binding via Mg^{2+} through interaction with the glutamine residues (Westfahl et al., 2008). Previous work found that mutation of the Q-loop glutamine in the NBD of the *E. coli* MsbA ABC transporter had no effect on ATPase activity (Westfahl et al., 2008). However, in the case of the Z0415-19 ATPase, the glutamine (Q86) was not mutated to another amino acid but instead lost entirely via the C→T substitution coding for a premature STOP

codon. The higher molecular weight seen for Z0417 by Western blot analysis is highly unlikely to be due to STOP codon readthrough as the proteins size was the same as both Z0416-17 expressed from pSUPROM. It is similarly unlikely that if Z0416 and Z0417 are independently expressed, which is possible considering Z0417 has an alternative start codon, the difference in theoretical and actual MW of Z0417 is the outcome of the proteins forming a complex post-translation. Such an interaction would be broken under denaturing conditions. Though this could be possible under non-denaturing conditions.

To discern the ability of the ATPase to bind and hydrolyse ATP across two-independent domains that interact, irrespective of Q86 loss, functional data would be required. This could include the expression and purification of both Z0416 and Z0417 for use in ATPase assays. Additionally, bacterial two-hybrid assays would also allow the binding-interactions between Z0416 and Z0417, as well as with TMDs, to be further explored. If the domains are able to interact irrespective of the SNP, the transporter may well still be able to import substrates. That being said, no phenotype was observed following deletion of the transporter in non-SNP harbouring strains grown on L-arabinose. Although, it is important to note that both CE10 and ZAP193 mutants still had an intact L-arabinose transport system via AraE and AraFGH.

In summary, it remains uncertain whether the lack of a phenotype for ΔZ0415-19 when grown on L-arabinose is either due to disruption in its transport caused by the SNP or the sugar simply not being important under the conditions tested. In addition to these explanations, it is possible that L-arabinose is not the substrate of Z0415-19 and instead AraC is responsible for cross-regulating the transport of an alternative substrate. Although being an example of negative regulation, this has been seen for the D-xylose transport, which is repressed by AraC (Koirala et al., 2015). Based upon the high sequence identity shared with the PBP of the *C. rodentium* Rbl system, Z0415-19 may transport D-ribulose. Despite no growth being observed on D-ribulose in minimal media, L-arabinose would not be present under these conditions to induce *araC* required for the expression of Z0415-19. The rational for AraC regulating Z0415-19 is analogous to the required phosphorylation of D-ribulose by a ribulokinase which is necessary for growth on the sugar, as seen in *C. rodentium*. Although lacking a D-ribulokinase, EHEC does possess the L-ribulokinase, AraB, also regulated by AraC.

As discussed earlier, L-ribulokinase is able to phosphorylate both the D- and L-isomers of ribulose. Consequently, the need for AraB to be co-expressed with Z0415-19 would create a shared dependency on AraC and be requisite to avoiding redundancy in transporter expression.

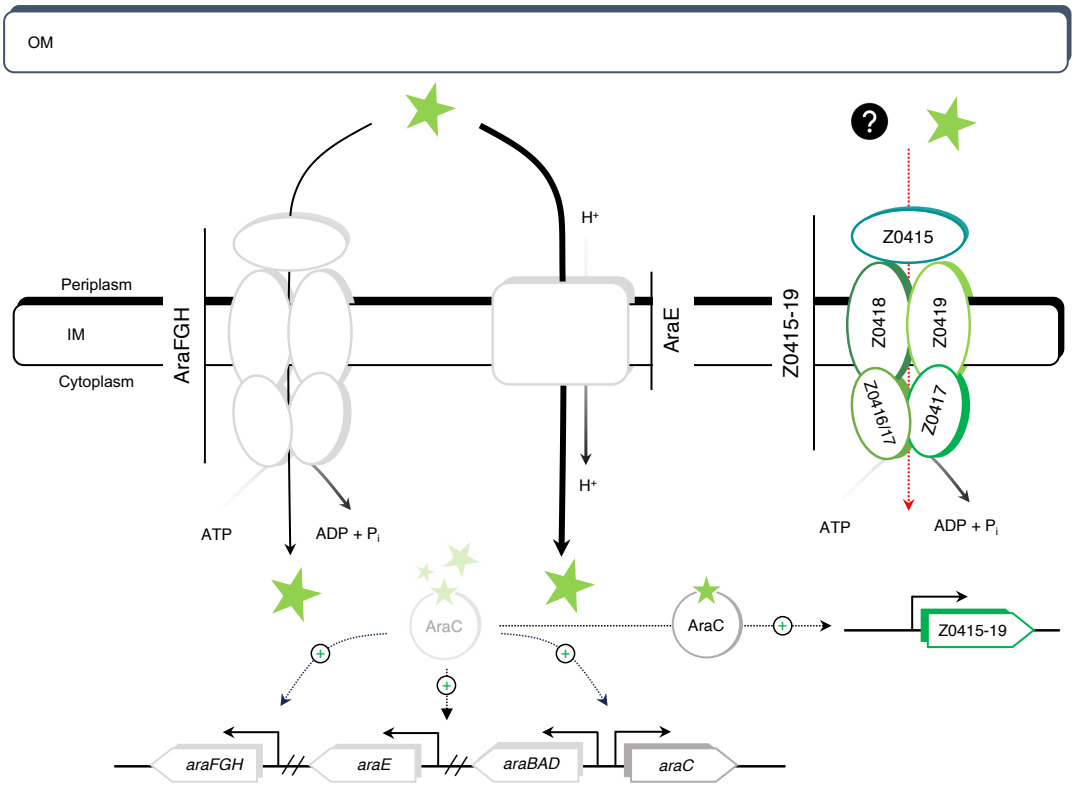


Figure 4-26. Regulatory overview of Z0415-19 expression by AraC in TUV93-0. L-arabinose (green star) is transported into the cytoplasm from the periplasm via the transporters of the canonical L-arabinose system. Transport of L-arabinose shows greater dependency upon AraE than AraFGH. In the cytoplasm, AraC binds L-arabinose activating expression of the L-arabinose system as well as the expression of Z0415-19. The transporter is predicted to encode an ABC transporter specific for the transporter of an unknown substrate (?) that is potentially L-arabinose. Arrow thickness represents substrate uptake.

Whilst investigating the contribution of transporters required to facilitate EHEC growth on L-arabinose, it is important to acknowledge that the conditions tested (i.e., aerobic growth on minimal media) are exceedingly simplistic and do not reflect the highly competitive environment of the gut. Due to the diverse bacterial consortia comprising the native gut microbiota of the host, each with its own unique nutritional requirements, EHEC must be able to outcompete these bacteria for nutrients. Therefore, whilst EHEC is able to grow on L-arabinose under isolated conditions, it must also do so when co-

grown with members of the microbiota, if it is to use the sugar for growth in the gut environment. To address this, EHEC was grown in co-culture with *Bt*.

One of the complexities that arises, is that members of the gut microbiota typically also encode equivalent systems specialised for the use of the same nutrient. As with EHEC, *Bt* encodes a system that is specific for the utilisation of L-arabinose. The system (BT0365-BT0350) encodes an IM transporter and associated metabolic enzymes for the import and metabolism of L-arabinose respectively (Schwalm et al., 2016). Despite this, EHEC was found to significantly outcompete *Bt* for L-arabinose, so much so that *Bt* grew worse in coculture than it did in monoculture. It is likely that the multiple systems dedicated to L-arabinose transport provides EHEC with the ability to use the sugar so rapidly that *Bt* is left with little for its own growth. This also provides a possible explanation as to why EHEC may encode an additional transporter. The impaired growth of EHEC through deletion of *araE* was found to limit the competitive advantage previously observed in co-culture. Rather than declining, *Bt* growth slowly but gradually increased over 12 h. Under these conditions, AraFGH alone was not sufficient to maintain the phenotype observed for the WT when grown on L-arabinose in co-culture. AraE therefore appeared to be the primary driver of this phenotype, likely attributable to its rapid induction and high capacity to transport L-arabinose. These data were supported by previous work that also details AraE to be the main route of L-arabinose uptake under anaerobiosis (Hasona et al., 2004).

Further to this, since *Bt* is a prolific degrader of complex polysaccharides (Martens et al., 2008; Sonnenburg et al., 2005), the affinity and overall kinetics of the L-arabinose utilisation system therefore likely differs to that of EHEC which can only utilise simple monosaccharides (Fabich et al., 2008). For example, the transcriptional activation and subsequent expression of the L-arabinose system in EHEC may be much more rapid than that of the *Bt* system. If so, this would explain the gradual dropping off of *Bt* counts observed for in coculture but not in monoculture. It has been previously shown that *araE* mRNA reaches 50 % of its maximum level within 1 minute of exposure to the sugar (Johnson & Schleif, 1995). In fact, the induction of the entire L-arabinose system of *E. coli* is extremely rapid, occurring within 15 to 30 seconds of exposure (Johnson & Schleif, 1995). Therefore, if EHEC is able to use the majority of the sugar even before the *Bt* system is fully expressed, then little of the sugar would remain for *Bt* to use.

Subsequently, the growth of *Bt* might have been sustained if L-arabinose was periodically re-supplemented into the co-culture across the course of growth.

Previous work has highlighted the dependency EHEC places upon *Bt* to provide an important source of nutrients and metabolites (some of which feed into the regulation of virulence) (Curtis et al., 2014; Pacheco et al., 2012) and may help to explain the source of L-arabinose in the gut. This is due to EHEC being completely incapacitated in its ability to use complex polysaccharides as a carbon source, arising from a lack of appropriately specialised enzymes (Conway & Cohen, 2015). Through its saccharolytic abilities *Bt* degrades complex polysaccharides, releasing monosaccharides that *E. coli* can then use (Conway & Cohen, 2015). Since the source of L-arabinose in the gut is likely to be sourced from the diet in the form of plant fibre, where the sugar is a major component of polysaccharides (arabinans) and glycoproteins (Crozier et al., 2021), EHEC would be unable to access the free sugar. Therefore, it could be hypothesised that gut microbiota members, such as *Bt*, may be able to release L-arabinose from arabinans and other complex structures. Mechanisms of cooperative sharing of resources have been described amongst *Bacteroides* (Feng et al., 2018) and contrast the sometimes-selfish mechanisms they use (Cuskin et al., 2015). However, the current model of *Bt* utilisation of arabinan (Schwalm et al., 2016) is reflective of the selfish mechanism employed for the utilisation of yeast mannan (Cuskin et al., 2015), whereby the polysaccharide is captured, transported, and degraded intracellularly. Therefore, the monomeric L-arabinose would not be liberated and made freely available in the gut environment as “public goods”. Although, small amounts of L-arabinose or the enzymes required for degradation may be present in the gut environment as a result of cell lysis over time.

Besides the possibility for microbiota-derived enzymes to liberate L-arabinose from complex plant biopolymers, there is the suggestion that these polysaccharides can be partially degraded during their passage through the gut (Zhang et al., 2003). This is based on findings that revealed up 10 % of L-arabinose to be liberated from hemicelluloses upon contact with acidities similar to those of the stomach (Zhang et al., 2003). Although less likely, there is the possibility that EHEC may encode currently unknown enzymes capable of cleaving L-arabinose residues. This has recently been seen for *Salmonella* whereby an arabinofuranosidase able to liberate L-arabinose from

dietary polysaccharides was expressed during colonisation of mice *in vivo* (Ruddle et al., 2023). The exact source and concentration of free L-arabinose in the human gut therefore remains to be determined.

4.15. Conclusions

The findings presented in this chapter demonstrate L-arabinose as a sugar that can support EHEC growth and confer a competitive advantage when co-cultured with *Bt*. This advantage is conferred by AraE and likely allows the rapid scavenging of L-arabinose in the human gut, important for colonisation and establishing a niche. Moreover, besides its canonical system, L-arabinose was also found to regulate the expression of a novel ABC transporter located on an OI in EHEC. Regulation of the locus was shown to be dependent on the transcriptional regulator, AraC. Although the exact function of the system remains unknown, the co-regulated expression with the L-arabinose utilisation system suggests that it is important during growth on the sugar. However, the exact conditions under which this system may play a phenotypic role are still unknown. This work therefore provides new insight into the molecular mechanisms underpinning the utilisation of L-arabinose and its role in EHEC.

5. Investigating the effects of L-arabinose on EDL933 virulence

5.1. Introduction

Regulation of the LEE, and therefore the virulence of EHEC, is greatly underpinned by the biochemical composition of the GI tract. Derived from the diet, host, and microbiota, nutrients and other metabolites act as important environmental stimuli for pathogens in the regulation of their virulence (Yang et al., 2023), particularly as LEE expression does not rely on a specific tissue-receptor tropism. The differential expression of genes and/or regulons specific for a particular nutrient or metabolite under *in vitro* T3SS inducing conditions such as growth in MEM-HEPES or *in vivo* infection is typically suggestive that they are important for pathogenesis (Connolly et al., 2014, 2018). Hence, these genes are normally co-expressed with the LEE and other important virulence-related traits (Connolly et al., 2018).

Through the phenomenon of nutrient sensing, EHEC determine the presence or absence of specific nutrients prior to integrating the chemical signals into complex regulatory circuits that are controlled by a plethora of transcriptional regulators (Jimenez et al., 2019; Kendall et al., 2012; Menezes-Garcia et al., 2020; Pacheco et al., 2012; Pifer et al., 2018). The terminal regulator often determines the outcome of expression, whether that be to enhance or repress virulence. Recent work has highlighted that accessory nutrient uptake systems specific to OIs can contribute to virulence (Yang et al., 2023). This enhancement likely occurs via these accessory systems by (a) providing a competitive advantage (Discussed previously) and (b) improving the scavenging of nutrients required for regulating virulence expression.

In the previous chapter, the ability of L-arabinose to induce the expression of a novel ABC transporter located on OI-17 was demonstrated. An earlier study which aimed to identify *in vivo* induced proteins during human infection highlighted the Z0419 permease component of Z0415-19 to be amongst those expressed (John et al., 2005). This finding coincided with the expression of established virulence factors in EHEC such as intimin, highlighting a potential role with L-arabinose via Z0415-19 in aiding pathogen adaptation and survival within the gut. Furthermore, both the aforementioned study and a more recent investigation demonstrated the significant upregulation of key AUGs (*araG*, *araC*, *araB*) during *in vivo* infection (Gardette et al., 2019; John et al.,

2005). Subsequently, there is substantial evidence to support the hypothesis that L-arabinose is a sugar important for not only colonisation but also the regulation of virulence in EHEC during human infection.

The work that follows in this chapter therefore aims to address the contribution of L-arabinose as a dietary-sourced signal in the regulation of virulence in EHEC, with specific emphasis towards LEE expression. Using a combination of bacterial genetics, transcriptomics, and infection models (both *in vitro* and *in vivo*) the downstream drivers of identified phenotypes were explored. The specificity of regulatory changes in response to L-arabinose were then expanded and tested to see if they were applicable to additional aldopentose sugars, D-ribose, and D-xylose.

5.2. Carriage of the LEE is strongly associated with Z0415-19

The potential connection between L-arabinose and the regulation of the LEE was investigated by examining the co-occurrence of the L-arabinose-specific locus Z0415-19 and the LEE (**Figure 5-1A**). Similar to the LEE, Z0415-19 was located on an OI and strongly associated with EHEC strains. Looking amongst the same strains used in the previous phylogenomic analyses, LEE carriage was identified to be largely restricted to phylogroup E (**Figure 5-1A**), with a total of 216 strains (97.7 %) identified as LEE+. None of the LEE- strains of phylogroup E were able to be assigned a pathotype apart from a single ETEC strain (str. 2845650). Carriage of the LEE was also observed across phylogroups B1 (n = 10), B2 (n = 5) and A (n = 6). Of those able to be assigned a pathotype, these were again EHEC/STEC with the addition of several EPEC and a single ETEC strain (str. TW07509) in phylogroup B1. The most frequently observed loci combination was LEE-/ Z0415-19-, occurring in 41.9 % of strains tested, whilst LEE+/ Z0415-19- was the least frequently observed, occurring in only 0.01 % of strains (**Figure 5-1B**). There was a strong positive correspondence between carriage of the LEE and Z0415-19 (odds ratio = 35.9; $P < 0.001$). Thus, LEE carriage was hardly ever observed without carriage of Z0415-19. Overall carriage of both loci was observed for phylogroups E, A, B1 and B2 (**Figure 5-1C**). Most of the strains able to be assigned a pathotype were EHEC/STEC with a small number of EPEC across phylogroups B1 (n = 1) and B2 (n = 2).

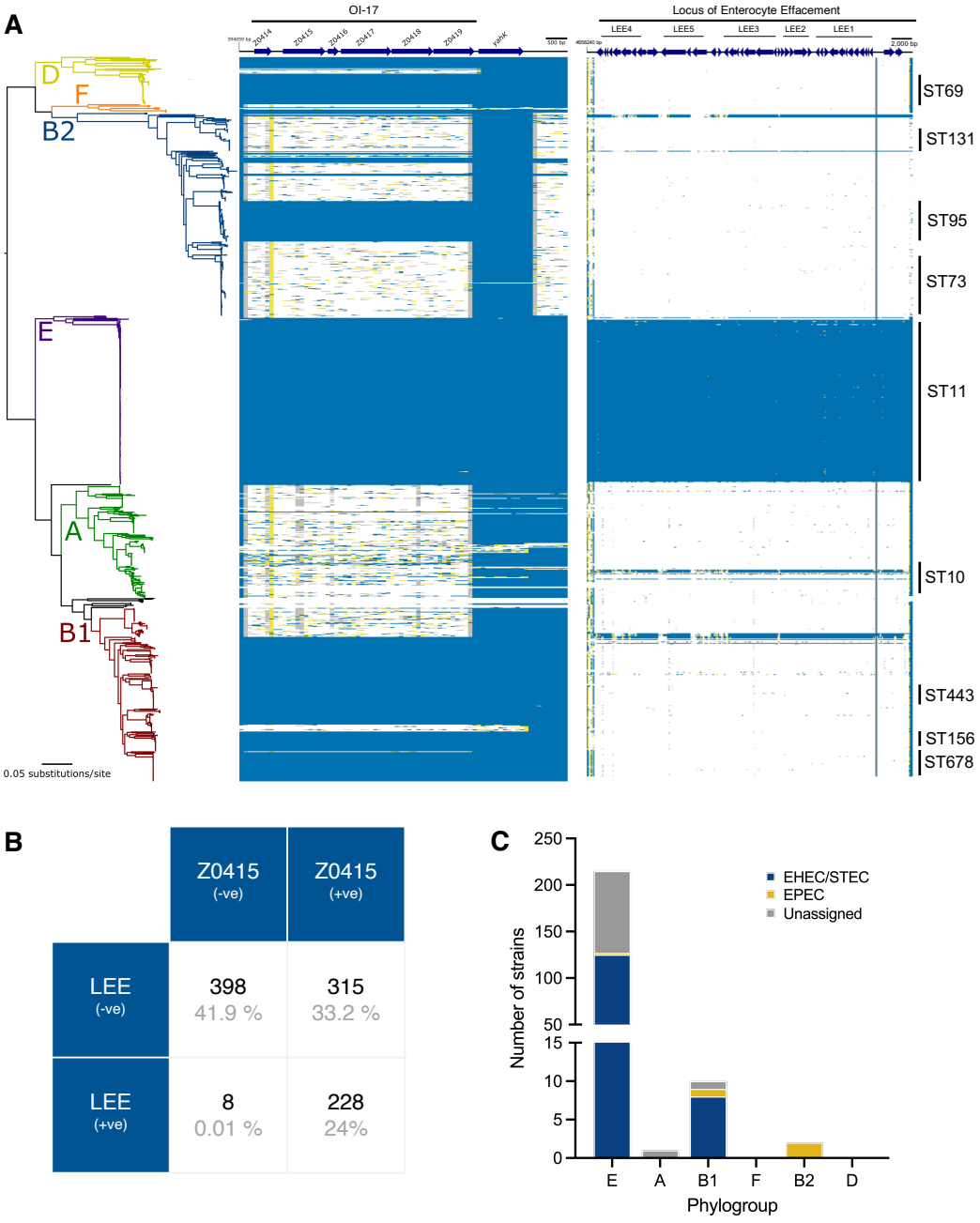


Figure 5-1. Association of Z0415-19 with LEE carriage across *E. coli*. (**A**) Maximum likelihood analysis built from 245,518 core-genome SNPs relative to the reference chromosome of EDL933. Phylogeny is rooted according to the actual root by *E. fergusonii* ATCC 35469 (Omitted for visualisation). Branch lengths and scale bar represent number of nucleotide substitutions per site. Coverage is shown as a heat map where ≥ 80 % identity is highlighted in blue, ≥ 50 % identity is highlighted in yellow, and ≥ 1 % is highlighted in grey. White plots indicate regions that are absent. Both Z0415-19 and LEE loci are indicated above the phylogeny. (**B**) Contingency matrix of LEE and Z0415-19 carriage association across strains used in this analysis. (**C**) Pathotype breakdown of strains positive for carriage of both loci across phylogroups.

5.3. Investigating the regulatory nature of L-arabinose on the LEE

Following the association identified between Z0415-19 and the LEE, it was hypothesised that L-arabinose uptake may act as a stimulatory signal of the LEE. A 400 bp stretch of sequence immediately upstream of *ler* (LEE1) was cloned into pMK1*lux* to enable the dynamics of LEE1 promoter activity to be assessed. Notably, this region included both the P1 and P2 promoters (**Figure 5-2A**). The corresponding pMK1*lux*-P_{LEE1} construct was transformed into WT EHEC and grown in MEM-HEPES (LEE-inducing conditions) in the presence or absence of L-arabinose (0, 0.5, 5 mg/mL). P_{LEE1} expression was found to be significantly enhanced when grown in the presence of L-arabinose, with activity increasing with the concentration of the sugar (**Figure 5-2B**). The enhancing effects of L-arabinose on P_{LEE1} expression was apparent post 5 h of growth, despite no significant differences in bacterial density across the different concentrations tested (**Figure 5-2CD**). P_{LEE1} expression was found to be sustained for several hours longer in the presence of L-arabinose (**Figure 5-2BD**). This enhanced P_{LEE1} expression was also still observed at concentrations as low as 0.05 mg/mL (**Figure 5-2E**).

Given that the reporter assays acted as a proxy for LEE expression and not a direct measure, the effect of L-arabinose on the transcription of all five LEE operons (LEE1-5) was validated directly by RT-qPCR. Although the carriage of *ler* by LEE1 makes it arguably the most influential operon of the LEE, the entire locus spans an additional four operons. RNA was extracted from TUV93-0 grown in MEM-HEPES supplemented with and without L-arabinose to late exponential (OD_{600} 0.9), and RT-qPCR was subsequently used to quantify the expression of genes across LEE1-5 (**Figure 5-3**). The expression of each gene was found to be significantly upregulated in the presence of L-arabinose when using the housekeeping gene, *mopA*, as an internal control (**Figure 5-3A**). The highest and lowest fold change in expression was observed for *espA* (~6-fold; LEE4) and *escV* (~3-fold; LEE3), respectively (**Figure 5-3**). As validation, all of the LEE genes tested except *escV* (LEE3) were also significantly upregulated when using a second housekeeping gene (*gapA*) (**Figure 5-3B**). The fold changes in expression were found to be relatively similar across all significantly upregulated genes (between 2 to 3-fold) (**Figure 5-3B**). These data therefore provide substantial evidence to show that L-arabinose is able to significantly enhance LEE expression in EHEC.

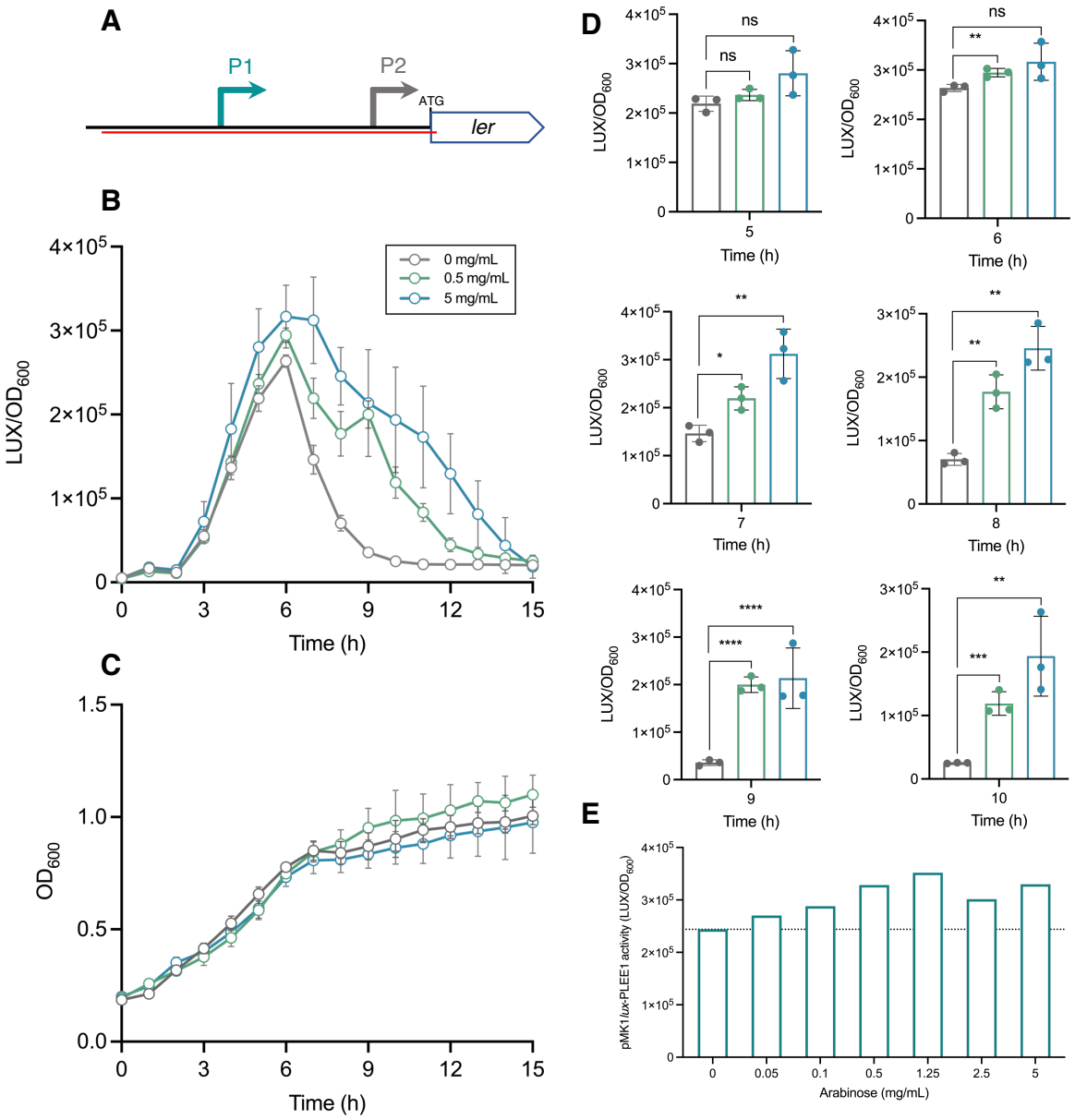


Figure 5-2. L-arabinose enhances P_{LEE1} activity in TUV93-0. (**A**) *ler* promoter region (red) used in this study. (**B**) Transcriptional reporter assay using the pMK1*lux*-P_{LEE1} fusion construct. WT TUV93-0 harbouring pMK1*lux*-P_{LEE1} was grown in MEM-HEPES supplemented with and without L-arabinose. Promoter activity was measured in RLU by normalising the raw luminescence against the OD_{600}. (**C**) Growth curves of WT TUV93-0 with pMK1*lux*-P_{LEE1} when grown in the presence and absence of L-arabinose. (**D**) Breakdown of mean pMK1*lux*-P_{LEE1} activity post 5 h of growth. (**E**) TUV93-0 pMK1*lux*-P_{LEE1} activity across a range of L-arabinose concentrations at mid-exponential growth. (*) $P < 0.05$, (**) $P < 0.01$, (****) $P < 0.001$ - calculated using a student's *t*-test. $n = 3$ biological replicates and error bars represent SD.

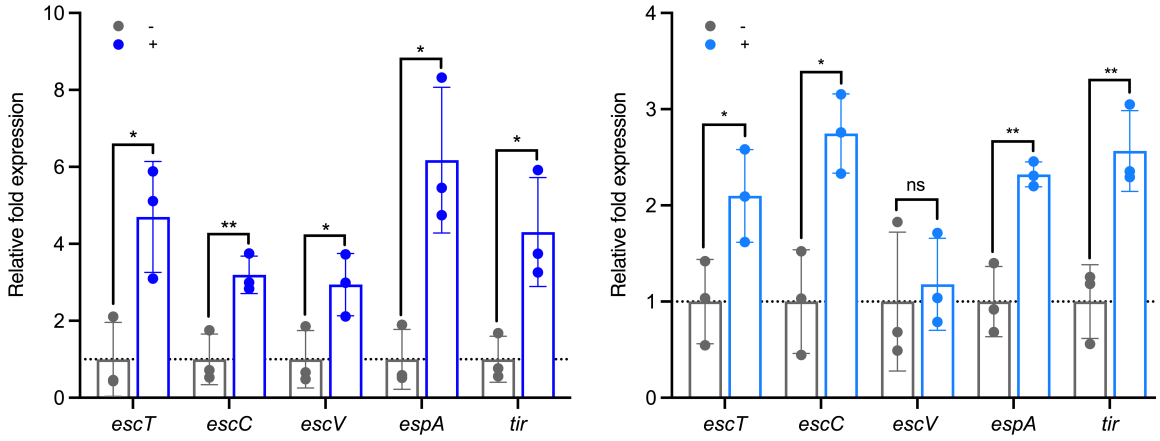


Figure 5-3. RT-qPCR of LEE gene expression in TUV93-0 after growth on L-arabinose. WT TUV93-0 was grown in MEM-HEPES supplemented with (+) and without (-) 5 mg/mL L-arabinose. Expression of LEE genes in the presence of L-arabinose was determined relative to in the absence of the sugar by RT-qPCR using the housekeeping genes (**A**) *mopA* (GroEL) and (**B**) *gapA*. The dashed line is indicative of the baseline expression in the WT grown without L-arabinose. (*) $P < 0.05$ and (**) $P < 0.01$ - calculated using a student's *t*-test. $n = 3$ biological replicates and error bars represent SD.

5.4. T3SS effector secretion is enhanced by L-arabinose

Enhanced transcription of the LEE in the presence of L-arabinose suggested increased T3SS assembly and secretion of effectors. To test this, secreted proteins were precipitated from the supernatant of TUV93-0 cultures grown in the presence and absence of L-arabinose and separated by SDS-PAGE. In the presence of L-arabinose, the known LEE-encoded effectors, EspD, EspA, EspC and Tir, were shown to be secreted to higher amounts (**Figure 5-4AB**). This enhanced secretion was confirmed by Western blotting for EspD in both secreted and whole cell fractions (**Figure 5-4C**). The housekeeping protein, GroEL, was also tested, using the same relative volume of whole cell lysate as per volume of supernatant to confirm that any observed differences were not the outcome of inconsistent loading (**Figure 5-4C**). Therefore, L-arabinose increased both the expression of the LEE and the secretion of T3SS effectors.

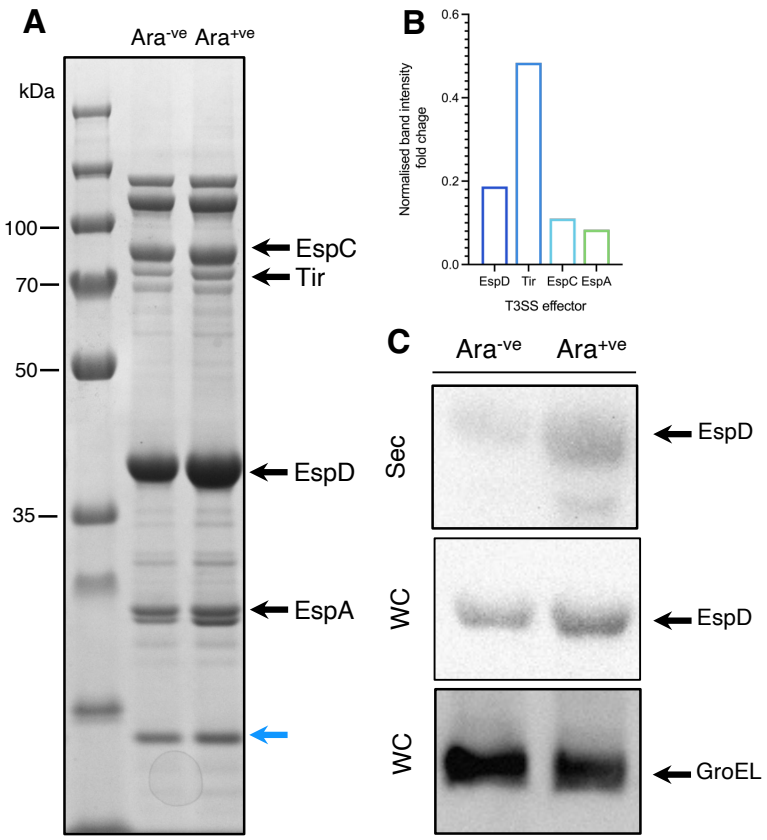


Figure 5-4. L-arabinose enhances T3SS activity in EDL933. (**A**) SDS-PAGE of the EDL933 secreted protein profile when grown in MEM-HEPEs supplemented with and without 5 mg/mL L-arabinose. Bands corresponding to the T3SS effectors are indicated by the arrows. The blue arrow represents an internal loading control (lysozyme; 1 μg/mL) (**B**) Band intensity fold change for the T3SS effectors shown in panel A, between control and treatment conditions. Fold changes were normalised using the band intensity of an internal loading control. (**C**) Western blot for EspD and GroEL in secreted (Sec) and whole cell lysate (WC) fractions respectively.

5.5. L-arabinose enhances adherence of EHEC to host cells

To test whether the L-arabinose enhanced T3SS expression increased host-cell interaction, the effect of the sugar on the adherence of EHEC to cultured HeLa cells was assessed (**Figure 5-5**). The adherence to host cells is a critical step in EHEC infection, necessary for T3SS activity and formation of A/E lesions. When media was supplemented with L-arabinose the number of EHEC adhered per infected cell was significantly higher (Median = 10) than in the absence of the sugar (Median = 6) ($P <$ 0.0001) (**Figure 5-6A**). The mean percentage of infected HeLa cells between conditions was found not to be significantly different ($P = 0.16$), although there was a

consistent increase within replicates in the presence of L-arabinose (**Figure 5-6B**). In agreement with the previous results for LEE expression and T3SS secretion, these data indicate L-arabinose enhances the ability of EHEC to adhere cells during *in vitro* infection.

A preliminary investigation to see if EHEC could be co-cultured with intestinal derived organoids was also conducted. In the presence of L-arabinose, higher CFU/mL counts were obtained from transwells which was indicative of greater adherence to the epithelium than without (**Figure 5-6C**). This was despite CFU/mL counts being lower in the apical media (containing planktonic cells). The TEER of the intestinal organoid monolayers was measured before and after coculture with EHEC ± L-arabinose. In cells exposed to EHEC with L-arabinose, ΔTEER was higher (i.e., more negative) than in those exposed to EHEC with no sugar (**Figure 5-6D**). A lower TEER suggested that when grown with L-arabinose, EHEC negatively affects the permeability and integrity of monolayers. Statistical analyses could however not be carried out on these data due to a lack of replicates.

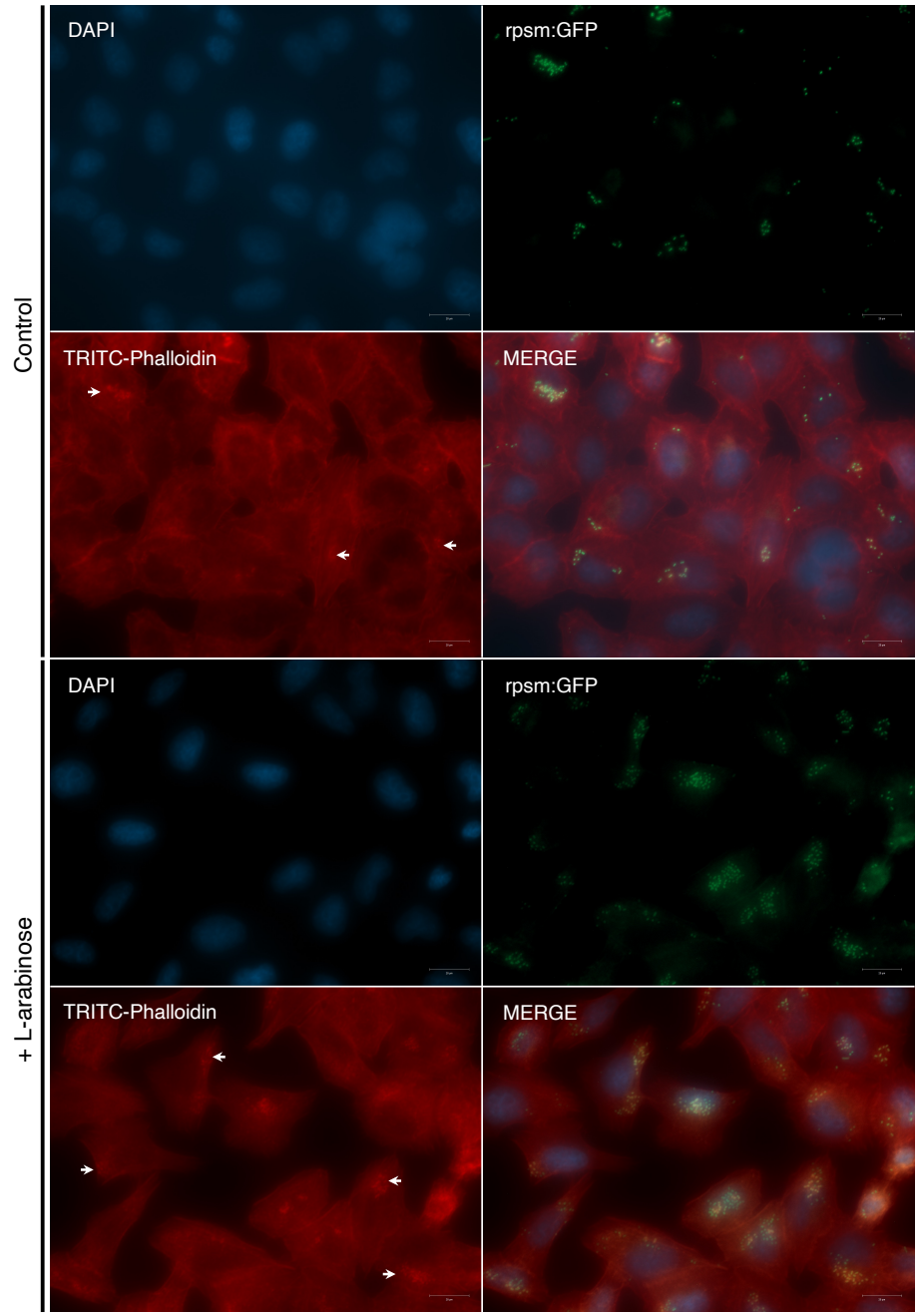


Figure 5-5. *In vitro* cell adhesion assay. Representative immunofluorescence images of HeLa cells infected with TUV93-0 grown in the absence or presence of 5 mg/mL L-arabinose. Images were captured by wide-field fluorescence microscopy. Host cell actin was stained with TRITC-Phalloidin (555 nm) and EHEC carried a plasmid constitutively expressing GFP (rpsm:GFP) (488 nm). Host cell nuclei were also stained with DAPI stain (360 nm). White arrows indicate regions of condensed actin characteristic of A/E lesions. Panels show images taken from individual microscope channels and when merged.

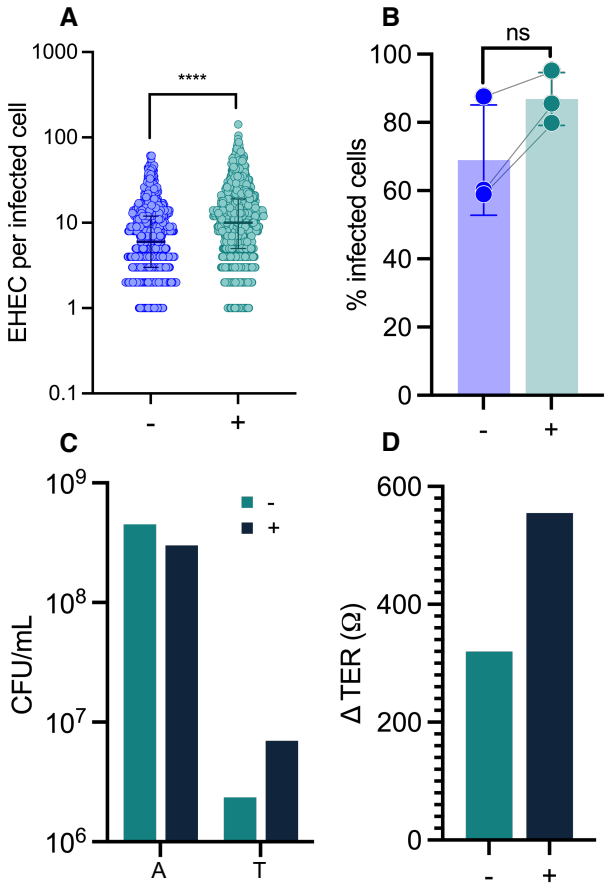


Figure 5-6. L-arabinose enhances adherence of EHEC to host cells. Comparison between (**A**) the number of EHEC per infected HeLa cell and (**B**) the proportion of HeLa cells infected by EHEC, in the presence and absence of L-arabinose. EHEC was co-cultured with HeLa cells for 5 h in MEM-HEPES ± L-arabinose at 37 °C (5 % CO_2). *n* = 3 biological replicates and error bars represent SD. (****) $P < 0.001$ - calculated using a Mann-Whitney U test. (**C**) CFU/mL EHEC recovered from the apical (A) and transwell (T) compartments after co-culture with intestinal organoids derived from human ileum in the presence and absence of L-arabinose. (**D**) Difference in TEER of intestinal organoids after co-culture with EHEC in the absence and presence of L-arabinose.

5.6. Deletion of *araC* abolishes P_{LEE1} activity in EHEC

Many cellular transcription factors directly affect regulation of the LEE in response to their cognate signals. Following determination that L-arabinose significantly enhances LEE expression in EHEC, the role of the cognate transcriptional regulator, AraC, was probed to test if L-arabinose mediated regulation directly overlapped with LEE control. Deletion of *araC* in EHEC resulted in the enhanced LEE1 phenotype no longer being observed, despite being grown in L-arabinose containing media (**Figure 5-7**). This suggested that the regulator may be an important contributor in the mechanism of LEE regulation via L-arabinose. However, it was unclear whether this was due to a direct regulatory effect of AraC or impaired L-arabinose uptake or metabolism via the corresponding AraC-regulated AUGs.

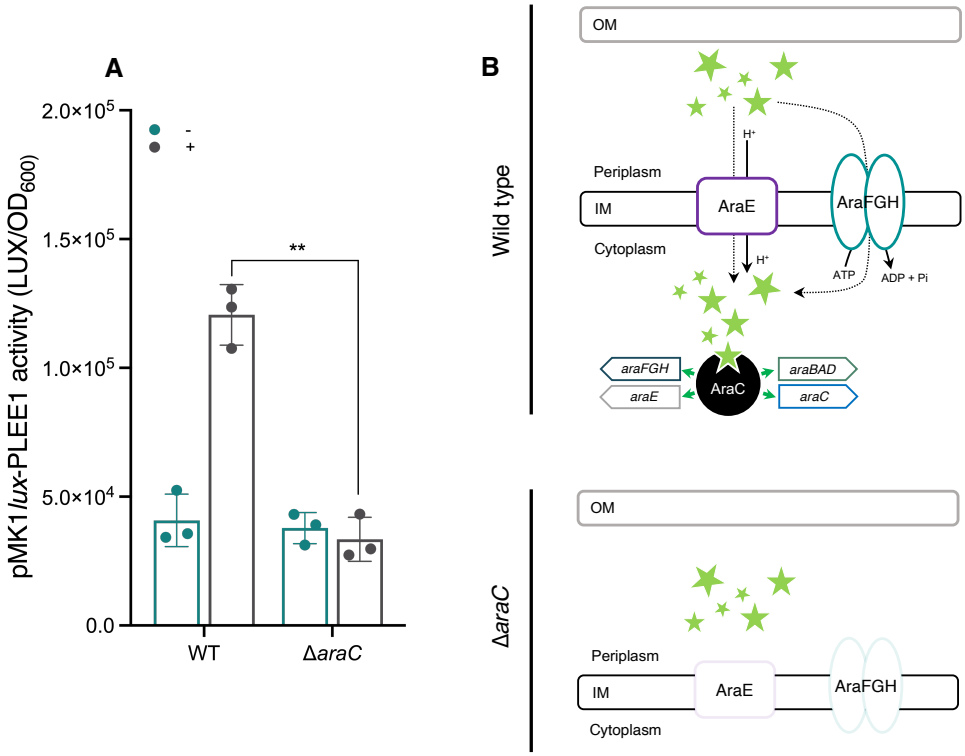


Figure 5-7. Deletion of *araC* prevents enhanced P_{LEE1} activity in TUV93-0. (**A**) WT and Δ*araC* with pMK1*lux*-P_{LEE1} were grown in MEM-HEPES supplemented with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. *n* = 3 biological replicates and error bars represent SD. (**) $P < 0.01$ - calculated using a student's *t*-test. (**B**) Schematic of L-arabinose utilisation in the WT and Δ*araC* mutant strain.

5.7. Deletion of key AUGs impairs P_{LEE1} activity

To next investigate whether the transport and metabolism of L-arabinose contributed to enhanced LEE expression, pMK1*lux*-P_{LEE1} was also introduced into mutants of the system (Δ*araBAD*, Δ*araE* and Δ*araFGH*). Whilst P_{LEE1} expression in the Δ*araFGH* mutant was found not to significantly differ to the WT when grown in L-arabinose containing media, deletion of *araBAD* and *araE* led to a significant reduction in LEE expression compared to the WT when grown with L-arabinose (**Figure 5-8**). However, unlike Δ*araBAD*, a significant increase in LEE expression was still observed for Δ*araE* when grown on L-arabinose, likely explained by the presence of multiple transport systems in EHEC.

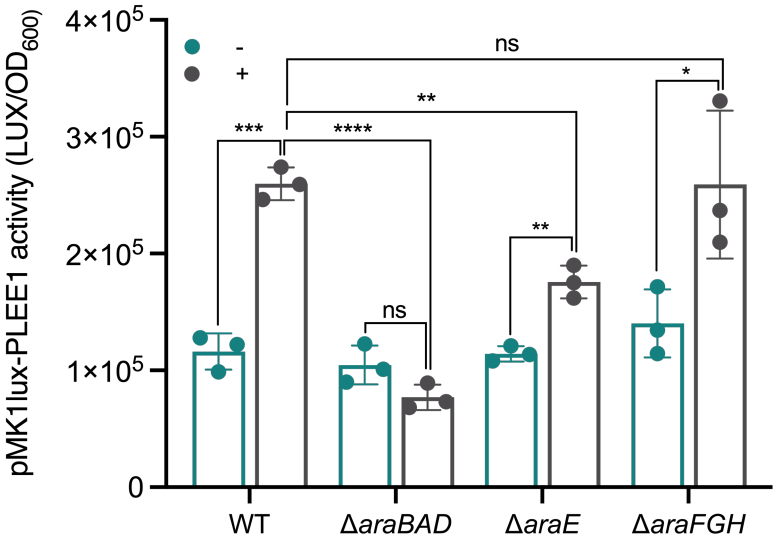


Figure 5-8. Deletion of *araBAD* and *araE* also prevent enhanced P_{LEE1} activity in TUV93-0. WT, Δ*araBAD*, Δ*araE* and Δ*araFGH* mutant strains with pMK1*lux*-P_{LEE1} were grown in MEM-HEPES supplemented with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. n = 3 biological replicates and error bars represent SD. (*) $P < 0.05$ (**) $P < 0.01$; (****) $P < 0.0001$ - calculated using a student's t-test.

5.8. Complementation of Δ*araC* restores enhanced P_{LEE1} activity

As AraC is responsible for regulating the expression of the canonical L-arabinose system, its absence would inhibit expression of both transport (AraE) and metabolism (AraBAD) components. It was therefore questioned whether the reduced P_{LEE1} expression observed in Δ*araC* was solely attributable to absence of the regulator, or the silenced expression of other AUGs. Using plasmid-based complementation,

components of the system were sequentially re-introduced into an Δ*araC* mutant background. Constitutive expression of *araC* in the Δ*araC* mutant was found to restore the WT phenotype, with P_{LEE1} expression once again being enhanced in the presence of L-arabinose (**Figure 5-9**). In contrast, re-introduction of *araE* did not allow for enhanced P_{LEE1} expression to be restored (**Figure 5-9**). To ensure that an absence of phenotype in this strain was not due to issues with the *araE* construct, WT, Δ*araE* and complemented Δ*araE* strains were grown in minimal media supplemented with L-arabinose (**Figure 5-10**). As seen previously, Δ*araE* displayed an extensive lag phase and was unable to grow at most concentrations (**Figure 5-10AB**). However, the mutant when complemented with pSU-*araE* was once again able to grow across all conditions (**Figure 10AC**) and had a similar final OD_{600} reading as the WT (**Figure 5-10D**). Therefore, transport via AraE, the main transporter of L-arabinose, was concluded as not being responsible for driving the enhanced P_{LEE1} expression phenotype observed.

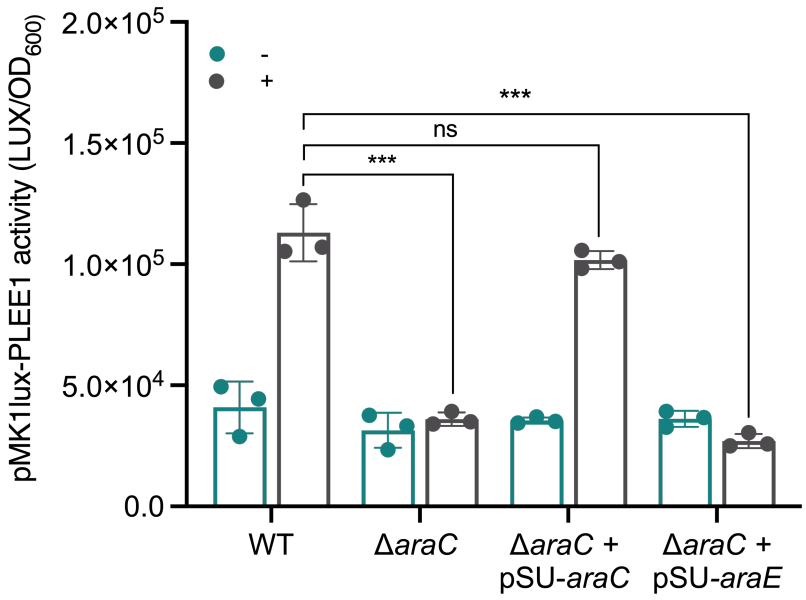
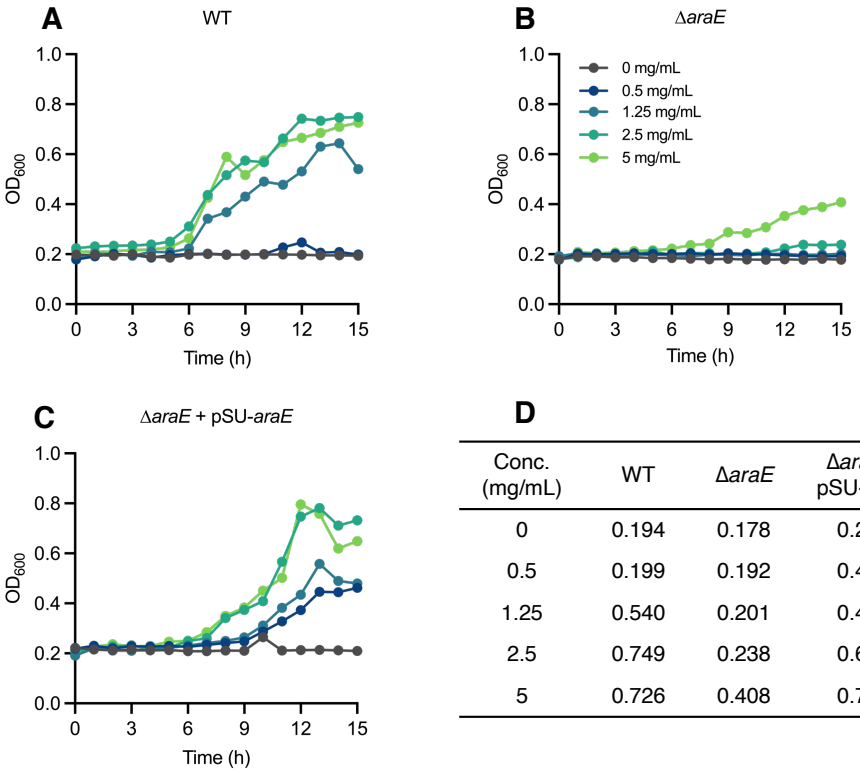


Figure 5-9. Transport of L-arabinose is not sufficient to drive enhanced P_{LEE1} expression in TUV93-0. WT, Δ*araC*, Δ*araC* with pSU-*araC* and Δ*araC* with pSU-*araE* TUV93-0 were grown in MEM-HEPES supplemented with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. n = 3 biological replicates and error bars represent SD. (***) $P <$ 0.001 - calculated using a student's *t*-test.



D

Conc. (mg/mL)	WT	Δ*araE*	Δ*araE* + pSU-*araE*
0	0.194	0.178	0.200
0.5	0.199	0.192	0.462
1.25	0.540	0.201	0.480
2.5	0.749	0.238	0.649
5	0.726	0.408	0.733

Figure 5-10. Complementation of *araE* restores the ability of TUV93-0 to grow on L-arabinose. Growth profiles of (**A**) WT, (**B**) Δ*araE* and (**C**) Δ*araE* with pSU-*araE* TUV93-0 when grown in M9 minimal media supplemented with L-arabinose for 15 h. (**D**) Final OD_{600} readings across the WT, mutant, and complemented strains across the various concentrations tested for growth.

5.9. L-arabinose metabolism drives enhanced P_{LEE1} expression

The fact that L-arabinose transport was not responsible for enhanced LEE expression suggested that the sugar is not simply sensed as a regulatory signal itself. To test the hypothesis that L-arabinose metabolism was responsible for the enhanced P_{LEE1} expression, *araBAD* was re-introduced into the Δ*araC* background. As L-arabinose metabolism requires uptake of the sugar into the cell, it was necessary to simultaneously re-introduce *araE*. This would allow for transport of L-arabinose into the cell. Co-expression of both *araBAD* and *araE* in the Δ*araC* mutant via a constitutive promoter restored the enhanced LEE1 phenotype, with significantly higher P_{LEE1} expression observed in EHEC when grown in the presence of L-arabinose (**Figure 5-11**). This indicated that L-arabinose metabolism is essential for stimulating the enhanced LEE1 expression phenotype.

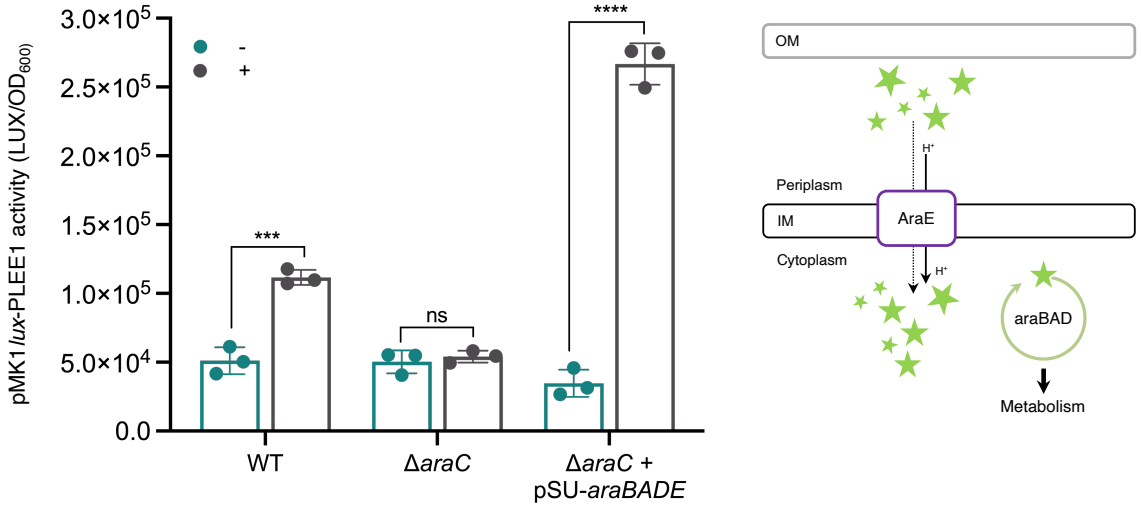


Figure 5-11. Expression of *araBAD* in the Δ*araC* background restores enhanced P_{LEE1} expression. WT, Δ*araC* and Δ*araC* with pSU-*araBADE* were grown in MEM-HEPES supplemented with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. $n = 3$ biological replicates and error bars represent SD. (***) $P < 0.001$; (****) $P < 0.0001$ - calculated using a student's *t*-test.

5.10. Primary L-arabinose metabolism does not drive P_{LEE1} expression

The metabolism of L-arabinose is mediated via the gene products of *araA*, *araB* and *araD*, encoding an isomerase (EC.5.3.1.4), L-ribulokinase (EC.2.7.1.16) and epimerase (EC.5.1.3.4), respectively (**Figure 5-12A**). Together these enzymes yield xylulose-5-phosphate for entry into the PPP. To determine whether any of these intermediate metabolites act as stimuli for driving LEE expression in EHEC, the necessary enzymes were re-introduced into an Δ*araBAD* mutant background (**Figure 5-12B**). *araA* was first re-introduced to allow for the generation of L-ribulose. However, despite being grown in the presence of L-arabinose, enhanced P_{LEE1} expression was not observed (**Figure 5-12C**). Similarly, enhanced P_{LEE1} expression was not restored in strains co-expressing *araA* and *araB*, grown on L-arabinose (**Figure 5-12C**). As such, neither L-ribulose nor L-ribulose-5-phosphate appear to be the primary driver of enhanced P_{LEE1} expression in EHEC.

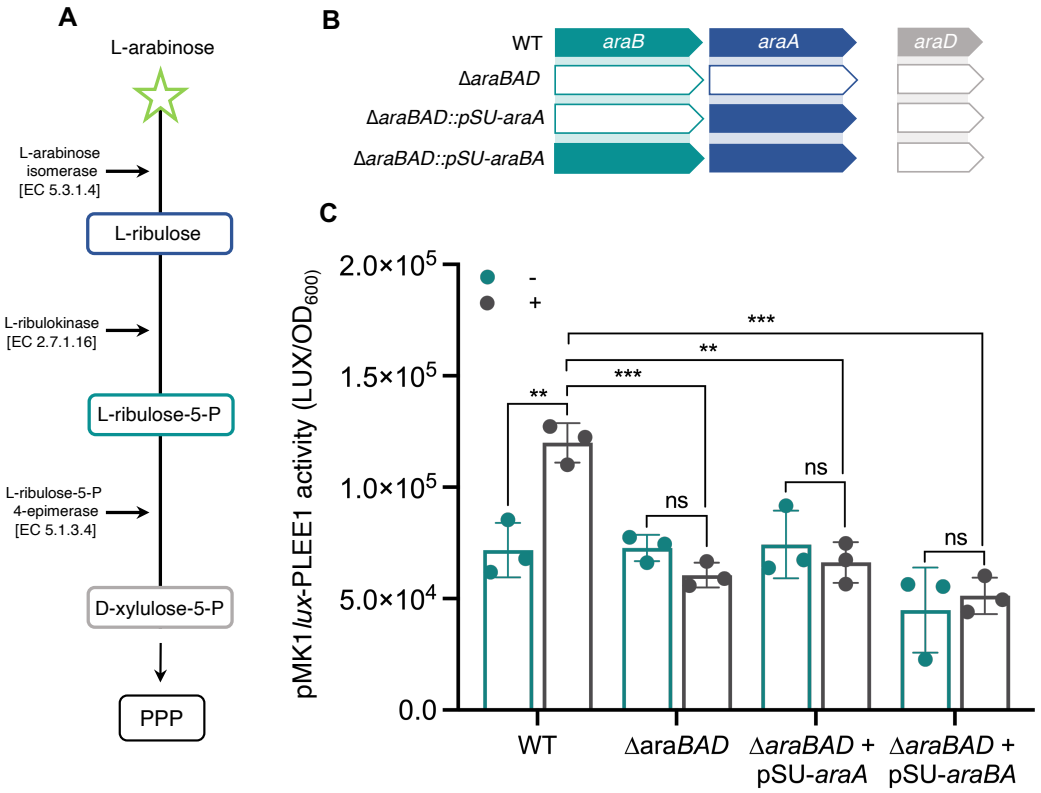


Figure 5-12. Metabolites of primary L-arabinose metabolism do not drive enhanced P_{LEE1} expression. (**A**) Overview of primary L-arabinose metabolism in *E. coli*. (**B**) Schematic of gene combinations across WT, Δ*araBAD*, Δ*araBAD* with pSU-*araA* and Δ*araBAD* with pSU-*araBA* TUV93-0 strains; (**C**) WT, Δ*araBAD*, Δ*araBAD* with pSU-*araA* and Δ*araBAD* with pSU-*araBA* were grown in MEM-HEPES with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. $n = 3$ biological replicates and error bars represent SD. (**) $P < 0.005$; (***) $P < 0.001$ - calculated using a student's *t*-test.

5.11. Enhanced P_{LEE1} activity coincides with P_{araB} activity

To further understand the context of L-arabinose metabolism in enhancing LEE expression, a transcriptional reporter fusion construct was made using the promoter of *araBAD* (P_{araB}), to determine the point at which L-arabinose metabolism is induced. Growth of EHEC harbouring pMK1*lux*-P_{araB} under the same conditions (MEM-HEPES + L-arabinose) for which enhanced P_{LEE1} expression had been previously observed, revealed the peak of P_{araB} activity to coincide with the peak of P_{LEE1} promoter activity (6h), with a strong positive correlation observed between the two ($R = 0.92$, $P < 0.001$) (**Figure 5-13A**). Activity from P_{araB} was found to be greatly influenced by sugar concentration, with promoter activation being more rapid at 5 mg/mL compared to 0.5 mg/mL (not shown). P_{araB} expression also appeared to be sustained following its onset and remained constant between hours 11 and 15. In the absence of L-arabinose, no activity was observed for pMK1*lux*-P_{araB} as expected.

As MEM-HEPES contains D-glucose, which is known to repress expression of AUGs via catabolite repression, utilisation of D-glucose by WT EHEC grown in MEM-HEPES was qualitatively assessed by TLC (**Figure 5-13BC**). D-glucose was largely depleted within 7 h of culture in MEM-HEPES (T_7) (**Figure 5-13B**). When supplemented with 5 mg/mL L-arabinose the profile of sugar utilisation differed and sugar was observed across all time points (**Figure 5-13C**). However, both L-arabinose and D-glucose standards ascended at similar rates on the silica plate. Therefore, the point at which EHEC switched to L-arabinose cannot be conclusively determined. However, based on no sugar being detected after 7 h of growth for the media only condition, and sugar being detected at all time points when supplemented with L-arabinose, it is likely that L-arabinose utilisation occurs post 7 h. These data agree with the P_{araB} reporter results and offer logical explanation as to why the enhanced LEE expression phenotype is observed at this phase of growth – i.e., when L-arabinose metabolism would be activated in glucose containing MEM-HEPES.

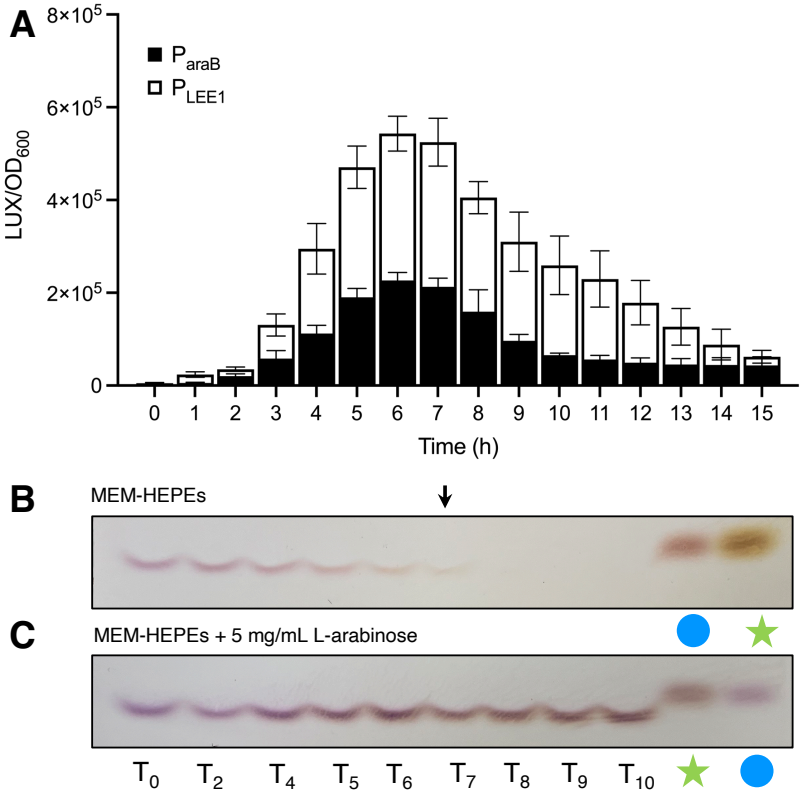


Figure 5-13. Enhanced P_{LEE1} expression coincides with the onset of P_{araB} expression. (**A**) P_{araB} activity expressed as RLU, overlayed with P_{LEE} activity in WT EHEC, when grown on MEM-HEPES supplemented with 5 mg/mL L-arabinose. $n = 3$ biological replicates and error bars represent SD. (**B**) TLC analysis of sugar utilisation (D-glucose) by WT EHEC when grown on MEM-HEPES (**C**) TLC of L-arabinose utilisation by WT EHEC when grown on MEM-HEPEs supplemented with 5 mg/mL L-arabinose. All growth for TLCs was done over 10 h. Symbols correspond to a 5 mM D-glucose (blue circle) and L-arabinose (green star) standard. The black arrow denotes the point of which sugar is depleted in the media only control.

5.12. Role of pyruvate in enhancing P_{LEE1} expression

Pyruvate has been previously shown to play a role in positively regulating LEE transcription (Carlson-Banning & Sperandio, 2016). In addition, pyruvate blocks the repressor (PdhR) of the pyruvate dehydrogenase complex responsible for allowing pyruvate to enter the TCA cycle (Anzai et al., 2020). Based on this, it was speculated that pyruvate, a far downstream metabolite of L-arabinose metabolism, could be responsible for the AraBAD dependent LEE phenotype. The LEE phenotype was therefore measured in the presence of L-arabinose and pyruvate. Accordingly, supplementation of MEM-HEPES with either pyruvate or L-arabinose enhanced P_{LEE1} expression (**Figure 5-14**). When supplemented together, P_{LEE1} expression was seen

to be enhanced even further. As the use of pyruvate relies on the activity of PdhR, its role as a regulator of LEE expression was tested. The Δ*pdhR* mutant with pMK1*lux*-P_{LEE1} was grown in MEM-HEPES supplemented in combination with L-arabinose and pyruvate. Despite its deletion, growth in the presence of both L-arabinose and pyruvate still enabled enhanced P_{LEE1} expression (**Figure 5-14**). Furthermore, enhanced P_{LEE1} expression for Δ*pdhR* was similarly still observed for when grown in the presence of L-arabinose and pyruvate independently, eliminating the regulator as a possible mediator (**Figure 5-14**). Important to note is that whilst these data were normalised against OD_{600}, strains grown in media supplemented with 0.2 % pyruvate did grow to a higher density than those without.

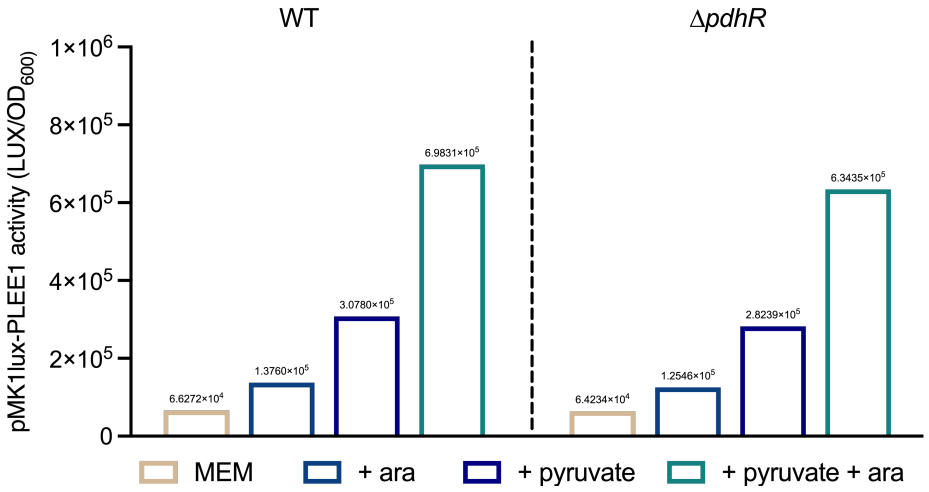


Figure 5-14. Enhanced LEE expression is unaffected in a Δ*pdhR* mutant. WT and Δ*pdhR* TUV93-0 was grown in MEM-HEPES supplemented either with or without 5 mg/mL L-arabinose and 0.2 % sodium pyruvate, as well as with both. RLU was calculated by normalising the raw luminescence against the OD_{600}.

5.13. L-arabinose drives significant changes to the EDL933 transcriptome

To determine a potential mechanistic basis of enhanced LEE expression by L-arabinose, transcriptomics was used to measure global changes in gene expression following growth on the sugar. Based on a positive/negative fold change (FC) of > 1.5 and FDR of $P \leq 0.05$ (see methods), a total of 1187 DEGs were identified in response to growth in MEM-HEPES with L-arabinose (**Appendix Table 8-1 & 8-2**). There was a strong correlation in transcriptome profiles between replicates within a condition (**Figure 5-15**). This was further confirmed by the distinct separation of replicates based

on condition by principal component analysis (not shown), demonstrating L-arabinose to have a significant effect on the overall transcriptome of TUV93-0.

Of the DEGs, 490 genes were significantly upregulated, whilst a greater number of 697 genes were significantly downregulated. As expected, genes for the transport (*araE*; *araFGH*) and metabolism (*araBAD*) of the sugar were amongst the most strongly upregulated, with a FC of > 42. Additionally, *araJ* and *ygeA*, known to be regulated by AraC, were also found to be significantly upregulated with a FC of 12.75 and 149.37, respectively. The expression of *araC* however, was significantly downregulated (FC = -1.84), suggesting EDL933 to be in the later stages of L-arabinose utilisation. Genes of Z0415-19 were also found to be significantly upregulated: Z0415 - FC = 26.51; $P < 0.001$; Z0417 - FC = 43.15; $P < 0.001$; Z0418 - FC = 7; $P < 0.001$). Z0419 was found not to be significantly upregulated (FC = 1.54; $P = 0.06$). An overview of DEGs following growth on L-arabinose is summarised by the volcano plot in **Figure 5-16**.

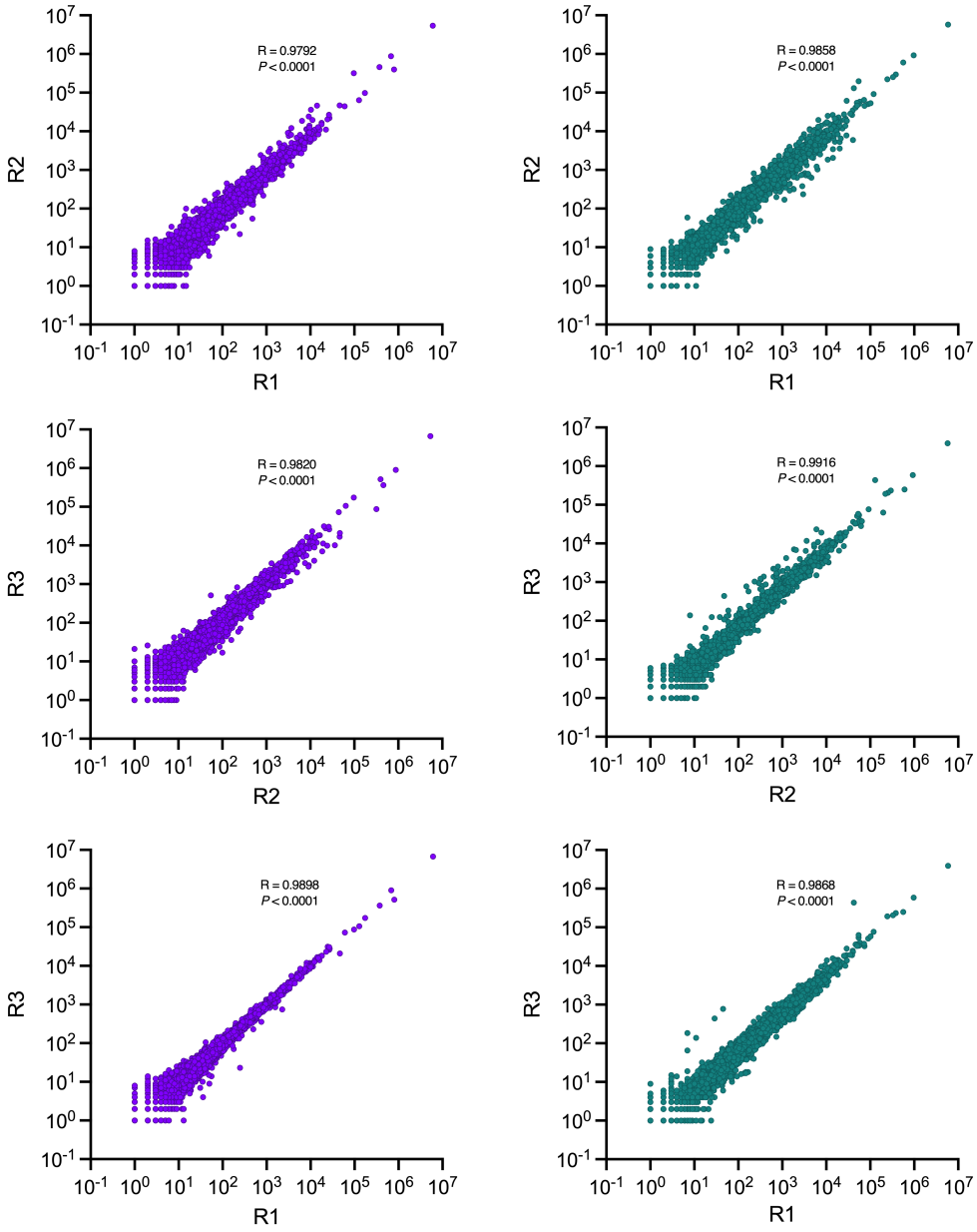


Figure 5-15. Spearman's rank analysis of TUV93-0 transcriptome profiles across replicates within conditions. Purple and green data points correspond to transcriptome profiles after growth in the absence and presence of L-arabinose respectively.

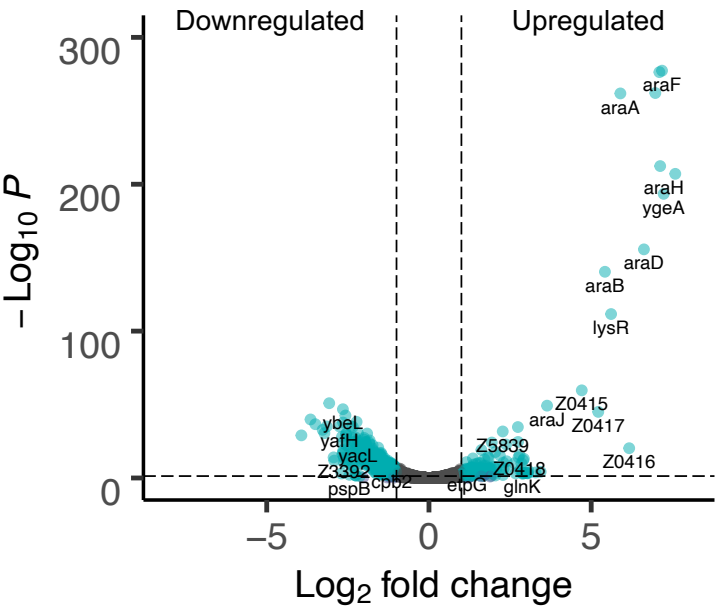


Figure 5-16. Differential gene expression post growth on L-arabinose in TUV93-0. Volcano plot of DEGs in TUV93-0 after growth in MEM-HEPES supplemented with 5 mg/mL L-arabinose. Data is representative of three biological replicates. The vertical dashed line indicates a *P*-value cut-off of $P < 0.05$. Horizontal lines equal < -1 (left) and > 1 (right) Log_2 fold change in gene expression.

5.14. L-arabinose induces differential expression across the LEE

Further exploring the RNA-seq dataset, the effect of L-arabinose on the differential expression of the LEE revealed the majority of genes to be upregulated (80 %), albeit to varying degrees (**Figure 5-17A**). Although not all meeting the significance cut-off after adjustment for multiple comparisons ($P \leq 0.05$), LEE1 was the only operon to display complete upregulation across all genes (**Figure 5-17A**). Those genes found to be significantly upregulated included *map* ($P < 0.01$), *espG* ($P < 0.05$), *escT* (LEE1) ($P < 0.05$) and *escL* (LEE1) ($P < 0.05$). Only a single gene, *sepZ* (LEE2) ($P < 0.01$), was found to be significantly downregulated across the LEE operons.

Despite carriage of some effectors by the LEE, a larger number of T3SS effectors are non-LEE encoded (NLEs) and located on OIs elsewhere on the chromosome. To determine whether L-arabinose might confer only LEE-specific changes, the expression of these NLE effectors was determined across the transcriptome. Similar to LEE expression, NLE effectors displayed variable expression. In total, 13/20 (65 %) NLEs were upregulated, however, it was only *nleF* ($P = 0.0001$) and *nleG6-3* ($P = 0.01$) that displayed significantly increased expression (**Figure 5-17B**). Only *espN* (*P*

< 0.05) and *nleG2-3* ($P < 0.01$) were significantly downregulated (**Figure 5-17B**). Additionally, although not effectors of the T3SS, two key regulators (Hfq and H-NS) involved in overseeing the expression of the LEE were found to be significantly downregulated (**Appendix Table 8-2**). Notably, these regulators act to repress LEE expression.

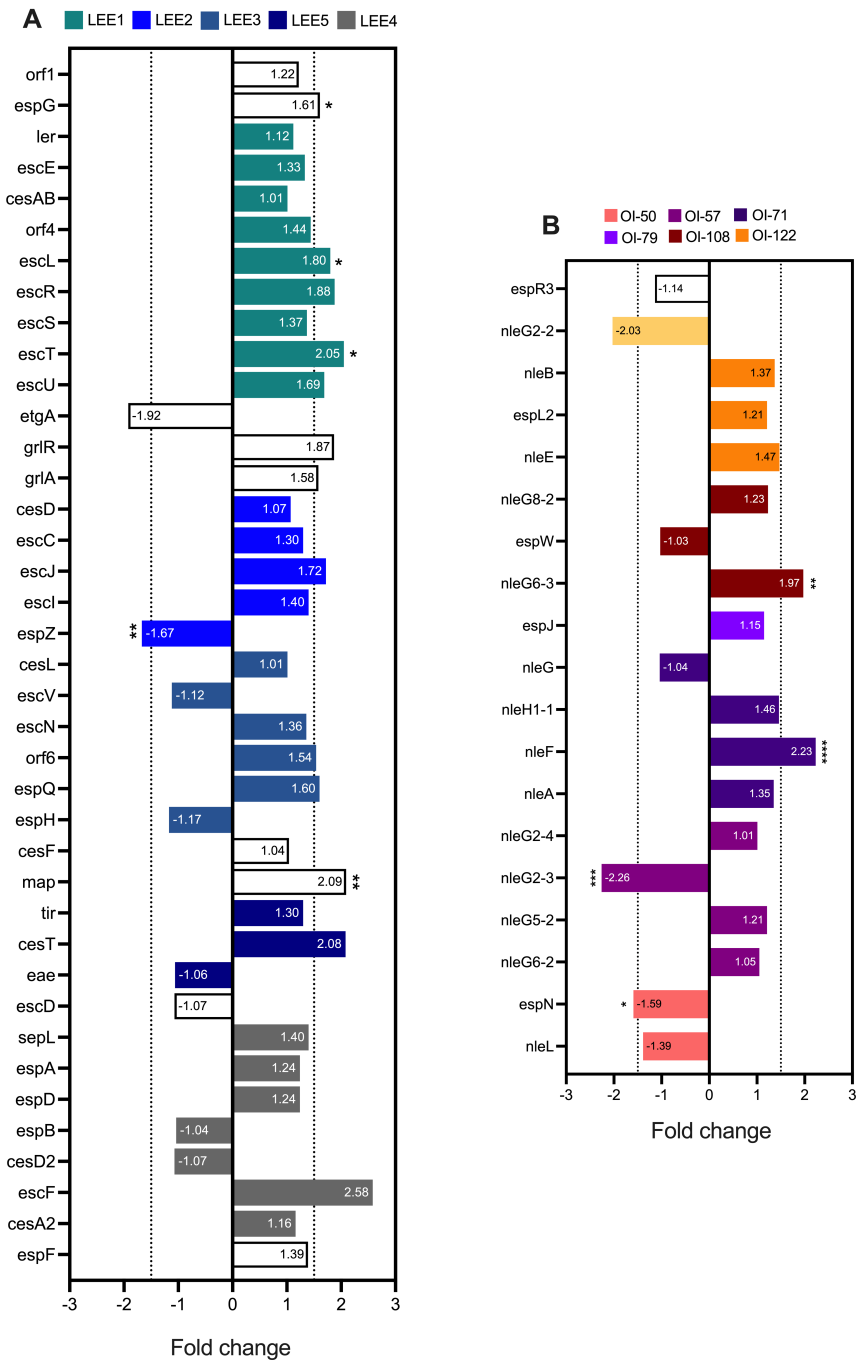


Figure 5-17. Differential gene expression across the LEE and NLE effectors in TUV93-0 in response to L-arabinose. (A) Absolute fold changes in the expression of genes across the LEE. Bars for the genes encoded within each of the LEE operons are colour coded respectively. Uncoloured bars are those genes not found within an individual operon but carried on the island. (**B**) Absolute fold changes in NLE effector expression. Bars for the genes carried on the same genomic OIs are colour coded. Uncoloured bars are those genes not carried on a genomic OIs. Dashed lines are representative of an absolute fold change cut-off of > 1.5 and < -1.5. (*) $P < 0.05$; (**) $P < 0.01$, (***) $P < 0.001$, (****) $P < 0.0001$. All data is representative of three biological replicates.

5.15. Distinct biological pathways are enriched by L-arabinose in EHEC

Functional analyses were performed on DEGs to identify any enriched pathways, where a network of interactions could be statistically inferred. A more stringent criteria of > 1 Log_2FC ($P \leq 0.05$) was used to filter DEGs as part of these analyses to reduce the number of nodes in the network and improve the clarity of interactions. Excluding disconnected nodes, a total of 332 downregulated genes were mapped into the network using the default parameters, highlighting 730 potential protein-protein interactions (**Figure 5-18**). Distinct pathways related to the signaling and regulation of biofilm formation (purple), cellular response to stress (orange), and metabolism of tyrosine (red), fatty acids, pyruvate (yellow) and hydrogen sulfide (green) were shown to be greatly enriched. More broadly, these interactions were grouped into 4 clusters using a hierarchical clustering that summarizes the correlation among significant pathways, whereby pathways that have many shared genes are clustered together (**Figure 5-18**). Whilst mostly containing mixed and uncharacterised genes, Cluster 1 significantly comprised of genes related to signalling. Cluster 2 was the smallest and contained genes related to both stress responses and protein refolding. Finally, Clusters 3 and 4 were largely related to metabolism of fatty acids and hydrogen sulphide, respectively.

Fewer nodes were mapped for those genes upregulated (208/229) compared to those downregulated (**Figure 5-19**). Differences in pathway enrichment were also observed, with significant networks related to the metabolism of galactose, monosaccharides (mainly xylose and L-arabinose) and nitrogen, as well as biotin biosynthesis, amino acid transport and QS. The significant enrichment of metabolic pathways was reflected when performing hierarchical clustering analyses (**Figure 5-19**). Cluster 1 was the largest and comprised of monosaccharide transport and metabolism. Although specific for amino acids, Cluster 3 was similarly enriched with genes related to transport and metabolism. In contrast, Cluster 2 contained genes involved in the utilisation of nitrogen.

Whilst the majority of enriched pathways were largely related to metabolism, there were a number of genes related to biofilm regulation and quorum sensing that were also differentially expressed, as identified by the STRING analysis. This included the downregulation of major biofilm regulatory factors such as *bssR* (FC = -3.75; $P <$

0.001), *bssS* (FC = -2.79; $P < 0.001$), *mcbR* (FC = -2.29; $P < 0.001$), and *glgS* (FC = -2.53; $P < 0.001$). Conversely, the following *lsr* QS genes were significantly upregulated: *lsrB* (FC = 2.41; $P < 0.001$), *lsrC* (FC = 1.62; $P < 0.05$), *lsrD* (FC = 1.89; $P < 0.01$), *lsrF* (FC = 3.12; $P < 0.001$), *lsrG* (FC = 3.23; $P < 0.001$), *lsrK* (FC = 2.83; $P < 0.001$), and *lsrR* (FC = 2.68; $P < 0.001$). In addition, genes encoding enzymes involved in the synthesis of QS signalling molecule AI-2 were also significantly upregulated (e.g., *metB* and *metK*). There was found to be no obvious pathways that were enriched linking the metabolism of L-arabinose to the enhanced LEE phenotype observed.

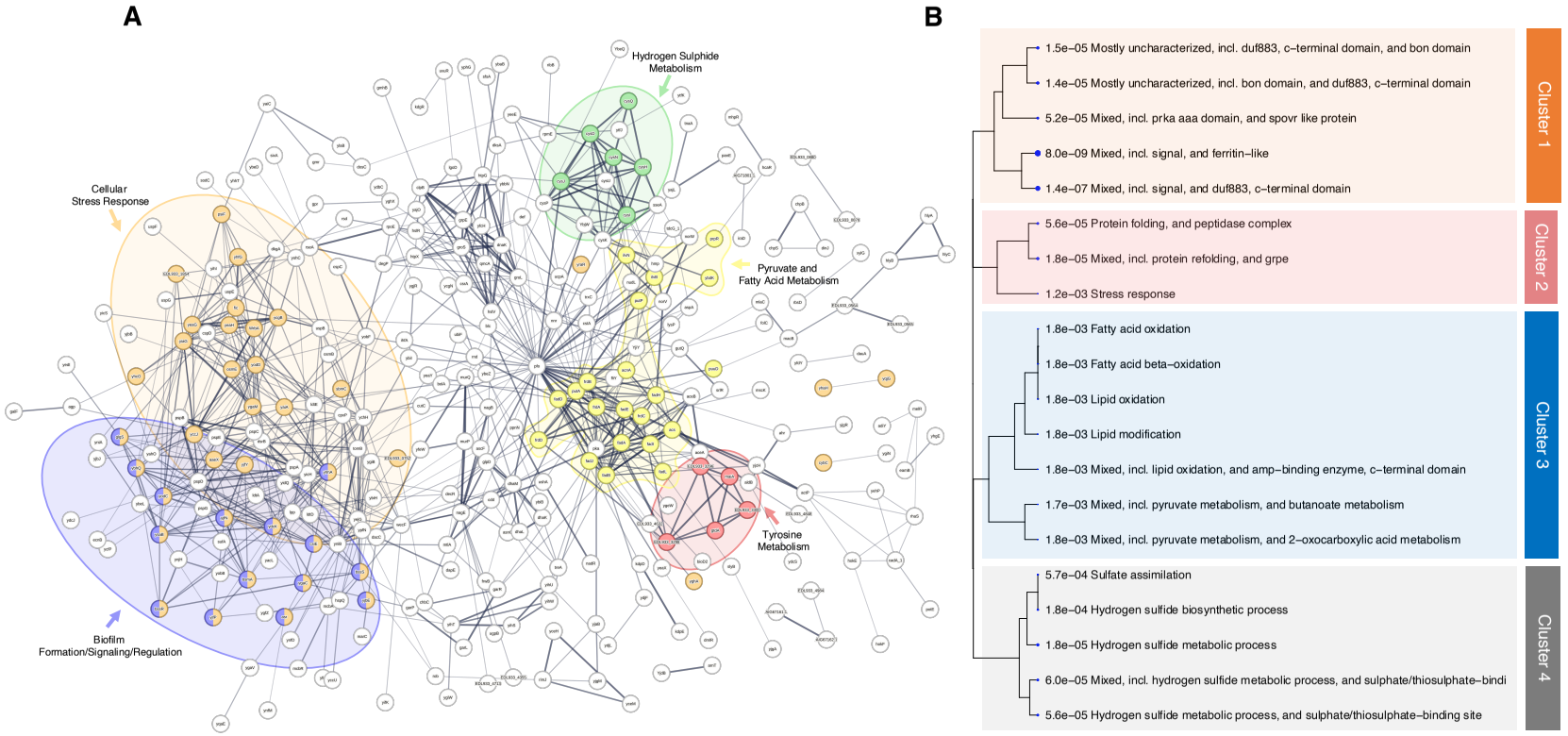


Figure 5-18. Functional enrichment analysis of significantly down-regulated genes in TUV93-0 when grown on L-arabinose. (**A**) STRING network analysis of predicted protein-protein interactions based upon upregulated genes. Line thickness is indicative of confidence in the interactions between proteins. Disconnected nodes were removed from the network to improve clarity. Distinct clusters that are enriched are in coloured bubbles. (**B**) Hierarchical clustering tree summarising the correlation among significant pathways generated using ShinyGO. Pathways that are enriched with shared genes are clustered together. Bigger dots indicate more significant *P*-values

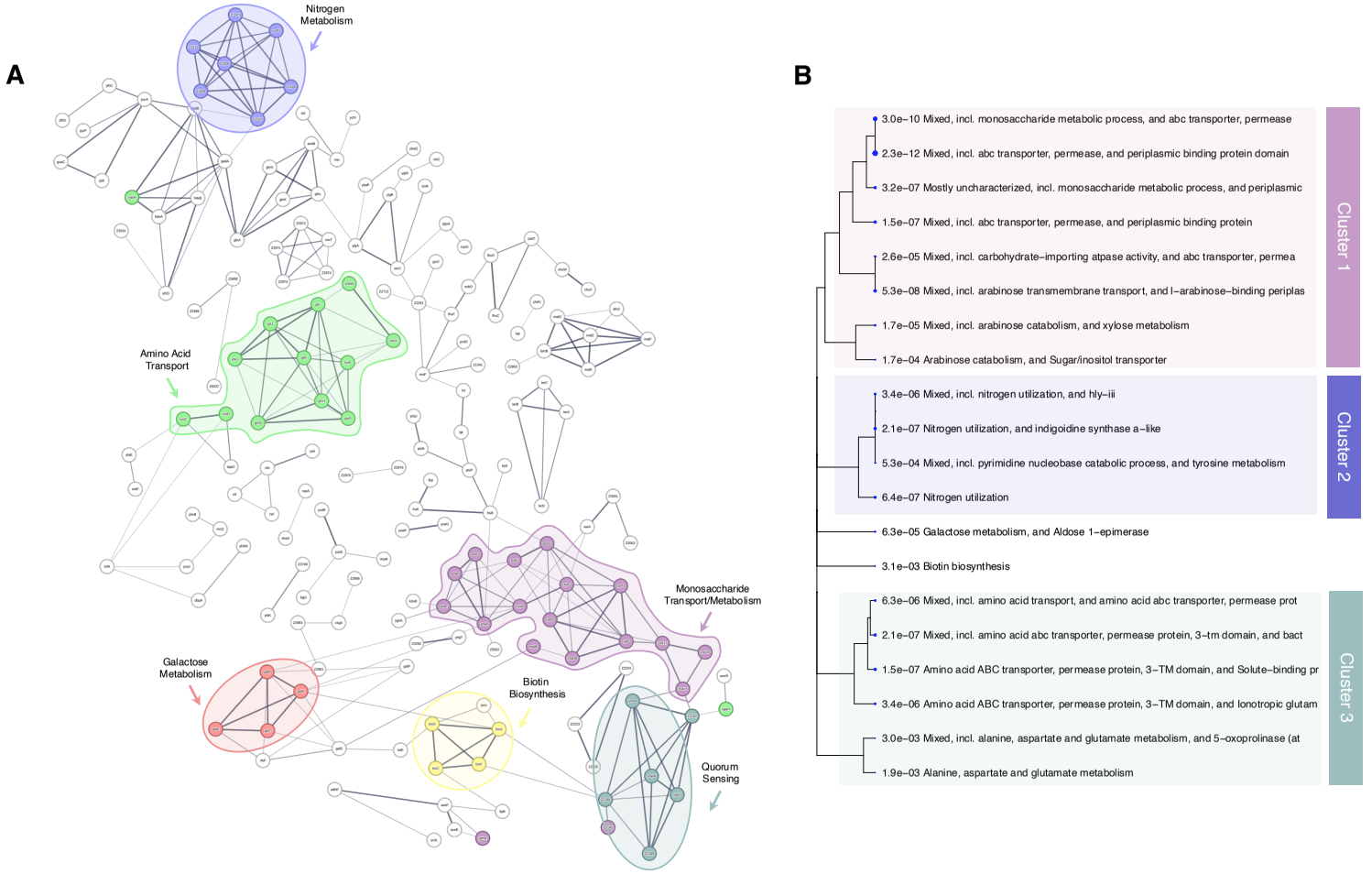


Figure 5-19. Functional enrichment analysis of significantly upregulated genes in TUV93-0 when grown on L-arabinose. (**A**) STRING network analysis of predicted protein-protein interactions based upon upregulated genes. Line thickness is indicative of confidence in the interactions between proteins. Disconnected nodes were removed from the network to improve clarity. Distinct clusters that are enriched are in coloured bubbles. (**B**) Hierarchical clustering tree summarising the correlation among significant pathways generated using ShinyGO. Pathways that are enriched with shared genes are clustered together. Bigger dots indicate more significant *P*-values

5.16. Investigating the effects of L-arabinose on biofilm regulators

To further explore and validate the effects of L-arabinose on biofilm formation identified by RNA-seq, reporters were constructed for the two biofilm repressors, BssS and BssR. The promoter regions for both genes were subsequently cloned into pMK1*lux*. Analysis of both P_{bssR} and P_{bssS} expression demonstrated activity to decrease over time, before eventually ceasing entirely, when media was supplemented with 5 mg/mL L-arabinose. These data suggested that the utilisation of L-arabinose is responsible for suppressing *bssR* and *bssS*. To further confirm whether the repression of *bssR* and *bssS* was L-arabinose-dependent, P_{bssR} and P_{bssS} expression was measured in Δ*araC* under the same conditions. When compared against the WT, P_{bssR} and P_{bssS} expression was significantly higher and no longer completely repressed in the presence of L-arabinose at 5 mg/mL. However, deletion of *araC* did not enable expression levels to reach those observed in the absence of the sugar, although being more similar for *bssR*. Comparatively, the maximal expression of P_{bssS} was found to be lower than that of P_{bssR}.

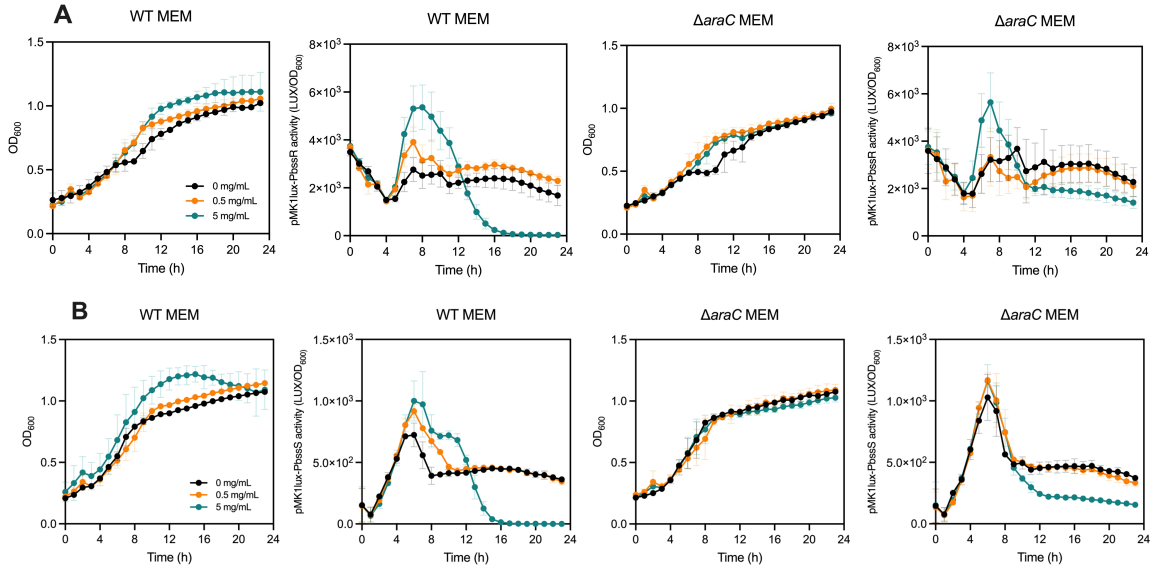


Figure 5-20. Regulatory biofilm proteins are suppressed by L-arabinose in TUV93-0. WT and Δ*araC* TUV93-0 with pMK1*lux*-P_{bssR} and pMK1*lux*-P_{bssR} in MEM-HEPES supplemented with 0, 0.5 and 5 mg/mL L-arabinose. Growth (OD_{600}) was measured over 24 h with readings taken hourly. The mean RLU was calculated by normalising the raw luminescence against the OD_{600} for each replicate. $n = 3$ biological replicates and error bars represent SD.

5.17. L-arabinose enhances TUV93-0 biofilm formation

Given that two known biofilm regulators were repressed by L-arabinose, the ability of TUV93-0 to form biofilms in the presence of L-arabinose was investigated (**Figure 5-21**). Biofilms were quantified following growth in MEM-HEPES supplemented with and without L-arabinose (**Figure 5-21A**). Significantly more biofilm was formed by the WT at increasing concentrations of L-arabinose ($P < 0.05$). Deletion of *araC* and therefore the inability to use L-arabinose prevented enhanced biofilm formation, irrespective of L-arabinose concentration. Complementation of Δ*araC* with pSU-*araC* restored enhanced biofilm formation at 5 mg/mL L-arabinose. These findings were also validated in M9 minimal media, where biofilm formation was found to be higher than in MEM-HEPES, regardless of sugar concentration (**Figure 5-21B**).

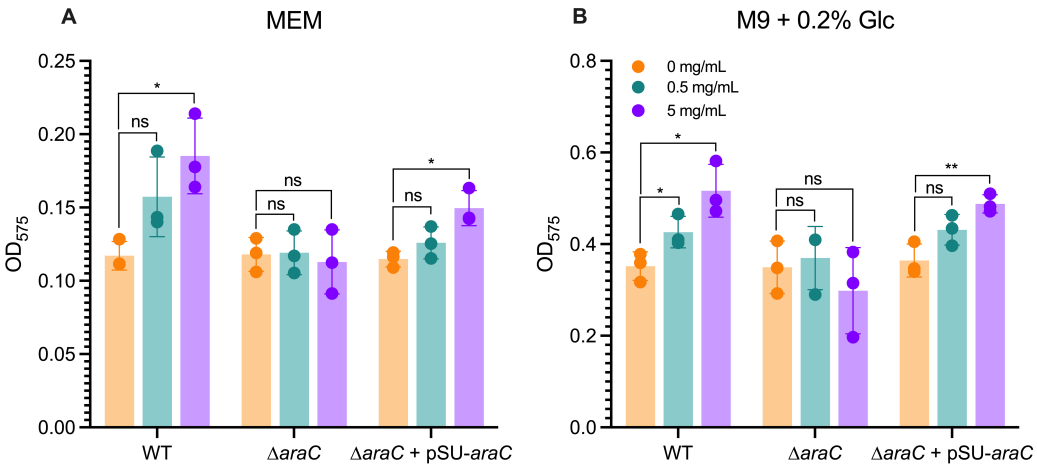


Figure 5-21. L-arabinose enhances biofilm formation in TUV93-0. WT, Δ*araC* and Δ*araC* complemented with *araC*-pSUPROM were grown in either (**A**) MEM-HEPES supplemented with and without L-arabinose or (**B**) M9 minimal media (+ 0.2 % glucose) statically at 37 °C for 24 h. Biofilm formation was quantified by the solubilisation of crystal violet with EtOH at OD_{575}. $n = 3$ biological replicates and error bars represent SD. (*) $P < 0.05$; (**) $P < 0.01$ - calculated using a student's *t*-test.

5.18. Deletion of *bssR* and *bssS* enhances biofilm formation in TUV93-0

BssR and BssS have been characterised as repressors of biofilm formation in *E. coli* K-12 (Domka et al., 2006). Qualitative assessment of the biofilms formed by TUV93-0 when grown in M9 minimal media and MEM-HEPES supported this claim and increased biofilm formation could be observed against the WT, particularly for Δ*bssS* (**Figure 5-22A-C**). This was then validated through the quantification of biofilm formed (**Figure 5-22DE**). A significant increase in biofilm formation was not observed for Δ*bssR* in MEM-HEPES when compared with the WT and could only be seen in M9 minimal media when supplemented with 5 mg/mL. Deletion of *bssS* however resulted in enhanced biofilm formation at 0.5 and 5 mg/mL in MEM-HEPES and across all sugar concentrations when grown in M9 minimal media. Whilst both are described as repressors of biofilm formation, based on these data BssS appears to have a more significant role in EHEC.

In addition to testing the effects of Δ*bssS* and Δ*bssR* on biofilm formation, the effect of their deletion on P_{LEE1} expression was tested due to their differential expression coinciding with that of the LEE identified via RNA-seq. However, it was found that no difference in the LEE phenotype was observed following their deletion, with enhanced P_{LEE1} expression still being observed, despite their absence (**Figure 5-23**). L-arabinose therefore acts independently to enhance biofilm formation and LEE expression.

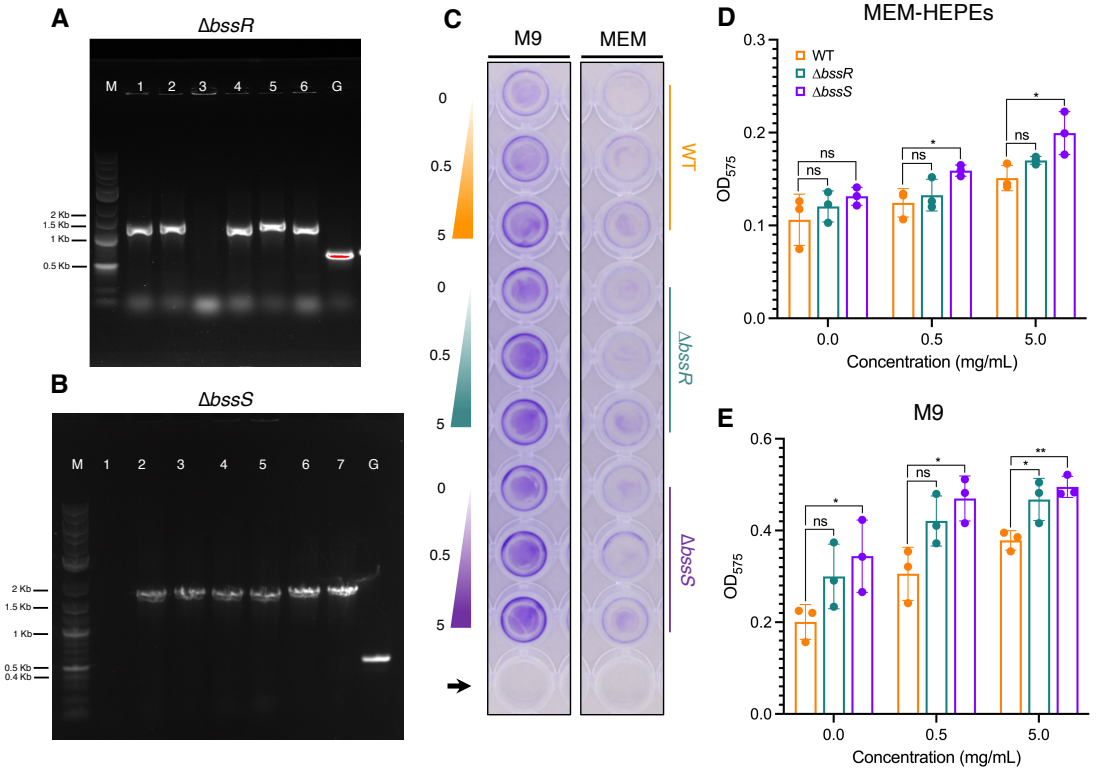


Figure 5-22. Deletion of *bssR* and *bssS* increases biofilm formation in TUV93-0. Gel electrophoresis confirmation of (**A**) Δ*bssR* and (**B**) Δ*bssS* for with the Cm^R (1.1 Kb) and Kan^R (1.6 Kb) cassettes, respectively. Numbered lanes correspond to colonies tested. Lanes M and G correspond to the 1 Kb plus DNA ladder and WT gDNA control, respectively. (**C**) Crystal violet staining of biofilm formation across WT, Δ*bssR* and Δ*bssS* TUV93-0 strains when grown in either M9 minimal media or MEM-HEPES supplemented with and without L-arabinose. Biofilm formation across (**D**) MEM-HEPES and (**E**) M9 minimal media was quantified by the solubilisation of crystal violet with EtOH at OD_{575}. The black arrow indicates wells with no bacteria (negative control). All growth was done statically at 37 °C for 24 h. $n = 3$ biological replicates and error bars represent SD. (*) $P < 0.05$; (**) $P < 0.01$ - calculated using a student's *t*-test.

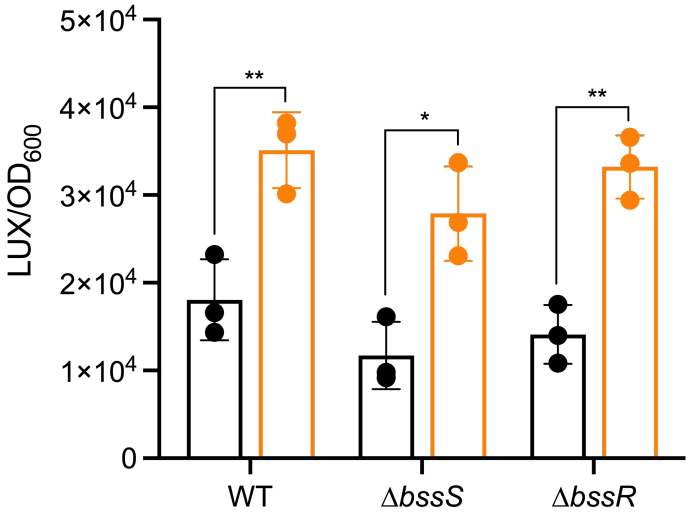


Figure 5-23. P_{LEE1} expression in Δ*bssS* and Δ*bssR*. WT, Δ*bssS* and Δ*bssR* TUV93-0 were grown in MEM-HEPES supplemented with and without 5 mg/mL L-arabinose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. $n = 3$ biological replicates and error bars represent SD. (*) $P < 0.05$; (**) $P < 0.01$ - calculated using a student's *t*-test.

5.19. L-arabinose and other *E. coli* pathotypes

As no single regulator or metabolite downstream of L-arabinose metabolism could be identified as being responsible for the enhanced LEE phenotype so far, it was asked whether the unknown driver was EHEC-specific. Following transformation of pMK1*lux*-P_{LEE1} into CFT073, CE10 and MG1655 assays were conducted to decipher whether P_{LEE1} enhancement was conserved amongst non-EHEC strains. When supplemented with increasing concentrations of L-arabinose, CE10 similarly to TUV93-0 displayed significantly increased P_{LEE1} expression (**Figure 5-24**). Expression was particularly enhanced between 7 and 9 h for 5 mg/mL when compared with 0 mg/mL. For MG1655, the effects of L-arabinose were less definitive than for TUV93-0 and CE10, however, a distinct period of significantly enhanced P_{LEE1} expression could be seen between 9 and 11 h for 5 mg/mL (**Figure 5-24**). In contrast, the reciprocal was seen for CFT073 harbouring pMK1*lux*-P_{LEE1}. Instead, P_{LEE1} expression reduced with increasing L-arabinose concentrations (**Figure 5-24**). The peak of P_{LEE1} expression in CFT073 at 5 mg/mL was therefore significantly lower than for CE10 (> 850-fold) and MG1655 (> 20-fold). This suggested that the stimulus responsible for enhancing LEE expression in TUV93-0 was not EHEC-specific and conserved in CE10 and MG1655 (to a lesser extent), but not in CFT073.

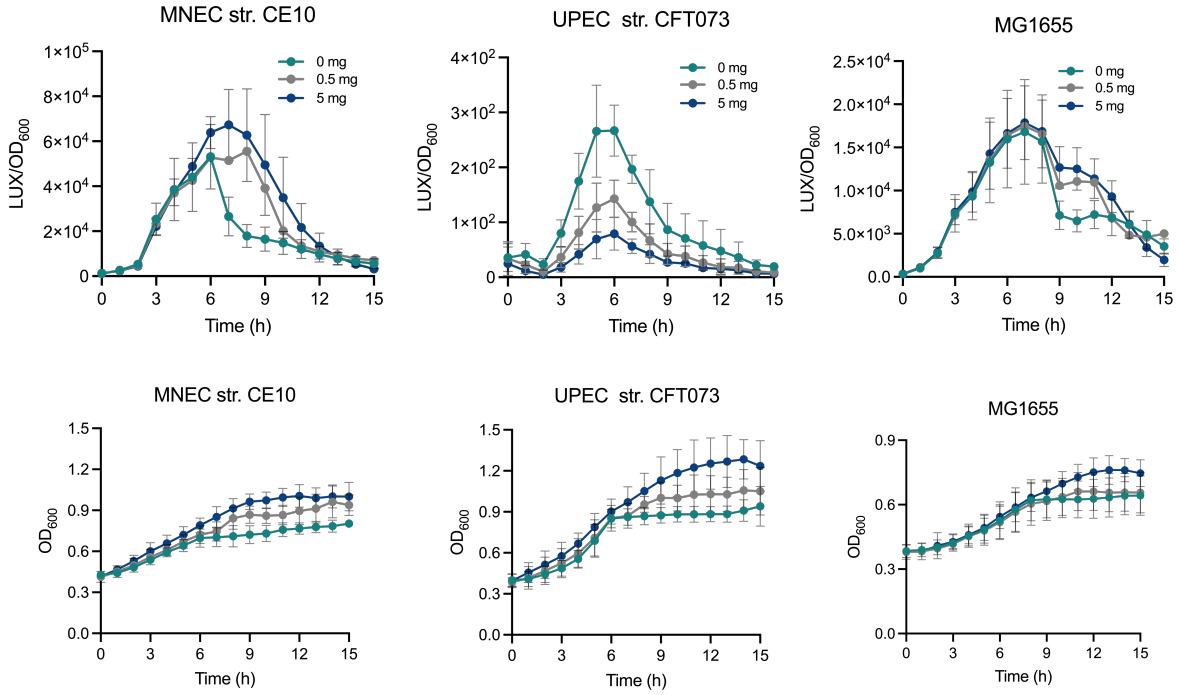


Figure 5-24. P_{LEE1} activity across non-EHEC strains. The UPEC str. CFT073, NMEC str. CE10 and lab str. MG1655 with pLEE1*lux* were grown in MEM-HEPES supplemented with and without L-arabinose for 15 h. Top panels show the mean RLU of each strain, which was calculated by normalising the raw luminescence against the OD_{600}. Bottom panels show the OD_{600} of each strain across all conditions tested. Measurements were taken every hour for 15 h. n = 3 biological replicates and error bars represent SD.

5.20. Exploring the effects of other aldopentose sugars on LEE expression

As L-arabinose was found to enhance LEE expression, the effect of other aldopentose on LEE expression was explored. Using the same reporter-based system as previous, EHEC was exposed to D-xylose and D-ribose. Both sugars exhibit the same chemical structure to each other and L-arabinose, with the only difference being the position of a single hydroxyl (-OH) group. When grown in the presence of D-ribose, P_{LEE1} expression was especially enhanced, with expression being sustained for an extended period time (**Figure 5-25A**). Although to a lesser extent, P_{LEE1} expression was also enhanced when grown in the presence of D-xylose (**Figure 5-25B**). In both cases enhanced P_{LEE1} expression corresponded with increasing sugar concentration. Notably, L-arabinose, D-ribose, and D-xylose all feed into the PPP and converge at D-xylulose-5-phosphate (**Figure 5-25C**), indicating a likely common mechanism for enhanced LEE expression.

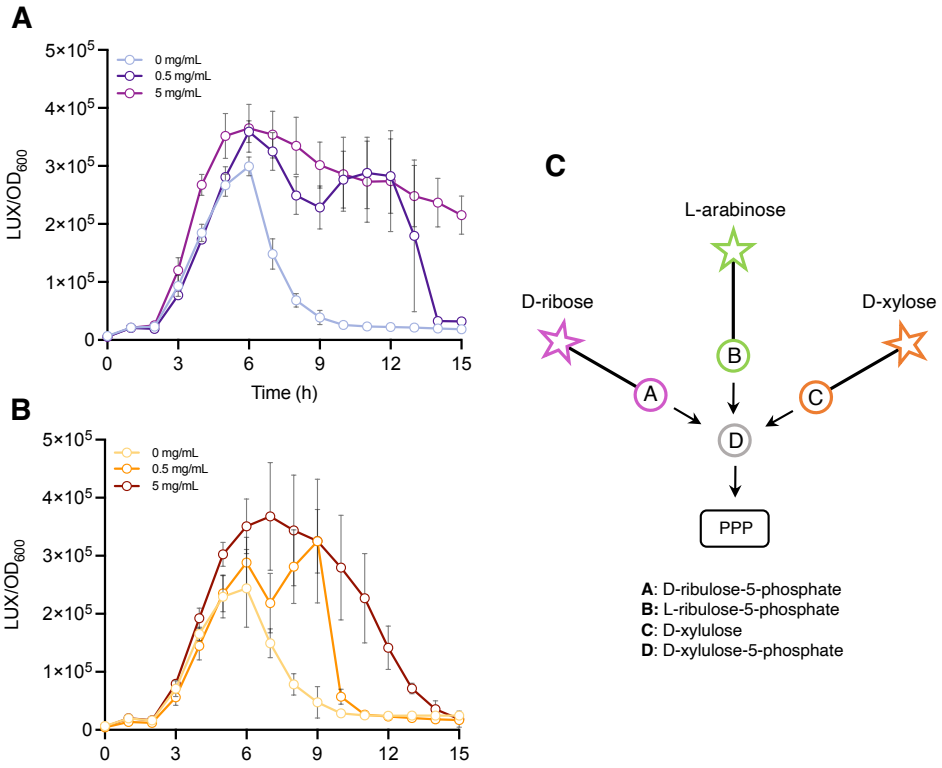


Figure 5-25. Other aldopentose sugars enhance P_{LEE1} expression in TUV93-0. pMK1*lux*-P_{LEE1} reporter assays when grown in MEM-HEPES supplemented with either (**A**) D-ribose or (**B**) D-xylose. The mean RLU was calculated by normalising the raw luminescence against the OD_{600} for each replicate. $n = 3$ biological replicates and error bars represent SD. (**C**) Overview of aldopentose sugar metabolism and entry into the PPP in *E. coli*.

As components of the D-ribose utilisation system had previously been shown to be upregulated during *C. rodentium* infection *in vivo* (Connolly et al., 2018), it suggested that the sugar might be important for virulence. Taking this into consideration, the enhanced P_{LEE1} expression observed following growth of D-ribose was further explored. The Rbs system, unlike the ara system, clusters at a single locus and encodes the necessary apparatus for regulation (*rbsR*), uptake (*rbsACB*) and metabolism (*rbsK*) of the sugar (**Figure 5-26A**). In addition, the locus carries a gene, *rbsD*, which is predicted to encode a D-pyranase (**Figure 5-26A**). Deletion of the entire *rbs* locus (**Figure 5-26B**) led to an inability to grow in M9 minimal media supplemented with D-ribose across a concentration range (**Figure 5-26C**), supporting its characterised role. Then, when transformed with pMK1*lux*-P_{LEE1}, Δ*rbs* no longer displayed a significant enhancement in P_{LEE1} expression (**Figure 5-26D**). This suggested the same LEE enhancing effects seen for L-arabinose to be shared amongst other aldopentose sugars.

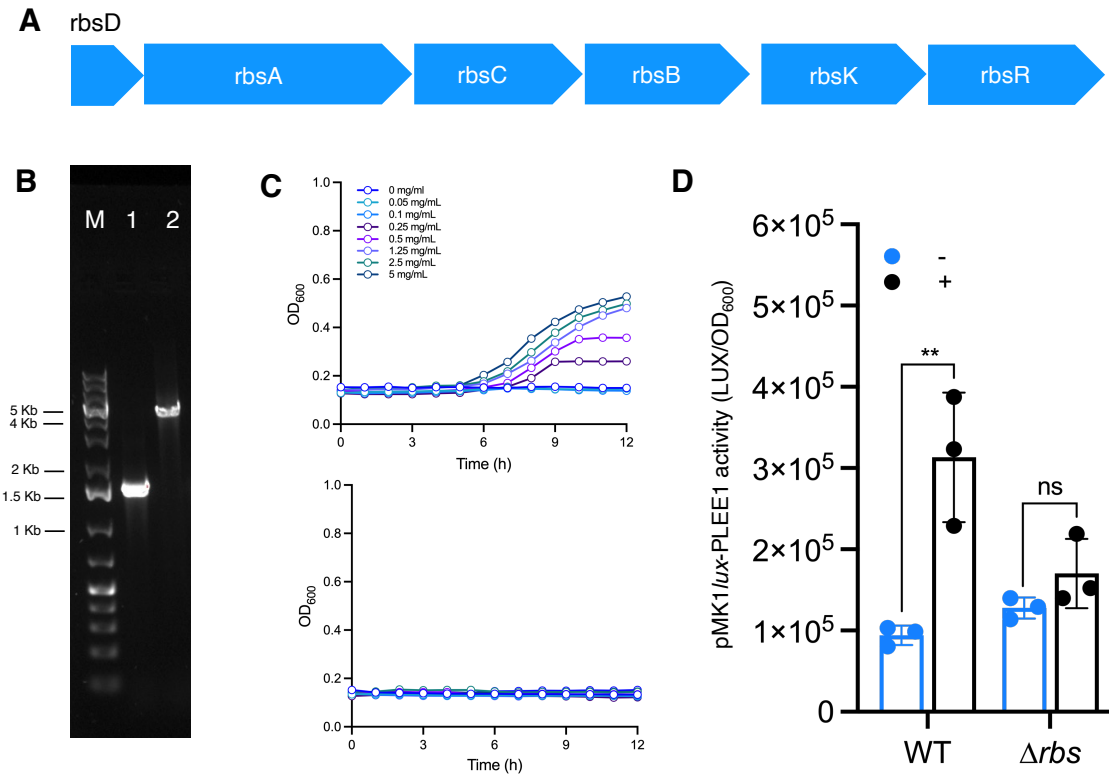


Figure 5-26. Deletion of the *rbs* locus prevents enhanced P_{LEE1} expression in TUV93-0. (**A**) *rbs* locus encoding the D-ribose utilisation system in *E. coli*. (**B**) Agarose gel confirmation of Δ*rbs* in TUV93-0. Lane 1 corresponds to the colony tested for replacement of *rbs* with the Kan^R cassette (1.6 Kb). Lane 2 corresponds to the WT gDNA control. Lane M corresponds with the 1 Kb plus DNA ladder (**C**) WT and *rbs* mutants were grown in M9 minimal media supplemented with D-ribose across a concentration range. OD_{600} was measured every hour for 12 h. (**D**) WT and Δ*rbs* TUV93-0 were grown in MEM-HEPES with and without 5 mg/mL D-ribose for 7 h. The mean RLU was calculated by normalising the raw luminescence against the OD_{600}. *n* = 3 biological replicates and error bars represent SD. (*) $P < 0.05$, (**) $P < 0.01$; (***) $P < 0.001$ - calculated using a student's *t*-test.

5.21. Testing the role of identified sugar systems during *in vivo* infection

The data so far had revealed that L-arabinose metabolism via AraBAD resulted in pleiotropic phenotypes (related to nutrition, LEE, and biofilm) that could all benefit EHEC infection of the gut. To address whether L-arabinose metabolism via AraBAD is important *in vivo*, BALB/c mice were mono-infected with WT *C. rodentium* (n = 10) and Δ*araBAD* (n = 10). Colonisation of the host by WT and mutant *C. rodentium* strains was indistinguishable during the early and peak phases of infection, with no significant differences in the faecal shedding (**Figure 5-27A**). As infection progressed through peak infection, the Δ*araBAD* mutant was much more rapidly cleared, and significant

difference could be observed in the faecal shedding of WT and Δ*araBAD* from day 13 onwards ($P < 0.01$) (**Figure 5-27A**).

To test the relative contribution of *araBAD* to *C. rodentium* fitness under competitive conditions, BALB/c mice (n = 10) were co-infected with the mutant and WT *C. rodentium* at a 1:1 ratio. When calculating the competitive indices, a significant difference in the faecal shedding between WT and mutant was similarly observed ($P < 0.01$) (**Figure 5-27B**). The *araBAD* mutant was also cleared four days earlier than seen for during mono-infection (Day 13). Correspondingly, counts for Δ*araBAD* were significantly lower than for the WT when harvested directly from colonic tissue ($P < 0.01$) at day 13 (**Figure 5-27C**), suggesting AraBAD to have an important role during infection of the host.

In addition to Δ*araBAD*, the effect of Δ*rbl* and Δ*rbs* during *in vivo* infection was also investigated. The Δ*bssS* mutant was also included due to its differential expression identified by RNA-seq and its demonstrated role in suppressing biofilm formation *in vitro*. As previous, BALB/c mice were mono-infected with WT or Δ*rbl*, Δ*rbs* or Δ*bssS* mutants (n = 4). No significant differences were seen in the faecal shedding across any of the time points for the mutants when compared with the WT (**Figure 5-28**). However, it is worth noting that Δ*bssS* and Δ*rbs* displayed an earlier onset of clearance compared to the WT, with the latter also trending towards a lower pathogen burden within the mice. Further replicates would be required to confidently determine the roles of these genes during infection.

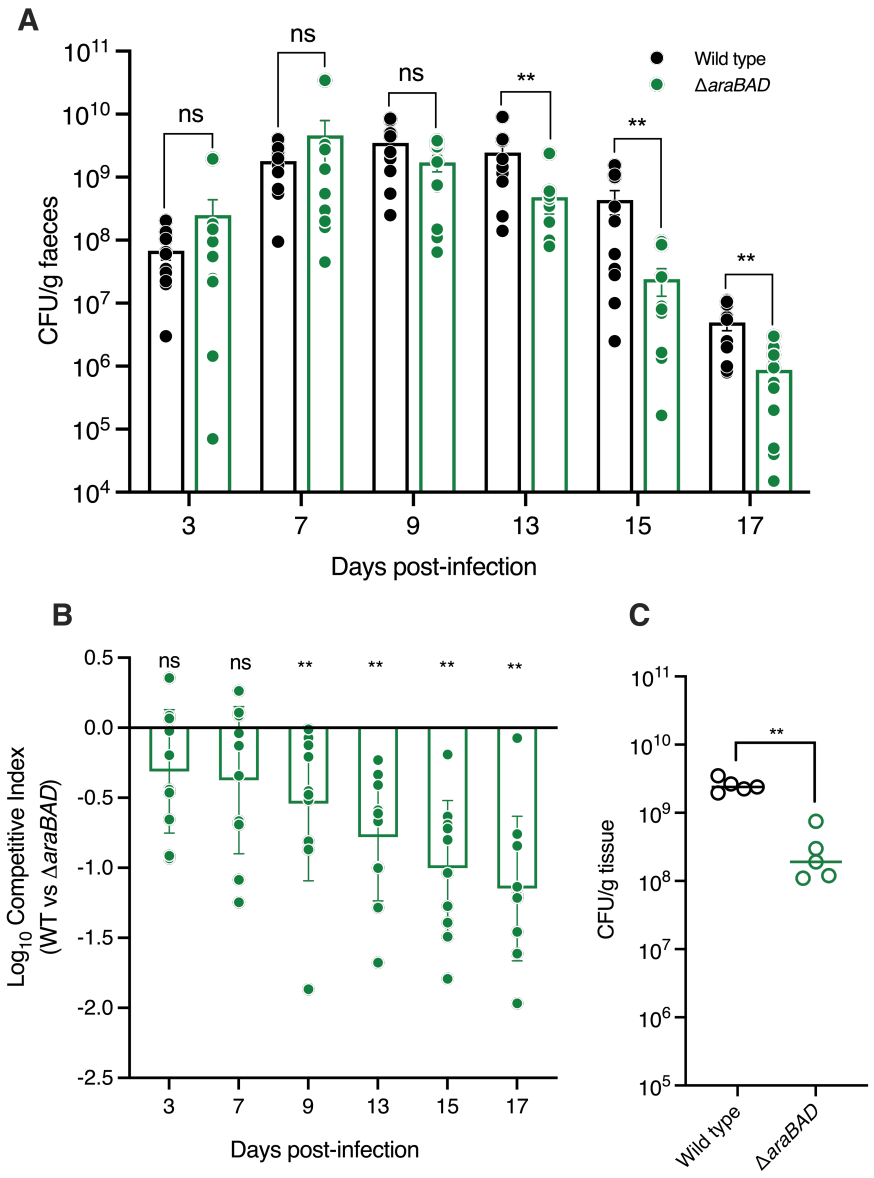


Figure 5-27. L-arabinose metabolism is required for *in vivo* infection. (**A**) Mono-infection of mice by WT or Δ*araBAD C. rodentium* (groups of n = 10). Bars represent the mean bacterial load in faeces of infected mice (CFU/g) across 17 days. Statistical significance was determined using a Mann-Whitney U-test. (**B**) Competitive index of WT *C. rodentium* versus Δ*araBAD* during infection. Mice were co-infected with both WT and Δ*araBAD C. rodentium* 1:1 (n = 10). Statistical significance was determined using the Wilcoxon signed rank test. (**C**) CFU obtained per gram of colonic mouse tissue after co-infection with both WT and Δ*araBAD C. rodentium* (n = 5) for 13 days. Statistical significance was determined using a Mann-Whitney U-test. Error bars represent SEM. (**) $P < 0.05$.

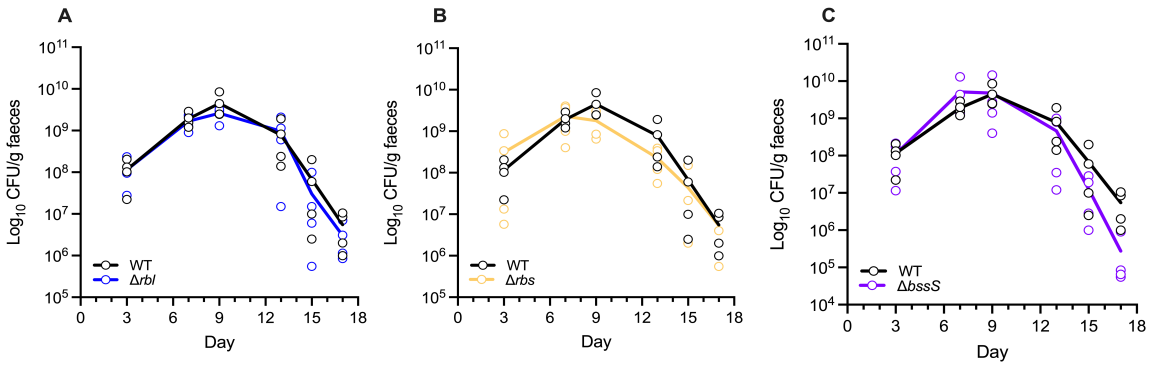


Figure 5-28. Deletion of *rbl*, *rbs* and *bssS* have no significant effect on *C. rodentium* infection *in vivo*. Mono-infection of mice with either (**A**) Δ*rbl,* (**B**) Δ*rbs* or (**C**) Δ*bssS* compared with WT *C. rodentium* (n = 4). Points represent the CFU/g of faeces for individual replicates across 17 days.

5.22. Discussion

Nutrients in the intestinal environment have been extensively demonstrated to regulate the expression of key virulence traits in EHEC, mainly acting to exert regulatory influence over the LEE (Jimenez et al., 2019; Njoroge et al., 2012; Pacheco et al., 2012). Typically, nutrients are sensed by a regulator that either (a) directly interacts with regulatory regions of the LEE (i.e., the *ler* promoter) (Menezes-Garcia et al., 2020) or (b) affects the activity of core regulators already implicated in controlling LEE expression (Nakanishi et al., 2009). Here, L-arabinose was shown to enhance LEE expression, specifically via its metabolism as a nutrient rather than being sensed as a signal.

Prior to this work, there was evidence to suggest L-arabinose to be important for the pathogenic lifestyle of EHEC. This included the significant expression of AUGs during *in vivo* infection of both humans (John et al., 2005) and mice (Gardette et al., 2019), as well as the advantage conferred by being able to metabolise L-arabinose during colonisation of the streptomycin treated mouse gut (Fabich et al., 2008). The observation that LEE carriage almost never occurred without the OI-17 Z0415-19 locus, expressed specifically in response to L-arabinose, further highlighted a link between the sugar and EHEC pathogenesis. It could be that the carriage of the LEE acts as a selective pressure to retain Z0415-19, hence the strongly identified association. Despite these indications, how L-arabinose is implicated in EHEC virulence has never been directly addressed.

In this study, LEE1 was shown to be up-regulated across reporter-based assays, RT-qPCR and the RNA-seq dataset in response to growth on L-arabinose. The significant effect of L-arabinose on LEE1 suggested the sugar to potentially confer a regulatory role via an interaction with the LEE1 regulatory region. However, deletion of the genes required for L-arabinose uptake, metabolism and associated regulation revealed the phenotype to rely upon L-arabinose breakdown and not just merely "sensing" its presence by AraC or an alternative regulator. Altered transport via Δ*araE* and Δ*araFGH* had contrasting effects on P_{LEE1} expression in EHEC. The absence of any change in expression following the deletion of *araFGH* could be attributed to the presence of *araE*, which as described in chapter 4, is the predominant route of L-arabinose uptake in EHEC. However, even with the deletion of *araE*, the capacity of L-arabinose to enhance P_{LEE1} expression was still observed and was confirmed also by complementation. It is likely that even in the absence of both transporters a similar phenotype would still be observed based on the ability of a double transporter mutant to still grow on L-arabinose as the sole carbon source.

As media such as MEM-HEPES contains D-glucose it was necessary to carefully consider the onset of enhanced LEE expression by L-arabinose as catabolite repression would block metabolism of the sugar. Using TLC, it was revealed that L-arabinose is not likely to be used until at least 7 h after the experimental start time, giving reason as to why more pronounced effects on LEE expression were not observed until later in growth. That being said, TLC is a means to qualitatively assess sugar utilisation in the media. Therefore, it may be more appropriate to quantify the kinetics of L-arabinose use over time using high performance liquid chromatography in follow-up experiments to allow the accurate detection of the sugar.

The metabolism of L-arabinose being necessary for enhanced LEE expression raised the question as to what the signal might specifically be for driving the phenotype. At its simplest L-arabinose is catabolised via a series of enzymes that terminate with the production of D-xylulose-5-P. This substrate then enters the PPP and via its conversion can enter various pathways of metabolism. Subsequently, the list of potential substrates that could drive enhanced LEE expression as a result of L-arabinose is extensive. RNA-seq revealed no enzymes or regulators related to the route L-arabinose would take through central metabolism that might be responsible for the LEE

phenotype other than the upregulation of *araBAD*. The contribution of each of these genes and the subsequent metabolites produced as a result of that reaction was tested by a series of complementation assays, revealing none of the by-products of 'primary' L-arabinose metabolism to be the signalling driver. The likelihood of a downstream metabolite being responsible for enhanced LEE expression was substantiated by a similar phenotype being observed for D-xylose and D-ribose, which all share the same metabolic fate in the cell (Sprenger, 1995). Though converging at D-xylulose-5-P, D-ribulose-5-P is able to be produced from all three sugars via this intermediate (Mayer & Boos, 2005). D-ribulose-5-P is a product of primary D-ribose metabolism formed from the isomerisation of D-ribose-5-phosphate via RpiA/B (Mayer & Boos, 2005). If responsible for enhancing the LEE, the production of D-ribulose-5-phosphate directly as part of D-ribose metabolism, may help to explain the extensively sustained P_{LEE1} expression seen for D-ribose in the reporter assays. To get from L-arabinose to D-ribulose-5-phosphate, *rpe*, encoding a ribulose-phosphate 3-epimerase, is required (Mayer & Boos, 2005). Whilst not significant, the gene was found to be modestly upregulated as part of the RNA-seq data set (FC = 1.41; $P = 0.24$).

Additionally, D-ribulose-5-P has been previously described to undergo spontaneous conversion to AI-2 (Tavender et al., 2008), also demonstrated to positively regulate LEE expression (Bansal et al., 2008). Although this spontaneous conversion from D-ribulose-5-P is considered to produce extremely negligible amounts of AI-2 (Tavender et al., 2008), this additional AI-2 could be a way of fine-tuning the LEE response, in addition to the AI-2 already produced by LuxS. Components of the system specific for AI-2 transport and breakdown were identified to be significantly upregulated as part of this RNA-seq dataset. That being said, it is unlikely that the enhanced LEE phenotype induced by L-arabinose metabolism arises from extremely small quantities of AI-2 being produced spontaneously, as this would prove greatly unreliable in allowing the repeatedly enhanced LEE expression seen for L-arabinose. Alternatively, L-arabinose may act via AI-2 to enhance LEE expression through the Lsr system. A mechanism for one such possibility could be linked to the positive regulatory activities of cAMP-CRP complex binding which is shared between both the L-arabinose and Lsr system. Important to note is that the influence of AI-2 on the LEE is conflicted by evidence to suggest that the signalling molecule is unable to activate the LEE (Sperandio et al., 2003). For example, the addition of exogenous AI-2 to media had no effect on LEE

expression (Sperandio et al., 2003). However, more recent work by Bansal *et al*., (2008) demonstrated AI-2 to temporally regulate virulence expression by positively upregulating 23 genes of the LEE and > 50 genes related to additional aspects of EHEC virulence (e.g., flagella biosynthesis, iron acquisition, colanic acid production) (Bansal et al., 2008). To address the contribution of D-ribulose-5-phosphate, P_{LEE1} reporter assays would need to be conducted in direct response to EHEC being exposed to the sugar phosphate. Alternatively, mutants in the enzymes required for the D-ribulose-5-P production could be used for RT-qPCR following growth on L-arabinose to determine the effects on LEE expression. Ultimately, there is evidence to suggest D-ribulose-5-P could be the metabolite driving enhanced LEE expression.

In theme with exploring candidate metabolites of downstream metabolism of L-arabinose, pyruvate in the presence of oxygen is known to enhance LEE expression in EHEC (Carlson-Banning & Sperandio, 2016) and was also demonstrated as part of this work. As a core metabolite of cellular metabolism, all three aldopentose (L-arabinose, D-xylose, D-ribose) could feasibly enhance LEE expression through their eventual conversion to pyruvate. Though how pyruvate confers these effects remains somewhat unclear, particularly as the deletion of *pdhR* had no effect on LEE expression in the presence of both L-arabinose and pyruvate. This was in spite of the regulator being significantly upregulated (FC = 2.04, $P < 0.01$) following growth in media supplemented with L-arabinose. Notably, pyruvate was found to increase the growth of EHEC despite it being previously shown not to (Carlson-Banning & Sperandio, 2016). That being said, these discrepancies could be attributed to differences in strain (EHEC str. 86-24) and media conditions (DMEM) used (Carlson-Banning & Sperandio, 2016).

Despite evidence to support the hypothesis that L-arabinose enhances LEE expression in EHEC via its conversion to core metabolites, no increase was observed for P_{LEE1} activity in UPEC str. CFT073. In fact, P_{LEE1} expression was shown to significantly decrease in the presence of L-arabinose. As seen for *E. coli* str. MG1655 and NMEC str. CE10, if the stimulus for increased LEE expression was part of central metabolism it would likely be also shared by CFT073, particularly as this strain has the metabolic capacity to produce both D-ribulose-5-P, AI-2, and pyruvate. In explanation, UPEC may not encode the specific regulatory protein that responds to downstream

metabolic products of L-arabinose, potentially responsible for exerting enhanced LEE expression. Although, this would not explain the decreased P_{LEE1} expression with increasing concentration of L-arabinose. To address this question, comparative transcriptomics could be conducted to compare regulatory protein expression profiles of EHEC and UPEC when grown L-arabinose. Furthermore, it is interesting to note that EDL933 and CE10 both encode Z0415-19 whilst CFT073 does not.

In concurrence with the upregulation of QS genes and pathways as identified through STRING analysis, differential expression was also observed in genes associated with biofilm formation in the presence of L-arabinose. Based upon the RNA-seq dataset it was initially thought that L-arabinose negatively affected biofilm formation as seen for *Salmonella* (Vasicek et al., 2021) due to related genes being significantly downregulated. However, these genes were identified to be known repressors of biofilm formation, suggesting L-arabinose in fact to promote biofilm production through repressing negative regulatory proteins, BssS and BssR. In *E. coli* K-12, these two small cytoplasmic proteins reduce biofilm formation by preventing AI-2 uptake (Domka et al., 2006). AI-2 promotes biofilm formation by blocking a downstream cascade of negative biofilm regulators (González Barrios et al., 2006; Wood, 2009). Notably, genes of the AI-2 uptake machinery were significantly upregulated in response to growth on L-arabinose. In further support of L-arabinose promoting EHEC biofilm formation, genes required for curli amyloid fibril biogenesis (namely *csgA* and *csgC*), which are major components of bacterial biofilms (Bhoite et al., 2019), were significantly upregulated.

The ability to enhance biofilm production was validated through phenotypic biofilm assays, whereby L-arabinose significantly enhanced biofilm formation *in vitro*. Comparatively, the effect of the sugar on *bssS* and *bssR* expression was similar, with L-arabinose rapidly blocking the expression of the two genes at high concentrations. Whilst deletion of *araC* significantly alleviated the repressive effect of the sugar, it was not entirely sufficient to restore WT expression levels. This suggested that perhaps a second regulatory factor was working in tandem to confer the repressive effects on *bssS* and *bssR* expression. Due to the implication of catabolite repression conferred by non-glucose sugars (Aidelberg et al., 2014), as well as the already discussed relationship of AraC with CRP, it is hypothesised that this additional factor could be

CRP. As such, cAMP-CRP has been previously described to promote biofilm formation in *E. coli* via the promotion of *csg* gene expression and repression of *rpoS* (Liu et al., 2020). Therefore, L-arabinose via cAMP-CRP could act to also encourage biofilm formation through the repression of associated regulatory proteins, BssS and BssR. The ability of these two proteins to repress biofilm formation in EHEC was evidenced by their deletion allowing for enhanced biofilm formation. In fact, Δ*bssS* displayed enhanced biofilm formation even prior to L-arabinose being supplemented when compared to the WT. This suggested BssS to exert greater regulatory influence over the ability to produce biofilms than BssR.

Whilst much of this work has focused on virulence expression in EHEC, a mechanism of how L-arabinose likely aids persistence of the pathogen via biofilm formation is proposed (**Figure 5-29**). Interestingly, there is overlap in the regulators that regulate LEE expression and biofilm formation (Sharma & Bearson, 2013), providing a link between two independent pathogenic strategies. Outside of the human host, EHEC have been demonstrated to form biofilms on several abiotic and biotic surfaces including plants (Kim et al., 2016), of which L-arabinose is a major constituent of (Crozier et al., 2021). There is also now evidence to suggest that enteric pathogens such as *E. coli* and *Salmonella* are able to form colonic biofilms associated with the mucus during colitis (Guo et al., 2023). As these biofilms display resistance to antimicrobials and host defences (Kim et al., 2016), a mechanistic understanding of the factors that contribute to biofilm formation need to be considered. Taken together, the results strongly support the idea of L-arabinose being an important dietary-sourced regulator of EHEC pathogenesis by promoting its colonisation, enhancing its T3SS expression and potentially allowing for longer-term gut persistence via biofilm production.

Finally, the implication of the systems discussed in this thesis in pathogenesis were addressed by *in vivo* infections using *C. rodentium. In vivo*, sugars have been demonstrated to play an important role in allowing host colonisation and regulation of virulence phenotypes (Fabich et al., 2008; Jimenez et al., 2019; Le Bihan et al., 2017; Yang et al., 2023). Here, deletion of *araBAD* enabled the pathogen to be cleared from mice far more rapidly than the WT, suggesting that the metabolism of L-arabinose is important during *C. rodentium* infection. From these results alone, it is not possible to

determine whether the defect conferred by the deletion of *araBAD* is due to an impaired ability to colonise, reduced LEE expression, or both. In support of the former, a study by Fabich *et al*. (2008) found that deletion of *araBAD* in EHEC negatively impacted colonisation of mice *in vivo* (Fabich et al., 2008). Additionally, based on the *in vitro* data presented here, there is also strong evidence to support LEE expression being affected. Constitutive expression of the LEE in Δ*araBAD* during in infection, as has been done previous (Connolly et al., 2018), could help to elucidate whether the differences observed were due to defective LEE expression, specifically.

Whilst mice were not supplemented with dietary L-arabinose, the sugar was in fact detected within tissue (Data not shown). Therefore, if follow up experiments were to modify the diet to include L-arabinose, it is likely that the difference observed would be more greatly pronounced. To better understand the mechanisms conferred by L-arabinose specific to EHEC, future work could focus on the use of streptomycin-treated mice. Using *C. rodentium*, a significant difference in pathogen clearance was seen earlier during co-infection than in mono-infection, suggesting a competitive element to L-arabinose utilisation and fitness during infection. Use of the streptomycin model could also provide insight into the potential competitive implications of the native gut microbiota in L-arabinose-mediated EHEC fitness (i.e., both colonisation and virulence) during infection.

No significant phenotype was identified for the other systems tested *in vivo*. It is likely that these experiments were underpowered and further replications, as per the Δ*araBAD* experiments, as well as co-infection experiments, may help to reveal any significant effects of these systems. Further to this, direct supplementation of the respective sugars into diet may be required to discern an effect. Overall, there is an evident importance of sugars to the pathogenic lifestyle of EHEC through a variety of mechanisms.

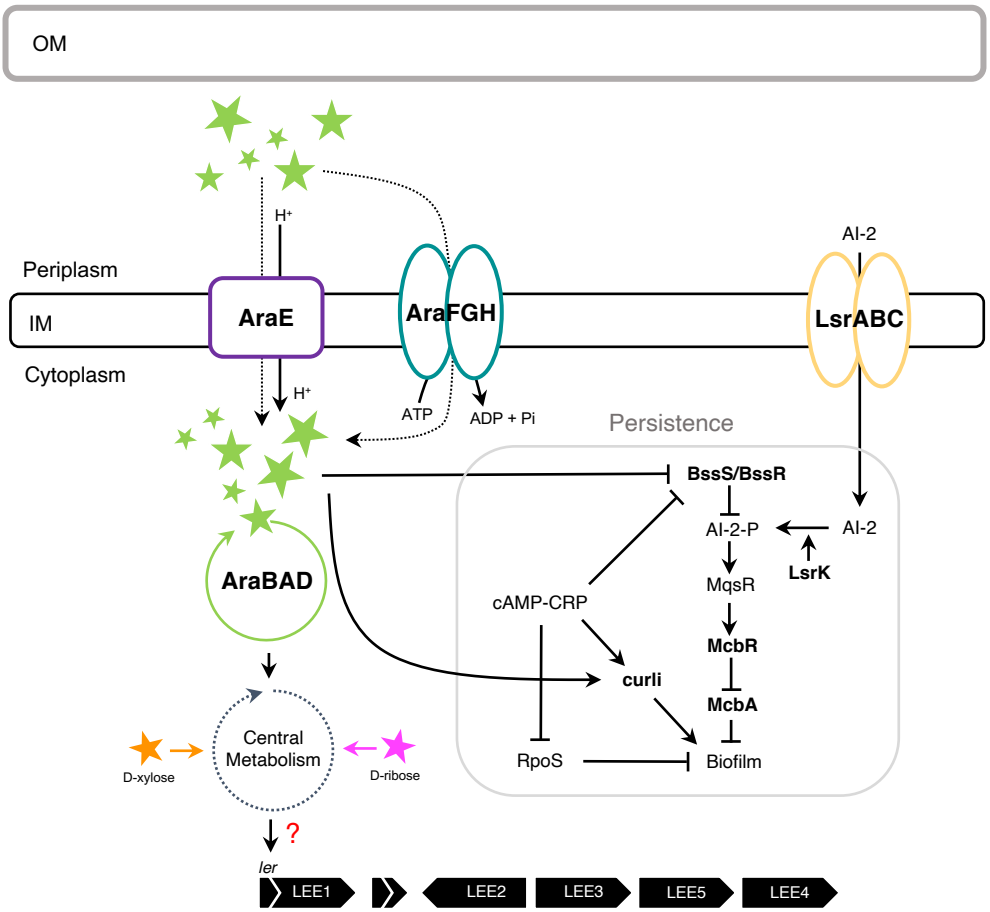


Figure 5-29. Mechanistic overview of enhanced virulence and persistence of EHEC by L-arabinose. In the cytoplasm, the metabolism of L-arabinose drives enhanced LEE expression via an unknown stimulus (?) that increases T3SS activity and adhesion to host cells. The aldopentose sugars, D-ribose and D-xylose also drive similar increases in LEE expression likely via their metabolism. L-arabinose also blocks the expression of biofilm repressor proteins, BssS and BssR. These proteins inhibit biofilm formation via a cascade of interactions (BssR/S – AI-2-P → MqsR → McbR – McbA). Additionally, L-arabinose enhances the expression of curli fimbriae genes and the Lsr AI-2 system, involved in biofilm formation. Those proteins highlighted in the grey box are those previously shown to be implicated in biofilm formation. Protein names in bold are those shown to be differentially expressed as part of this work.

5.23. Conclusions

The results in chapter 5 collectively highlight L-arabinose as dietary sugar with significant roles in enhancing the pathogenic lifestyle of EHEC. Through its metabolism and downstream products, L-arabinose acts as an important signal in controlling the expression of the LEE. By positively regulating the LEE and enhancing its expression, L-arabinose increases T3SS activity and colonisation of host cells. Furthermore, during *in vivo* infection, the inability to metabolise L-arabinose via *araBAD* results in a significant colonisation defect.

In addition to its effects on virulence, L-arabinose promotes biofilm formation through the downregulation of known repressors as well as the upregulation of genes important for initial biofilm attachment, maturation, and signalling. L-arabinose therefore represents a nutrient with possible roles in the persistence of EHEC in both the environment and within the host. These findings provide a powerful insight into the various implications dietary sugars have on pathogenesis, with a potential importance in the development of novel treatment strategies for EHEC infection via dietary intervention.

6. Final discussion and outlook

In vivo RNA-seq is a powerful tool for identifying novel candidate genes important for pathogenesis of the host. These candidates must then be investigated to gain a functional understanding of their roles during infection. In this work, characterisation of *C. rodentium* genes previously shown to be significantly upregulated during infection of the murine host (Connolly et al., 2018) led to the identification of a locus that encoded an ABC transport system and catabolic enzymes specific for D-ribulose. This was the first report of a system specific for D-ribulose amongst enteric pathogens that is absolutely essential for growth on the sugar.

A search for homologs revealed a system (Z0415-19) in EHEC that was also potentially specific for D-ribulose. In EHEC, the locus did, however, display key differences, mainly the absence of associated enzymes, but also in the arrangement of genes across which the transporter is encoded. Frequently associated with virulence traits (Jiang et al., 2021; Yang et al., 2023), the locus was located on a genomic OI in EDL933. Phylogenetic analysis of locus carriage and its association with distinct pathotypes suggested the locus to be strongly associated with enteric pathotypes, and predominantly EHEC strains. Despite displaying little expression under T3SS inducing conditions (i.e., growth in MEM-HEPES), the locus was significantly expressed in response to supplementation with L-arabinose. Through a multidisciplinary approach, L-arabinose was demonstrated to be of importance to EHEC, conferring several benefits that contributed to fitness by (a) acting as a source of nutrition, (b) aiding competition, (c) regulating the virulence arsenal and (d) promoting potential strategies of pathogen persistence through enhanced biofilm formation. A strong association between carriage of the locus and the LEE in EHEC strains further linked L-arabinose to virulence, and the sugar was shown to significantly enhance expression of the LEE in EHEC. Mechanistically, enhancement via L-arabinose occurs through metabolism of the sugar and the deletion of metabolic genes encoded by *araBAD* reduced the enhanced LEE expression conferred by L-arabinose. *In vivo*, deletion of *araBAD* led to increased pathogen clearance under both competitive and non-competitive conditions, supporting the role of the sugar and its metabolism for fitness during infection.

Though this work provides significant details into the role of aldopentose sugars on the fitness of A/E pathogens, through both their transport and metabolism, a number of questions are still outstanding. How L-arabinose confers its effect via metabolism specifically, still remains unknown. In terms of regulation, work has traditionally favoured the view that environmental stimuli control LEE expression through TCS and regulators that directly interact with said substrate (Njoroge & Sperandio, 2012; Pacheco et al., 2012; Rowley et al., 2020). The involvement of metabolism represents a unique slant on virulence regulation and there is substantial evidence to show that the metabolism of imported nutrients allows EHEC to fine-tune T3SS expression and respond to their environment (Connolly et al., 2018). Furthermore, modification of imported nutrients has also been demonstrated to have roles in LEE regulation. For example, host-derived arachidonic acid is processed to its acyl-CoA form and through FadR represses the LEE (Ellermann et al., 2021). Rather than being modified, L-arabinose is metabolised, and it is therefore more likely that its effect on the LEE is conferred by a single or mixture of downstream metabolites (via the aggregation of multiple aspects of the cells metabolic state) interacting either with novel or already known transcriptional regulators of the LEE. Interestingly the responsible regulator could be an enhancer and positively correlate with the effect of L-arabinose, or instead could be a suppressor of LEE expression that is somehow repressed by L-arabinose. Additional complexity is added to the regulation of the LEE via L-arabinose metabolism due the sugar feeding into central metabolic pathways that yield numerous substrates that may impact LEE expression, with potential synergistic or antagonistic effects. However, there are several instances whereby the mechanism of LEE regulation by nutrients remain incomplete, including how 1,2-propanediol-derived propionate enhances LEE expression (Connolly et al., 2018) and what the ligand is for the LEE transcriptional activator, YhaJ (Connolly et al., 2019). This incompleteness also extends to core regulators such as for GrlRA and can be attributed to the intricate dynamics of LEE regulation (Lara-Ochoa et al., 2023). Future work must subsequently seek the stimuli downstream of L-arabinose metabolism that drive enhanced LEE expression to then aid identification of regulatory factors responsible.

Additionally, there exists a significant lack of information regarding sugar concentrations, such as L-arabinose, within the human gut. This scarcity of data poses a considerable challenge in determining physiologically relevant concentrations for use

in *in vitro* experiments, and perhaps provides explanation for a lack of phenotype for Δ*rbl* and Δ*rbs* mutants *in vivo*. Considering dietary sources, L-arabinose is likely derived from ingested plant material and not from host mucus, as seen for several other sugars (Carlson-Banning & Sperandio, 2016). Interestingly, plant material provides the host with a source of fibre which has been extensively described and is generally thought to be of benefit by reducing the risk of diabetes, irritable bowel disease and colorectal cancers (Sauvaitre et al., 2021). However, the abundance of L-arabinose in plant material, which has been shown to promote colonisation and virulence of enteric pathogens, seems somewhat contradictory. That being said, butyrate, which is a by-product of intestinal fibre fermentation, promotes T3SS expression (Nakanishi et al., 2009; Takao et al., 2014), increases Stx binding to host cells and depletes commensal *E. coli* in the gut (Zumbrun et al., 2013). Therefore, whilst having known benefits to gut health, butyrate increases host susceptibility to EHEC infection.

Regulation of the LEE by nutrients is typically interlinked by the overcoming of CR and establishment of a novel niche. Due to the high rate of competition for nutrients in the gut, enteric pathogens such as EHEC must acquire ways to effectively scavenge them, particularly if they are to be metabolised. The acquisition of novel and/or additional transport systems such as the Rbl and Z0415-19 systems identified in this work provide a way to do so. Despite high similarity in sequence, architecture, and predicted function, the Rbl and Z0415-19 systems possess different specificities, demonstrating the exceptional fine-tuning these transporters have to their substrates. Differences in the substrate specificity is probably reflective of contrasting host conditions and the nutrients that are present, with specificities diverging relative to their different environments. For example, the diet of mice may offer more of nutrient A than it does of nutrient B, which could be vice versa in the human gut. Irrespectively, the presence of responsive transport systems is indicative that the specific substrate is available in the environment and has some importance to the pathogen.

The role of plant material and dietary fibre in enteric infection is important to consider, especially amongst those populations at greater risk of EHEC infection (Zumbrun et al., 2013), and as the west transitions to eating more plant-based diets (e.g., through veganism and sustainable practises). Creating further complications, EHEC is able to

colonise plants (Crozier et al., 2021), which can also act as a vehicle of infection to human hosts (Croxen et al., 2013). However, differences in diet, gut microbiota composition and genetics across individuals likely affects the outcome of nutrients on infection, with severity probably varying. This does highlight the possible role of dietary intervention as a means of not only decreasing infection risk but also as future treatment if stimulatory nutrients of virulence can be limited or their effects mitigated (reviewed in detail by Sauvaitre et al., 2021). For those sugars, derived from the diet, modification will be easier than those sourced from the host themselves (i.e., in mucus). However, this still poses a very complex situation as changes to diet often correspond with shifts in the microbiome, which similarly alter the risk of infection, particularly if beneficial microbes are lost (Forgie et al., 2019; Makki et al., 2018). The feasibility of dietary intervention is one which still requires thoughtful consideration.

7. References

Aidelberg, G., Towbin, B. D., Rothschild, D., Dekel, E., Bren, A., & Alon, U. (2014). Hierarchy of non-glucose sugars in *Escherichia coli*. *BMC Systems Biology*, *8*(1), 1–12

Akhtar, A. A., & Turner, D. P. (2022). The role of bacterial ATP-binding cassette (ABC) transporters in pathogenesis and virulence: Therapeutic and vaccine potential. *Microbial Pathogenesis*, *171*

Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P., Keller, K., Dubchak, I. L., & Arkin, A. P. (2005). The MicrobesOnline Web site for comparative genomics. *Genome Research*, *15*(7), 1015-1022

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology 2019 37:4*, *37*(4), 420–423

Alsharif, G., Ahmad, S., Islam, M. S., Shah, R., Busby, S. J., & Krachler, A. M. (2015). Host attachment and fluid shear are integrated into a mechanical signal regulating virulence in *Escherichia coli* O157:H7. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5503–5508

Ammar, E. M., Wang, X., & Rao, C. V. (2018). Regulation of metabolism in *Escherichia coli* during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. *Scientific Reports 2018 8:1*, *8*(1), 1–11

Anzai, T., Imamura, S., Ishihama, A., & Shimada, T. (2020). Expanded roles of pyruvate-sensing pdhr in transcription regulation of the *Escherichia coli* K-12 genome: Fatty acid catabolism and cell motility. *Microbial Genomics*, *6*(10), 1–14

Bansal, T., Jesudhasan, P., Pillai, S., Wood, T. K., & Jayaraman, A. (2008). Temporal regulation of enterohemorrhagic *Escherichia coli* virulence mediated by autoinducer-2. *Applied Microbiology and Biotechnology*, *78*(5), 811–819

Barba, J., Bustamante, V. H., Flores-Valdez, M. A., Deng, W., Finlay, B. B., & Puente, J. L. (2005). A positive regulatory loop controls expression of the locus of enterocyte effacement-encoded regulators Ler and GrlA. *Journal of Bacteriology*, *187*(23), 7918–7930

Battesti, A., Majdalani, N., & Gottesman, S. (2011). The RpoS-mediated general stress response in *Escherichia coli*. *Annual Review of Microbiology*, 65, 189-213.

Baümler, A. J., & Sperandio, V. (2016). Interactions between the microbiota and pathogenic bacteria in the gut. *Nature 2016 535:7610*, *535*(7610), 85–93

Beckwith, J. R. (1967). Regulation of the *lac* operon. *Science*, *156*(3775), 597–604

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*. *Journal of Bacteriology*, *187*(1), 349-357

Bhat, M. I., Sowmya, K., Kapila, S., & Kapila, R. (2019). *Escherichia coli* K12: An evolving opportunistic commensal gut microbe distorts barrier integrity in human intestinal cells. *Microbial Pathogenesis*, *133*, 103545

Bhatt, S., Egan, M., Ramirez, J., Xander, C., Jenkins, V., Muche, S., El-Fenej, J., Palmer, J., Mason, E., Storm, E., & Buerkert, T. (2017). Hfq and three Hfq-dependent small regulatory RNAs—MgrR, RyhB and McaS—coregulate the locus of enterocyte effacement in enteropathogenic *Escherichia coli*. *Pathogens and Disease*, *75*(1)

Bhatt, S., Romeo, T., & Kalman, D. (2011). Honing the message: Post-transcriptional and post-translational control in attaching and effacing pathogens. *Trends in Microbiology*, *19*(5), 217–224

Bhoite, S., van Gerven, N., Chapman, M. R., & Remaut, H. (2019). Curli biogenesis: Bacterial amyloid assembly by the type VIII secretion pathway. *EcoSal Plus*, *8*(2)

Bleich, A., Sundberg, J. P., Smoczek, A., Von Wasielewski, R., De Buhr, M. F., Janus, L. M., Julga, G., Ukena, S. N., Hedrich, H. J., & Gunzer, F. (2008). Sensitivity to *Escherichia coli* Nissle 1917 in mice is dependent on environment and genetic background. *International Journal of Experimental Pathology*, *89*(1), 45–54

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *ELife*, *4*. https://doi.org/10.7554/ELIFE.05826

Bohnhoff, M., Drake, B. L., & Miller, C. P. (1954). Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* Infection. *Proc Soc Exp Biol Med*, *86*(1), 132–137

Brockhurst, M. A., Harrison, E., Hall, J. P. J., Richards, T., McNally, A., & MacLean, C. (2019). The ecology and evolution of pangenomes. *Current Biology : CB*, *29*(20), R1094–R1103

Burland, V., Shao, Y., Perna, N. T., Plunkett, G., Sofia, H. J., & Blattner, F. R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Research*, *26*(18), 4196–4204

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., … Ioannidis, Y. E. (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, *47*(D1), D520–D528

Bustamante, V. H., Villalba, M. I., García-Angulo, V. A., Vázquez, A., Martínez, L. C., Jiménez, R., & Puente, J. L. (2011). PerC and GrlA independently regulate Ler expression in enteropathogenic *Escherichia coli*. *Molecular Microbiology*, *82*(2), 398–415

Byrne, L., Dallman, T. J., Adams, N., Mikhail, A. F. W., McCarthy, N., & Jenkins, C. (2018). Highly pathogenic clone of Shiga toxin–producing *Escherichia coli* O157:H7, England and Wales. *Emerging Infectious Diseases*, *24*(12), 2303-2308.

Bzdzion, L., Krezel, H., Wrzeszcz, K., Grzegorek, I., Nowinska, K., Chodaczek, G., & Swietnicki, W. (2017). Design of small molecule inhibitors of type III secretion system ATPase EscN from enteropathogenic *Escherichia coli*. *Acta Biochimica Polonica*, *64*(1), 49–63

Cameron, E. A., Curtis, M. M., Kumar, A., Dunny, G. M., & Sperandio, V. (2018). Microbiota and pathogen proteases modulate type III secretion activity in enterohemorrhagic *Escherichia coli*. *MBio*, *9*(6). https://doi.org/10.1128/MBIO.02204-18

Cameron, E. A., Sperandio, V., & Dunny, G. M. (2019). *Enterococcus faecalis* enhances expression and activity of the enterohemorrhagic *Escherichia coli* type III secretion system. *MBio*, *10*(6). https://doi.org/10.1128/MBIO.02547-19

Carlson-Banning, K. M., & Sperandio, V. (2016). Catabolite and oxygen regulation of enterohemorrhagic *Escherichia coli* virulence. *MBio*, *7*(6).

Chang, D. E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J., Anderson, A. B., Grissom, J. E., Laux, D. C., Cohen, P. S., & Conway, T. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7427-7432

Chen, G. Y., Thorup, N. R., Miller, A. J., Li, Y. C., & Ayres, J. S. (2023). Cooperation between physiological defenses and immune resistance produces asymptomatic carriage of a lethal bacterial pathogen. *Science Advances*, *9*(25), eadg8719

Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S. J., Webb, C., Wu, D., Varghese, N. J., Reddy, T. B. K., Mukherjee, S., Ovchinnikova, G., Nolan, M., Seshadri, R., Roux, S., Visel, A., Woyke, T., … Ivanova, N. N. (2023). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Research*, *51*(D1), D723–D732

Choudhury, D., & Saini, S. (2019). Evolution of *Escherichia coli* in different carbon environments for 2,000 generations. *Journal of Evolutionary Biology*, *32*(12), 1331–1341

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

Clermont, O., Condamine, B., Dion, S., Gordon, D. M., & Denamur, E. (2021). The E phylogroup of *Escherichia coli* is highly diverse and mimics the whole *E. coli* species population structure. *Environmental Microbiology*, *23*(11), 7139–7151

Collins, J. W., Keeney, K. M., Crepin, V. F., Rathinam, V. A. K., Fitzgerald, K. A., Finlay, B. B., & Frankel, G. (2014). *Citrobacter rodentium*: Infection, inflammation and the microbiota. *Nature Reviews Microbiology 2014 12:9*, *12*(9), 612–623

Connolly, J. P. R., Brett Finlay, B., & Roe, A. J. (2015). From ingestion to colonization: The influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic *Escherichia coli*. *Frontiers in Microbiology*, *6*(568)

Connolly, J. P. R., Gabrielsen, M., Goldstone, R. J., Grinter, R., Wang, D., Cogdell, R. J., Walker, D., Smith, D. G. E., & Roe, A. J. (2016). A highly conserved bacterial D-serine uptake system links host metabolism and virulence. *PLoS Pathogens*, *12*(1)

Connolly, J. P. R., Goldstone, R. J., Burgess, K., Cogdell, R. J., Beatson, S. A., Vollmer, W., Smith, D. G. E., & Roe, A. J. (2014). The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *The ISME Journal 2015 9:4*, *9*(4), 1039–1051

Connolly, J. P. R., O'Boyle, N., Turner, N. C. A., Browning, D. F., & Roe, A. J. (2019). Distinct intraspecies virulence mechanisms regulated by a conserved transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19695–19704

Connolly, J. P. R., Slater, S. L., O'Boyle, N., Goldstone, R. J., Crepin, V. F., Gallego, D. R., Herzyk, P., Smith, D. G. E., Douce, G. R., Frankel, G., & Roe, A. J. (2018). Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nature Communications 2018 9:1*, *9*(1), 1–14

Conway, T., & Cohen, P. S. (2015). Commensal and pathogenic *Escherichia coli* metabolism in the gut . *Microbiology Spectrum*, *3*(3)

Cooper, K. K., Mandrell, R. E., Louie, J. W., Korlach, J., Clark, T. A., Parker, C. T., Huynh, S., Chain, P. S., Ahmed, S., & Carter, M. Q. (2014). Comparative genomics of enterohemorrhagic *Escherichia coli* O145: H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157: H7. *BMC Genomics*, *15*(1), 1–17

Correa-Martinez, C. L., Leopold, S. R., Kö, R., Kossow, A., Bauwens, A., & Mellmann, A. (2022). Enterohemorrhagic *E. coli* (EHEC): Environmental-vehicle-human interface. https://doi.org/10.1007/978-3-030-85877-3_9-1

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology 2015 13:6*, *13*(6), 343–359

Cramer, J. P. (2014). Enterohemorrhagic *Escherichia coli* (EHEC): Hemorrhagic colitis and hemolytic uremic syndrome (HUS). *Emerging Infectious Diseases: Clinical Case Studies*, 213–227

Crepin, V. F., Collins, J. W., Habibzay, M., & Frankel, G. (2016). *Citrobacter rodentium* mouse model of bacterial infection. *Nature Protocols 2016 11:10*, *11*(10), 1851–1876

Crepin, V. F., Shaw, R., Abe, C. M., Knutton, S., & Frankel, G. (2005). Polarity of enteropathogenic *Escherichia coli* EspA filament assembly and protein secretion. *Journal of Bacteriology*, *187*(8), 2881–2889

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research*, *14*(6), 1188–1190

Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., & Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clinical Microbiology Reviews*, *26*(4), 822–880

Crozier, L., Marshall, J., Holmes, A., Wright, K. M., Rossez, Y., Merget, B., Humphris, S., Toth, I., Jackson, R. W., & Holden, N. J. (2021). The role of l-arabinose metabolism for *Escherichia coli* O157:H7 in edible plants. *Microbiology*, *167*(7), 1070. https://doi.org/10.1099/MIC.0.001070

Curtis, M. M., Hu, Z., Klimko, C., Narayanan, S., Deberardinis, R., & Sperandio, V. (2014). The gut commensal *Bacteroides thetaiotaomicron* exacerbates enteric infection through modification of the metabolic landscape. *Cell Host and Microbe*, *16*, 759–769

Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., … Gilbert, H. J. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature*, *517*(7533), 165-169

Dahan, S., Wiles, S., La Ragione, R. M., Best, A., Woodward, M. J., Stevens, M. P., Shaw, R. K., Chong, Y., Knutton, S., Phillips, A., & Frankel, G. (2005). EspJ is a prophage-carried type III effector protein of attaching and effacing pathogens that modulates infection dynamics. *Infection and Immunity*, *73*(2), 679–686

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–6645

Dean, P., & Kenny, B. (2009). The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. *Current Opinion in Microbiology*, *12*(1–3), 101-109

Denamur, E., Clermont, O., Bonacorsi, S., & Gordon, D. (2020). The population genetics of pathogenic *Escherichia coli*. *Nature Reviews Microbiology 2020 19:1*, *19*(1), 37–54

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., Strynadka, N. C. J., & Finlay, B. B. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology 2017 15:6*, *15*(6), 323–337

Deng, W., Yu, H. B., De Hoog, C. L., Stoynov, N., Li, Y., Foster, L. J., & Finlay, B. B. (2012). Quantitative proteomic analysis of type III secretome of enteropathogenic *Escherichia coli* reveals an expanded effector repertoire for attaching/effacing bacterial pathogens. *Molecular & Cellular Proteomics : MCP*, *11*(9), 692-709

Desvaux, M., Dalmasso, G., Beyrouthy, R., Barnich, N., Delmas, J., & Bonnet, R. (2020). Pathogenicity factors of genomic islands in intestinal and extraintestinal *Escherichia coli*. *Frontiers in Microbiology*, *11*, 2065

Dhiman, A., & Schleif, R. (2000). Recognition of overlapping nucleotides by AraC and the sigma subunit of RNA polymerase. *Journal of Bacteriology*, *182*(18), 5076-5081

Domka, J., Lee, J., & Wood, T. K. (2006). YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signalling. *Applied and Environmental Microbiology*, *72*(4), 2449–2459

Du, J., Reeves, A. Z., Klein, J. A., Twedt, D. J., Knodler, L. A., & Lesser, C. F. (2016). The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4794–4799

Dunn, T. M., & Schleif, R. (1984). Deletion analysis of the *Escherichia coli* ara PC and PBAD promoters. *Journal of Molecular Biology*, *180*(1), 201–204

Durack, J., & Lynch, S. V. (2019). The gut microbiome: Relationships with disease and opportunities for therapy. *The Journal of Experimental Medicine*, *216*(1), 20-40

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797

Egan, M., Critelli, B., Cleary, S. P., Marino, M., Upreti, C., Kalman, D., & Bhatt, S. (2019). Transcriptional and posttranscriptional regulation of the locus of enterocyte effacement in *Escherichia albertii*. *Microbial Pathogenesis*, *135*, 103643

Ellermann, M., Jimenez, A. G., Pifer, R., Ruiz, N., & Sperandio, V. (2021). The canonical long-chain fatty acid sensing machinery processes arachidonic acid to inhibit virulence in enterohemorrhagic *Escherichia coli*. *MBio*, *12*(1), 1–17

Ellermann, M., Pacheco, A. R., Jimenez, A. G., Russell, R. M., Cuesta, S., Kumar, A., Zhu, W., Vale, G., Martin, S. A., Raj, P., McDonald, J. G., Winter, S. E., & Sperandio, V. (2020). Endocannabinoids inhibit the induction of virulence in enteric pathogens. *Cell*, *183*(3), 650. https://doi.org/10.1016/J.CELL.2020.09.022

Elliott, S. J., Sperandio, V., Giron, J. A., Shin, S., Mellies, J. L., Wainwright, L., Hutcheson, S. W., McDaniel, T. K., & Kaper, J. B. (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*. *Infection and Immunity*, *68*(11), 6115-6126

Fabich, A. J., Jones, S. A., Chowdhury, F. Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J. W., Hightower, G. A., Smith, J. T., Autieri, S. M., Leatham, M. P., Lins, J. J., Allen, R. L., Laux, D. C., Cohen, P. S., & Conway, T. (2008). Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infection and Immunity*, *76*(3), 1143–1152

Feng, G., Flanagan, B. M., Mikkelsen, D., Williams, B. A., Yu, W., Gilbert, R. G., & Gidley, M. J. (2018). Mechanisms of utilisation of arabinoxylans by a porcine faecal inoculum: competition and co-operation. *Scientific Reports 2018 8:1*, *8*(1), 1–11

Filloux, A. (2022). Bacterial protein secretion systems: Game of types. *Microbiology 168*(5), 001193

Flockhart, A. F., Tree, J. J., Xu, X., Karpiyevich, M., McAteer, S. P., Rosenblum, R., Shaw, D. J., Low, C. J., Best, A., Gannon, V., Laing, C., Murphy, K. C., Leong, J. M., Schneiders, T., La Ragione, R., & Gally, D. L. (2012). Identification of a novel prophage regulator in *Escherichia coli* controlling the expression of type III secretion. *Molecular Microbiology*, *83*(1), 208–223

Forgie, A. J., Fouhse, J. M., & Willing, B. P. (2019). Diet-microbe-host interactions that affect gut mucosal integrity and infection resistance. *Frontiers in Immunology*, *10*, 448287

Foster-Nyarko, E., & Pallen, M. J. (2022). The microbial ecology of *Escherichia coli* in the vertebrate gut. *FEMS Microbiology Reviews*, *46*(3)

Frankel, G., Berger, C. N., Crepin, V. F., Baruch, K., Mousnier, A., & Rosenshine, I. (2012). EspZ of enteropathogenic and enterohemorrhagic *Escherichia coli* regulates type III secretion system protein translocation. *MBio*, *3*(5), e00317-12

Fraser, M. E., Fujinaga, M., Cherney, M. M., Melton-Celsa, A. R., Twiddy, E. M., O'Brien, A. D., & James, M. N. G. (2004). Structure of Shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7. *Journal of Biological Chemistry*, *279*(26), 27511–27517.

Freter, R., Brickner, H., Botney, M., Cleven, D., & Aranki, A. (1983). Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infection and Immunity*, *39*(2), 676–685

Fritz, G., Megerle, J. A., Westermayer, S. A., Brick, D., Heermann, R., Jung, K., Rädler, J. O., & Gerland, U. (2014). Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. coli*. *PLOS ONE*, *9*(2), e89532

Fukui, N., Oshima, T., Ueda, T., Ogasawara, N., & Tobe, T. (2016). Gene activation through the modulation of nucleoid structures by a horizontally transferred regulator, Pch, in enterohemorrhagic *Escherichia coli*. *PLOS ONE*, *11*(2), e0149718

Gardette, M., Le Hello, S., Mariani-Kurkdjian, P., Fabre, L., Gravey, F., Garrivier, A., Loukiadis, E., & Jubelin, G. (2019). Identification and prevalence of in vivo-induced genes in enterohaemorrhagic *Escherichia coli*. *Virulence*, *10*(1), 180–193

Gatsios, A., Kim, C. S., & Crawford, J. M. (2021). *Escherichia coli* small molecule metabolism at the host–microorganism interface. *Nature Chemical Biology 2021 17:10*, *17*(10), 1016–1026

Gaytán, M. O., Martínez-Santos, V. I., Soto, E., & González-Pedrajo, B. (2016). Type three secretion system in attaching and effacing pathogens. *Frontiers in Cellular and Infection Microbiology*, *6*, 129

Ge, S. X., Jung, D., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629

Gelalcha, B. D., Brown, S. M., Crocker, H. E., Agga, G. E., & Kerro Dego, O. (2022). Regulation mechanisms of virulence genes in enterohemorrhagic *Escherichia coli*. *Foodborne Pathogens and Disease*, *19*(9), 598–612

Geurtsen, J., de Been, M., Weerdenburg, E., Zomer, A., McNally, A., & Poolman, J. (2022). Genomics and pathotypes of the many faces of *Escherichia coli*. *FEMS Microbiology Reviews*, *46*(6)

Goldwater, P. N., & Bettelheim, K. A. (2012). Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Medicine*, *10*(1), 1–8

González Barrios, A. F., Zuo, R., Hashimoto, Y., Yang, L., Bentley, W. E., & Wood, T. K. (2006). Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). *Journal of Bacteriology*, *188*(1), 305–316

Gonzalez-Alba, J. M., Baquero, F., Cantón, R., & Galán, J. C. (2019). Stratified reconstruction of ancestral *Escherichia coli* diversification. *BMC Genomics*, *20*(1), 1–15

Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology 2008 6:8*, *6*(8), 613–624

Green, E. R., & Mecsas, J. (2016). Bacterial secretion systems – An overview. *Microbiology Spectrum*, *4*(1)

Groff, D., Benke, P. I., Batth, T. S., Bokinsky, G., Petzold, C. J., Adams, P. D., & Keasling, J. D. (2012). Supplementation of intracellular XylR leads to coutilization of hemicellulose sugars. *Applied and Environmental Microbiology*, *78*(7), 2221–2229

Gronbach, K., Eberle, U., Müller, M., Ölschläger, T. A., Dobrindt, U., Leithäuser, F., Niess, J. H., Döring, G., Reimann, J., Autenrieth, I. B., & Frick, J. S. (2010). Safety of probiotic *Escherichia coli* strain Nissle 1917 depends on intestinal microbiota and adaptive immunity of the host. *Infection and Immunity*, *78*(7), 3036-3046

Gruber, C. C., & Sperandio, V. (2015). Global analysis of posttranscriptional regulation by GlmY and GlmZ in enterohemorrhagic *Escherichia coli* O157:H7. *Infection and Immunity*, *83*(4), 1286–1295

Guo, X. K., Wang, J., van Hensbergen, V. P., Liu, J., Xu, H., & Hu, X. (2023). Interactions between host and intestinal crypt-resided biofilms are controlled by epithelial fucosylation. *Cell Reports*, *42*(7). https://doi.org/10.1016/j.celrep.2023.112754

Hallgren, J., Tsirigos, K. D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*, 2022.04.08.487609. https://doi.org/10.1101/2022.04.08.487609

Hansen, A. M., & Kaper, J. B. (2009). Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Molecular Microbiology*, *73*(3), 446–465

Hasona, A., Kim, Y., Healy, F. G., Ingram, L. O., & Shanmugam, K. T. (2004). Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of *Escherichia coli* on xylose. *Journal of Bacteriology*, *186*(22), 7593–7600

Hering, N. A., Richter, J. F., Fromm, A., Wieser, A., Hartmann, S., Günzel, D., Bücker, R., Fromm, M., Schulzke, J. D., & Troeger, H. (2014). TcpC protein from *E. coli* Nissle improves epithelial barrier function involving PKCζ and ERK1/2 signaling in HT-29/B6 cells. *Mucosal Immunology*, *7*(2), 369–378

Hernandez-Doria, J. D., & Sperandio, V. (2013). Nutrient and chemical sensing by intestinal pathogens. *Microbes and Infection*, *15*(12), 759–764

Heuel, H., Shakeri-Garakani, A., Turgut, S., & Lengeler, J. W. (1998). Genes for D-arabinitol and ribitol catabolism from *Klebsiella pneumoniae*. *Microbiology*, *144*(6), 1631–1639

Horler, R. S. P., Müller, A., Williamson, D. C., Potts, J. R., Wilson, K. S., & Thomas, G. H. (2009). Furanose-specific sugar transport: Characterization of a bacterial galactofuranose-binding protein. *Journal of Biological Chemistry*, *284*(45), 31156–31163

Hughes, D. T., Clarke, M. B., Yamamoto, K., Rasko, D. A., & Sperandio, V. (2009). The QseC adrenergic signaling cascade in enterohemorrhagic *E. coli* (EHEC). *PLOS Pathogens*, *5*(8), e1000553

Hughes, D. T., Terekhova, D. A., Liou, L., Hovde, C. J., Sahl, J. W., Patankar, A. V., Gonzalez, J. E., Edrington, T. S., Rasko, D. A., & Sperandio, V. (2010). Chemical sensing in mammalian host-bacterial commensal associations. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(21), 9831–9836

Islam, M. S., Bingle, L. E. H., Pallen, M. J., & Busby, S. J. W. (2011). Organization of the LEE1 operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Molecular Microbiology*, *79*(2), 468–483

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*. *Journal of Bacteriology*, *187*(12), 4086-4094

Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., & Sargent, F. (2004). Coordinating assembly and export of complex bacterial proteins. *The EMBO Journal*, *23*(20), 3962–3972

Jiang, L., Yang, W., Jiang, X., Yao, T., Wang, L., & Yang, B. (2021). Virulence-related O islands in enterohemorrhagic *Escherichia coli* O157:H7. *Gut Microbes*, *13*(1). https://doi.org/10.1080/19490976.2021.1992237

Jimenez, A. G., Ellermann, M., Abbott, W., & Sperandio, V. (2019). Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic *Escherichia coli* and *Citrobacter rodentium*. *Nature Microbiology 2019 5:2*, *5*(2), 368–378

John, M., Kudva, I. T., Griffin, R. W., Dodson, A. W., McManus, B., Krastins, B., Sarracino, D., Progulske-Fox, A., Hillman, J. D., Handfield, M., Tarr, P. I., & Calderwood, S. B. (2005). Use of in vivo-induced antigen technology for identification of *Escherichia coli* O157:H7 proteins expressed during human infection. *Infection and Immunity*, *73*(5), 2665–2679

Johnson, C. M., & Schleif, R. F. (1995). In vivo induction kinetics of the arabinose promoters in *Escherichia coli*. *Journal of Bacteriology*, *177*(12), 3438–3442

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240

Joseph, A., Cointe, A., Kurkdjian, P. M., Rafat, C., & Hertig, A. (2020). Shiga toxin-associated hemolytic uremic syndrome: A narrative review. *Toxins*, *12*(2). https://doi.org/10.3390/TOXINS12020067

Kakoullis, L., Papachristodoulou, E., Chra, P., & Panos, G. (2019). Shiga toxin-induced haemolytic uraemic syndrome and the role of antibiotics: a global overview. *Journal of Infection*, *79*(2), 75–94

Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, *30*(1), 42–46

Kaper, J. B., Nataro, J. P., & Mobley, H. L. T. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology 2004 2:2*, *2*(2), 123–140

Kaplan, S., Bren, A., Zaslaver, A., Dekel, E., & Alon, U. (2008). Diverse two-dimensional input functions control bacterial sugar genes. *Molecular Cell*, *29*(6), 786–792

Karavolos, M. H., Spencer, H., Bulmer, D. M., Thompson, A., Winzer, K., Williams, P., Hinton, J. C. D., & Khan, C. M. A. (2008). Adrenaline modulates the global transcriptional profile of *Salmonella* revealing a role in the antimicrobial peptide and oxidative stress resistance responses. *BMC Genomics*, *9*(1), 1–14

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., & Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, *20*(4), 1085–1093

Katsowich, N., Elbaz, N., Pal, R. R., Mills, E., Kobi, S., Kahan, T., & Rosenshine, I. (2017). Host cell attachment elicits posttranscriptional regulation in infecting enteropathogenic bacteria. *Science*, *355*(6326), 735–739

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols 2015 10:6*, *10*(6), 845–858

Kendall, M. M., Gruber, C. C., Parker, C. T., & 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, M. M., Gruber, C. C., Rasko, D. A., Hughes, D. T., & Sperandio, V. (2011). Hfq virulence regulation in enterohemorrhagic *Escherichia coli* O157:H7 Strain 86-24. *Journal of Bacteriology*, *193*(24), 6843–6851

Kenny, B. (2001). The enterohaemorrhagic *Escherichia coli* (serotype O157:H7) Tir molecule is not functionally interchangeable for its enteropathogenic *E. coli* (serotype O127:H6) homologue. *Cellular Microbiology*, *3*(8), 499–510

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D. J., Frey, E. A., & Finlay, B. B. (1997). Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell*, *91*(4), 511–520

Khan, I., Bai, Y., Zha, L., Ullah, N., Ullah, H., Shah, S. R. H., Sun, H., & Zhang, C. (2021). Mechanism of the gut microbiota colonization resistance and enteric pathogen infection. *Frontiers in Cellular and Infection Microbiology*, *11*, 716299

Kim, Y. G., Lee, J. H., Gwon, G., Kim, S. Il, Park, J. G., & Lee, J. (2016). Essential oils and eugenols inhibit biofilm formation and the virulence of *Escherichia coli* O157:H7. *Scientific Reports*, *6*, 36377

Koirala, S., Wang, X., & Rao, C. V. (2015). Reciprocal regulation of l-arabinose and d-xylose metabolism in *Escherichia coli*. *Journal of Bacteriology*, *198*(3), 386–393

Koli, P., Sudan, S., Fitzgerald, D., Adhya, S., & Kar, S. (2011). Conversion of commensal *Escherichia coli* K-12 to an invasive form via expression of a mutant histone-like protein. *MBio*, *2*(5), 12, e00182-11

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549

Kyle, J. L., Parker, C. T., Goudeau, D., & Brandl, M. T. (2010). Transcriptome analysis of Escherichia coli O157:H7 exposed to lysates of lettuce leaves. *Applied and Environmental Microbiology*, *76*(5), 1375–1387

Lagerstrom, K. M., & Hadly, E. A. (2023). Under-appreciated phylogroup diversity of *Escherichia coli* within and between animals at the urban-wildland interface. *Applied and Environmental Microbiology*, *89*(6). https://doi.org/10.1128/aem.00142-23

Lara-Ochoa, C., Huerta-Saquero, A., Medrano-López, A., Deng, W., Finlay, B. B., Martínez-Laguna, Y., & Puente, J. L. (2023). GrlR, a negative regulator in enteropathogenic *E. coli*, also represses the expression of LEE virulence genes independently of its interaction with its cognate partner GrlA. *Frontiers in Microbiology*, *14*, 279, 1063368

Le Bihan, G., Sicard, J. F., Garneau, P., Bernalier-Donadille, A., Gobert, A. P., Garrivier, A., Martin, C., Hay, A. G., Beaudry, F., Harel, J., & Jubelin, G. (2017). The NAG Sensor NagC regulates LEE gene expression and contributes to gut

colonization by *Escherichia coli* O157:H7. *Frontiers in Cellular and Infection Microbiology*, *7*, 134, 257203.

Lease, R. A., Smith, D., McDonough, K., & Belfort, M. (2004). The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *Journal of Bacteriology*, *186*(18), 6179–6185

Lee, D. J., Bingle, L. E., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J., & Hobman, J. L. (2009). Gene doctoring: A method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiology*, *9*(1), 1–14

Lee, N., Cielow, W., Martin, R., Hamilton, E., & Fowler, A. (1986). The organization of the araBAD operon of *Escherichia coli*. *Gene*, *47*, 23–244

Lee, S. M., Donaldson, G. P., Mikulski, Z., Boyajian, S., Ley, K., & Mazmanian, S. K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature 2013 501:7467*, *501*(7467), 426–429

Lee, L. V., Gerratana, B., & Cleland, W. W. (2001). Substrate specificity and kinetic mechanism of *Escherichia coli* ribulokinase. *Archives of Biochemistry and Biophysics*, *396*(2), 219–224

Leimbach, A., Hacker, J., & Dobrindt, U. (2013). E. coli as an all-rounder: The thin line between commensalism and pathogenicity. *Current Topics in Microbiology and Immunology*, *358*, 3–32

Liang, Q., & Vallance, B. A. (2021). What's for dinner? How *Citrobacter rodentium's* metabolism helps it thrive in the competitive gut. *Current Opinion in Microbiology*, *63*, 76–82

Lim, J. Y., Yoon, J. W., & Hovde, C. J. (2010). A Brief Overview of *Escherichia coli* O157:H7 and Its Plasmid O157. *Journal of Microbiology and Biotechnology*, *20*(1), 5-14

Litvak, Y., & Bäumler, A. J. (2019). The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. *PLoS Pathogens*, *15*(2), e1007563

Liu, C., Sun, D., Zhu, J., Liu, J., & Liu, W. (2020). The regulation of bacterial biofilm formation by cAMP-CRP: A mini-review. *Frontiers in Microbiology*, *11*, 802, 532558

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods*, *25*(4), 402–408

Lobell, R. B., & Schleif, R. F. (1990). DNA looping and unlooping by AraC protein. *Science*, *250*(4980), 528–532

Locher, K. P., Lee, A. T., & Rees, D. C. (2002). The *E. coli* BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science*, *296*(5570), 1091–1098

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21

Luzader, D. H., Clark, D. E., Gonyar, L. A., & Kendall, M. M. (2013). EutR is a direct regulator of genes that contribute to metabolism and virulence in enterohemorrhagic *Escherichia coli* O157: H7. *Journal of Bacteriology*, *195*(21), 4947–4953

Luzader, D. H., Willsey, G. G., Wargo, M. J., & Kendall, M. M. (2016). The type three secretion system 2-encoded regulator EtrB modulates enterohemorrhagic *Escherichia coli* virulence gene expression. *Infection and Immunity*, *84*(9), 2555–2565

Makino, F., Shen, D., Kajimura, N., Kawamoto, A., Pissaridou, P., Oswin, H., Pain, M., Murillo, I., Namba, K., & Blocker, A. J. (2016). The architecture of the cytoplasmic region of type III secretion systems. *Scientific Reports 2016 6:1*, *6*(1), 1–12

Makki, K., Deehan, E. C., Walter, J., & Bäckhed, F. (2018). The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host and Microbe*, *23*(6), 705–715

Maqbool, A., Horler, R. S. P., Muller, A., Wilkinson, A. J., Wilson, K. S., & Thomas, G. H. (2015). The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochemical Society Transactions*, *43*(5), 1011–1017

Martens, E. C., Chiang, H. C., & Gordon, J. I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host & Microbe*, *4*(5), 447–457

Martins, F. H., Kumar, A., Abe, C. M., Carvalho, E., Nishiyama-Jr, M., Xing, C., Sperandio, V., & Elias, W. P. (2020). Espfu-mediated actin assembly enhances enteropathogenic *Escherichia coli* adherence and activates host cell inflammatory signaling pathways. *MBio*, *11*(2), e00617-20

Martinson, J. N. V., & Walk, S. T. (2020). *Escherichia coli* residency in the gut of healthy human adults. *EcoSal Plus*, *9*(1)

Mayer, C., & Boos, W. (2005). Hexose/pentose and hexitol/pentitol metabolism. *EcoSal Plus*, *1*(2). doi.10.1128/ecosalplus.3.4.1

Mellmann, A., Harmsen, D., Cummings, C. A., Zentz, E. B., Leopold, S. R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S. F., Henkhaus, J. K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P. M., Moore, R. L., Guenther, S., Rothberg, J. M., & Karch, H. (2012).Prospective genomic

characterisation of the German enterohaemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE*, *6*(7). https://doi.org/10.1371/JOURNAL.PONE.0022751

Menezes-Garcia, Z., Kumar, A., Zhu, W., Winter, S. E., & Sperandio, V. (2020). L-arginine sensing regulates virulence gene expression and disease progression in enteric pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(22), 12387–12393

Meredith, T. C., & Woodard, R. W. (2006). Characterization of *Escherichia coli* D-arabinose 5-phosphate isomerase encoded by *kpsF*: implications for group 2 capsule biosynthesis. *Biochemical Journal*, *395*, 427-432

Miyada, C. G., Stoltzfus, L., & Wilcox, G. (1984). Regulation of the *araC* gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on *araBAD* expression. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(13), 4120-4124

Morgan, J. K., Carroll, R. K., Harro, C. M., Vendura, K. W., Shaw, L. N., & Riordan, J. T. (2016). Global regulator of virulence A (GrvA) coordinates expression of discrete pathogenic mechanisms in enterohemorrhagic *Escherichia coli* through interactions with GadW-GadE. *Journal of Bacteriology*, *198*(3), 394–409

Mühlen, S., & Dersch, P. (2020). Treatment strategies for infections with Shiga toxin-producing *Escherichia coli*. *Frontiers in Cellular and Infection Microbiology*, *10*, 169. https://doi.org/10.3389/FCIMB.2020.00169

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), 521–530

Naylor, S. W., Low, J. C., Besser, T. E., Mahajan, A., Gunn, G. J., Pearce, M. C., McKendrick, I. J., Smith, D. G. E., & Gally, D. L. (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. *Infection and Immunity*, *71*(3), 1505–1512

Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C., Monack, D. M., & Sonnenburg, J. L. (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*, *502*(7469), 96-99

Njoroge, J., & Sperandio, V. (2012). Enterohemorrhagic *Escherichia coli* virulence regulation by two bacterial adrenergic kinases, QseC and QseE. *Infection and Immunity*, *80*(2), 688–703

Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G., & Sperandio, V. (2012). Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli*. *MBio*, *3*(5), e00280-12

Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L., & Chen, J. (2007). Crystal structure of a catalytic intermediate of the maltose transporter. *Nature 2007 450:7169*, *450*(7169), 515–521

Oliveira, R. A., Cabral, V., Torcato, I., & Xavier, K. B. (2023). Deciphering the quorum-sensing lexicon of the gut microbiota. *Cell Host & Microbe*, *31*(4), 500–512

Ortega, Á., Matilla, M. A., & Krell, T. (2022). The repertoire of solute-binding proteins of model bacteria reveals large differences in number, type, and ligand range. *Microbiology Spectrum*, *10*(5), e0205422

Pacheco, A. R., Munera, D., Waldor, M. K., Sperandio, V., & Ritchie, J. M. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature 2012 492:7427*, *492*(7427), 113–117

Pacheco, A. R., & Sperandio, V. (2015). Enteric pathogens exploit the microbiota-generated nutritional environment of the gut. *Microbiology Spectrum*, *3*(3). doi.10.1128/microbiolspec.MBP-0001-2014

Padavannil, A., Jobichen, C., Mills, E., Velazquez-Campoy, A., Li, M., Leung, K. Y., Mok, Y. K., Rosenshine, I., & Sivaraman, J. (2013). Structure of GrlR–GrlA complex that prevents GrlA activation of virulence genes. *Nature Communications 2013 4:1*, *4*(1), 1–10

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods 2017 14:4*, *14*(4), 417–419

Perna, N. T., Mayhew, G. F., Pósfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B., & Blattner, F. R. (1998). Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infection and Immunity*, *66*(8), 3810–3817

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature 2001 409:6819*, *409*(6819), 529–533

Persad, A. K., & LeJeune, J. T. (2014). Animal reservoirs of Shiga toxin-producing *Escherichia coli*. *Microbiology Spectrum*, *2*(4). doi.10.1128/microbiolspec.EHEC-0027-2014

Persat, A., Inclan, Y. F., Engel, J. N., Stone, H. A., & Gitai, Z. (2015). Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7563–7568

Pifer, R., Russell, R. M., Kumar, A., Curtis, M. M., & Sperandio, V. (2018). Redox, amino acid, and fatty acid metabolism intersect with bacterial virulence in the gut. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), E10712–E10719

Pilla, G., & Tang, C. M. (2018). Going around in circles: virulence plasmids in enteric pathogens. *Nature Reviews Microbiology 2018 16:8*, *16*(8), 484–495

Platenkamp, A., & Mellies, J. L. (2018). Environment controls LEE regulation in enteropathogenic *Escherichia coli*. *Frontiers in Microbiology*, *9*, 1694

Pokharel, P., Dhakal, S., & Dozois, C. M. (2023). The diversity of *Escherichia coli* pathotypes and vaccination strategies against this versatile bacterial pathogen. *Microorganisms*, *11*(2). https://doi.org/10.3390/MICROORGANISMS11020344

Pradhan, S., & Weiss, A. A. (2020). Probiotic properties of *Escherichia coli* nissle in human intestinal organoids. *MBio*, *11*(4), 1–16

Prager, R., Lang, C., Aurass, P., Fruth, A., Tietze, E., & Flieger, A. (2014). Two Novel EHEC/EAEC Hybrid Strains Isolated from Human Infections. *PLOS ONE*, *9*(4), e95379

Proença, J. T., Barral, D. C., & Gordo, I. (2017). Commensal-to-pathogen transition: One-single transposon insertion results in two pathoadaptive traits in *Escherichia coli* -macrophage interaction. *Scientific Reports 2017 7:1*, *7*(1), 1–12

Pruimboom-Brees, I. M., Morgan, T. W., Ackermann, M. R., Nystrom, E. D., Samuel, J. E., Cornick, N. A., & Moon, H. W. (2000). Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10325–10329

Rangarajan, A. A., & Schnetz, K. (2018). Interference of transcription across H-NS binding sites and repression by H-NS. *Molecular Microbiology*, *108*(3), 226–239

Reading, N. C., Torres, A. G., Kendall, M. M., Hughes, D. T., 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. *Journal of Bacteriology*, *189*(6), 2468-2476

Reeder', T., & Schleif1, R. (1991). Mapping, sequence, and apparent lack of function of araJ, a gene of the *Escherichia coli* arabinose regulon. *JOURNAL OF BACTERIOLOGY*, *173*(24), 7765–7771

Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K., & Whittam, T. S. (2000). Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature 2000 406:6791*, *406*(6791), 64–67

Riley, L. W. (2020). Distinguishing Pathovars from Nonpathovars: *Escherichia coli*. *Microbiology Spectrum*, *8*(4). https://doi.org/10.1128/MICROBIOLSPEC.AME-0014-2020

Riley, L. W., Remis, R. S., Helgerson, S. D., McGee, H. B., Wells, J. G., Davis, B. R., Hebert, R. J., Olcott, E. S., Johnson, L. M., Hargrett, N. T., Blake, P. A., & Cohen, M. L. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *The New England Journal of Medicine*, *308*(12), 681–685 https://doi.org/10.1056/NEJM198303243081203

Ritchie, J. M., & Waldor, M. K. (2005). The locus of enterocyte effacement-encoded effector proteins all promote enterohemorrhagic *Escherichia coli* pathogenicity in infant rabbits. *Infection and Immunity*, *73*(3), 1466–1474

Robinson, A. M., Gondalia, S. V., Karpe, A. V., Eri, R., Beale, D. J., Morrison, P. D., Palombo, E. A., & Nurgali, K. (2016). Fecal Microbiota and Metabolome in a Mouse Model of Spontaneous Chronic Colitis: Relevance to Human Inflammatory Bowel Disease. *Inflammatory Bowel Diseases*, *22*(12), 2767–2787

Roe, A.J., Tysall, S.W., Martin, J., Smith, D.G.E., & Gally, D.L. (2003). Heterogeneous surface expression of EspA translocon filaments by *Escherichia coli* O157:H7 is controlled at the surface level. *Infection and Immunity, 7*(10), 5900-5909

Rowley, C. A., Sauder, A. B., & Kendall, M. M. (2020). The ethanolamine-sensing transcription factor EutR promotes virulence and transmission during *Citrobacter rodentium* intestinal infection. *Infection and Immunity*, *88*(9), e00137-20

Ruano-Gallego, D., Sanchez-Garrido, J., Kozik, Z., Núñez-Berrueco, E., Cepeda-Molero, M., Mullineaux-Sanders, C., Clark, J. N. B., Slater, S. L., Wagner, N., Glegola-Madejska, I., Roumeliotis, T. I., Pupko, T., Fernández, L. Á., Rodríguez-Patón, A., Choudhary, J. S., & Frankel, G. (2021). Type III secretion system effectors form robust and flexible intracellular virulence networks. *Science*, *371*(6534), eabc9531

Ruddle, S. J., Massis, L. M., Cutter, A. C., & Monack, D. M. (2023). *Salmonella*-liberated dietary L-arabinose promotes expansion in superspreaders. *Cell Host and Microbe*, *31*(3), 405-417

Rumer, L., Jores, J., Kirsch, P., Cavignac, Y., Zehmke, K., & Wieler, L. H. (2003). Dissemination of pheU- and pheV-located genomic islands among enteropathogenic (EPEC) and enterohemorrhagic (EXEC) *E. coli* and their possible role in the horizontal transfer of the locus of enterocyte effacement (LEE). *International Journal of Medical Microbiology*, *292*(7–8), 463–475

Sal-Man, N., Setiaputra, D., Scholz, R., Deng, W., Yu, A. C. Y., Strynadka, N. C. J., & Finlay, B. B. (2013). EscE and EscG are cochaperones for the type III needle protein EscF of enteropathogenic *Escherichia coli*. *Journal of Bacteriology*, *195*(11), 2481–2489

Sanchez-Garrido, J., Alberdi, L., Chatterjee, S., Frankel, G., & Mullineaux-Sanders, C. (2021). Type III secretion system effector subnetworks elicit distinct host immune responses to infection. *Current Opinion in Microbiology*, *64*, 19–26

Sanchez-Garrido, J., Ruano-Gallego, D., Choudhary, J. S., & Frankel, G. (2022). The type III secretion system effector network hypothesis. *Trends in Microbiology*, *30*(6), 524–533

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

Sauder, A. B., & Kendall, M. M. (2018). After the fact(or): Posttranscriptional gene regulation in enterohemorrhagic *Escherichia coli* O157:H7. *Journal of Bacteriology*, *200*(19), e00228-18

Sauvaitre, T., Etienne-Mesmin, L., Sivignon, A., Mosoni, P., Courtin, C. M., Van De Wiele, T., & Blanquet-Diot, S. (2021). Tripartite relationship between gut microbiota, intestinal mucus and dietary fibers: towards preventive strategies against enteric infections. *FEMS Microbiology Reviews*, *45*(2), 1–36

Schleif, R. (2000). Regulation of the l-arabinose operon of *Escherichia coli*. *Trends in Genetics*, *16*(12), 559–565

Schraidt, O., & Marlovits, T. C. (2011). Three-dimensional model of *Salmonella's* needle complex at subnanometer resolution. *Science*, *331*(6021), 1192–1195

Schüller, S. (2011). Shiga toxin interaction with human intestinal epithelium. *Toxins*, *3*(6), 626-639

Schwalm, N. D., Townsend, G. E., & Groisman, E. A. (2016). Multiple signals govern utilization of a polysaccharide in the gut bacterium *Bacteroides thetaiotaomicron*. *MBio*, *7*(5), e01342-16

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLOS Biology*, *14*(8), e1002533

Serapio-Palacios, A., & Finlay, B. B. (2020). Dynamics of expression, secretion and translocation of type III effectors during enteropathogenic *Escherichia coli* infection. *Current Opinion in Microbiology*, *54*, 67–76

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., & Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols 2009 4:2*, *4*(2), 206–223

Sharma, V. K., & Bearson, B. L. (2013). Hha controls *Escherichia coli* O157:H7 biofilm formation by differential regulation of global transcriptional regulators FlhDC and CsgD. *Applied and Environmental Microbiology*, *79*(7), 2384–2396

Sharma, V. K., Carlson, S. A., & Casey, T. A. (2005). Hyperadherence of an *hha* mutant of *Escherichia coli* O157:H7 is correlated with enhanced expression of LEE-encoded adherence genes. *FEMS Microbiology Letters*, *243*(1), 189–196

Sharma, V. K., & Casey, T. A. (2014). Determining the relative contribution and hierarchy of *hha* and *qseBC* in the regulation of flagellar Motility of *Escherichia coli* O157:H7. *PLOS ONE*, *9*(1), e85866

Shealy, N. G., Yoo, W., & Byndloss, M. X. (2021). Colonization resistance: metabolic warfare as a strategy against pathogenic Enterobacteriaceae. *Current Opinion in Microbiology*, *64*, 82–90

Shimada, T., Fujita, N., Yamamoto, K., & Ishihama, A. (2011). Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS ONE*, *6*(6), 20081. https://doi.org/10.1371/JOURNAL.PONE.0020081

Shimada, T., Kori, A., & Ishihama, A. (2013). Involvement of the ribose operon repressor RbsR in regulation of purine nucleotide synthesis in *Escherichia coli*. *FEMS Microbiology Letters*, *344*(2), 159–165

Shin, M. (2017). The mechanism underlying Ler-mediated alleviation of gene repression by H-NS. *Biochemical and Biophysical Research Communications*, *483*(1), 392–396

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539–539

Sims, G. E., & Kim, S. H. (2011). Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(20), 8329–8334

Singh, C., Glaab, E., Linster, C. L., & Denu, J. M. (2017). Molecular identification of D-ribulokinase in budding yeast and mammals. *Journal of Biological Chemistry*, *292*(3), 1005–1028

Singh, S. S., & Grainger, D. C. (2013). H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. *PLOS Genetics*, *9*(6), e1003589

Sirisaengtaksin, N., Odem, M. A., Bosserman, R. E., Flores, E. M., & Krachler, A. M. (2020). The *E. coli* transcription factor GrlA is regulated by subcellular compartmentalization and activated in response to mechanical stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9519–9528

Slater, S. L., & Frankel, G. (2020). Advances and challenges in studying type III secretion effectors of attaching and effacing pathogens. *Frontiers in Cellular and Infection Microbiology*, *10*, 337. doi.10.3389/fcimb.2020.00337

Sommaruga, S., Gioia, L. De, Tortora, P., & Polissi, A. (2009). Structure prediction and functional analysis of KdsD, an enzyme involved in lipopolysaccharide biosynthesis. *Biochemical and Biophysical Research Communications*, *388*(2), 222–227

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C. H., Westover, B. P., Weatherford, J., Buhler, J. D., & Gordon, J. I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science*, *307*(5717), 1955–1959

Sorbara, M. T., & Pamer, E. G. (2019). Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunology*, *12*(1), 1-9

Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P., & Kaper, J. B. (2003). Bacteria-host communication: The language of hormones. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 8951–8956

Sprenger, G. A. (1995). Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. *Archives of Microbiology*, *164*(5), 324–330

Srikant, S. (2020). Evolutionary history of ATP-binding cassette proteins. *FEBS Letters*, *594*(23), 3882–3897

Stringer, A. M., Currenti, S., Bonocora, R. P., Baranowski, C., Petrone, B. L., Palumbo, M. J., Reilly, A. A., Zhang, Z., Erill, I., & Wade, J. T. (2014). Genome-scale analyses of *Escherichia coli* and *Salmonella enterica* AraC reveal noncanonical targets and an expanded core regulon. *Journal of Bacteriology*, *196*(3), 660-671

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & Von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646

Takao, M., Yen, H., & Tobe, T. (2014). LeuO enhances butyrate-induced virulence expression through a positive regulatory loop in enterohaemorrhagic *Escherichia coli*. *Molecular Microbiology*, *93*(6), 1302–1313

Tang, F., & Saier, M. H. (2014). Transport proteins promoting *Escherichia coli* pathogenesis. *Microbial Pathogenesis*, *0*(1), 41-55

Tang, S. Y., Fazelinia, H., & Cirino, P. C. (2008). AraC regulatory protein mutants with altered effector specificity. *Journal of the American Chemical Society*, *130*(15), 5267–5271

Tatsuno, I., Horie, M., Abe, H., Miki, T., Makino, K., Shinagawa, H., Taguchi, H., Kamiya, S., Hayashi, T., & Sasakawa, C. (2001). toxB gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. *Infection and Immunity*, *69*(11), 6660-6669

Tavender, T. J., Halliday, N. M., Hardie, K. R., & Winzer, K. (2008). LuxS-independent formation of AI-2 from ribulose-5-phosphate. *BMC Microbiology*, *8*(1), 1–8

Tejeda-Dominguez, F., Huerta-Cantillo, J., Chavez-Dueñas, L., & Navarro-Garcia, F. (2017). A novel mechanism for protein delivery by the type 3 secretion system for extracellularly secreted proteins. *MBio*, *8*(2), e00184-17

Ter Beek, J., Guskov, A., & Slotboom, D. J. (2014). Structural diversity of ABC transporters. *The Journal of General Physiology*, *143*(4), 419-435

Thomas, C., & Tampé, R. (2020). Structural and mechanistic principles of ABC transporters. *Annual Reviews Biochemistry*, *89*, 605–636

Tierrafría, V. H., Rioualen, C., Salgado, H., Lara, P., Gama-Castro, S., Lally, P., Gómez-Romero, L., Peña-Loredo, P., López-Almazo, A. G., Alarcón-Carranza, G., Betancourt-Figueroa, F., Alquicira-Hernández, S., Polanco-Morelos, J. E., García-Sotelo, J., Gaytan-Nuñez, E., Méndez-Cruz, C. F., Muñiz, L. J., Bonavides-Martínez, C., Moreno-Hagelsieb, G., Collado-Vides, J. (2022). RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation *in Escherichia coli* K-12. *Microbial Genomics*, *8*(5), 000833

Tobe, T., Beatson, S. A., Taniguchi, H., Abe, H., Bailey, C. M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., & Pallen, M. J. (2006). An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(40), 14941-14946

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., El Karoui, M., Frapy, E., Denamur, E. (2009). Organised genome dynamics in *the Escherichia coli* species results in highly diverse adaptive paths. *PLOS Genetics*, *5*(1), e1000344

Touchon, M., Perrin, A., De Sousa, J. A. M., Vangchhia, B., Burn, S., O'Brien, C. L., Denamur, E., Gordon, D., & Rocha, E. P. C. (2020). Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLOS Genetics*, *16*(6), e1008866

Tsang, A. K. L., Lee, H. H., Yiu, S. M., Lau, S. K. P., & Woo, P. C. Y. (2017). Failure of phylogeny inferred from multilocus sequence typing to represent bacterial phylogeny. *Scientific Reports 2017 7:1*, *7*(1), 1–12

Ty, F., Cj, S., 2#, A., Jm, Rl, W., Ra, B., Kj, G.-A., Mk, E., & Jf, P. (2019). A novel human enteroid-anaerobe co-culture system to study microbial-host interaction under physiological hypoxia. *BioRxiv*, 555755. https://doi.org/10.1101/555755

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115–e115

Van Der Hooft, J. J. J., Goldstone, R. J., Harris, S., Burgess, K. E. V., & Smith, D. G. E. (2019). Substantial extracellular metabolic differences found between phylogenetically closely related probiotic and pathogenic strains of *Escherichia coli*. *Frontiers in Microbiology*, *10*(FEB), 252. doi.10.3389/fcmicb.2019.00252

Vanaja, S. K., Bergholz, T. M., & Whittam, T. S. (2009). Characterization of the *Escherichia coli* O157:H7 Sakai GadE Regulon. *Journal of Bacteriology*, *191*(6), 1868-1877

Vasicek, E. M., O'Neal, L., Parsek, M. R., Fitch, J., White, P., & Gunn, J. S. (2021). L-arabinose transport and metabolism in *Salmonella* influences biofilm formation. *Frontiers in Cellular and Infection Microbiology*, *11*, 698146

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607

Wan, B., Zhang, Q., Tao, J., Zhou, A., Yao, Y. F., & Ni, J. (2016). Global transcriptional regulation by H-NS and its biological influence on the virulence of enterohemorrhagic *Escherichia coli*. *Gene*, *588*(2), 115–123

Warr, A. R., Hubbard, T. P., Munera, D., Blondel, C. J., Zur Wiesch, P. A., Abel, S., Wang, X., Davis, B. M., & Waldor, M. K. (2019). Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization. *PLOS Pathogens*, *15*(8), e1007652

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191

Westfahl, K. M., Merten, J. A., Buchaklian, A. H., & Klug, C. S. (2008). Functionally important ATP binding and hydrolysis sites in *Escherichia coli* MsbA. *Biochemistry*, *47*(52), 13878-13886

Wick, L. M., Qi, W., Lacher, D. W., & Whittam, T. S. (2005). Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *Journal of Bacteriology*, *187*(5), 1783–1791

Wiles, S., Hanage, W. P., Frankel, G., & Robertson, B. (2006). Modelling infectious disease — time to think outside the box? *Nature Reviews Microbiology 2006 4:4*, *4*(4), 307–312

Willing, B. P., Vacharaksa, A., Croxen, M., Thanachayanont, T., & Finlay, B. B. (2011). Altering host resistance to infections through microbial transplantation. *PLOS ONE*, *6*(10), e26988

Wong, A. R. C., Pearson, J. S., Bright, M. D., Munera, D., Robinson, K. S., Lee, S. F., Frankel, G., & Hartland, E. L. (2011). Enteropathogenic and enterohaemorrhagic

Escherichia coli: even more subversive elements. *Molecular Microbiology*, *80*(6), 1420–1438

Wood, T. K. (2009). Insights on *Escherichia coli* biofilm formation and inhibition from whole-transcriptome profiling. *Environmental Microbiology*, *11*(1), 1–15

Yang, D., Yang, Y., Qiao, P., Jiang, F., Zhang, X., Zhao, Z., Cai, T., Li, G., & Cai, W. (2023). Genomic island-encoded histidine kinase and response regulator coordinate mannose utilization with virulence in enterohemorrhagic *Escherichia coli*. *MBio*. https://doi.org/10.1128/MBIO.03152-22

Yang, W., Sun, H., Yan, J., Kang, C., Wu, J., & Yang, B. (2023). Enterohemorrhagic *Escherichia coli* senses microbiota-derived nicotinamide to increase its virulence and colonization in the large intestine. *Cell Reports*, *42*(6). https://doi.org/10.1016/J.CELREP.2023.112638

Yang, X., Dong, N., Liu, X., Yang, C., Ye, L., Chan, E. W. C., Zhang, R., & Chen, S. (2021). Co-conjugation of virulence plasmid and KPC plasmid in a clinical *Klebsiella pneumoniae* strain. *Frontiers in Microbiology*, *12*, 739461. doi.10.3389/fcmicb.2021.739461

Yara, D. A., Greig, D. R., Gally, D. L., Dallman, T. J., & Jenkins, C. (2020). Comparison of Shiga toxin-encoding bacteriophages in highly pathogenic strains of Shiga toxin-producing *Escherichia coli* O157:H7 in the UK. *Microbial Genomics*, *6*(3). https://doi.org/10.1099/MGEN.0.000334

Zhang, H., Lyu, Z., Fan, Y., Evans, C. R., Barber, K. W., Banerjee, K., Igoshin, O. A., Rinehart, J., & Ling, J. (2020). Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(36), 22167–22172

Zhang, J., Zhang, Y., Xia, Y., & Sun, J. (2021). Imbalance of the intestinal virome and altered viral-bacterial interactions caused by a conditional deletion of the vitamin D receptor. *Gut Microbes 13*(1), 1957408 https://doi.org/10.1080/19490976.2021.1957408

Zhang, P., Zhang, Q., & Whistler, R. L. (2003). L-arabinose release from arabinoxylan and arabinogalactan under potential gastric acidities. *Cereal Chemistry*, *80*(3), 252–254

Zhang, X., McDaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K., & Acheson, D. W. K. (2000). Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *The Journal of Infectious Diseases*, *181*(2), 664–670

Zhang, Y., Zagnitko, O., Rodionova, I., Osterman, A., & Godzik, A. (2011). The FGGY carbohydrate kinase family: Insights into the evolution of functional specificities. *PLoS Computational Biology*, *7*(12), 1002318. https://doi.org/10.1371/JOURNAL.PCBI.1002318

Zumbrun, S. D., Melton-Celsa, A. R., Smith, M. A., Gilbreath, J. J., Merrell, D. S., & O'Brien, A. D. (2013). Dietary choice affects Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 colonization and disease. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(23), E2126–E2133

8. Appendices

Table 8.1. Significantly upregulated genes in TUV93-0 following growth in MEM-HEPES supplemented with L-arabinose.

Feature ID	Fold change	Log_2 FC	FDR *P* value	Product
araH	151.81	7.25	3.49E-136	High-affinity L-arabinose transport system
ygeA	149.37	7.22	1.08E-126	Amino acid racemase
araF	145.09	7.18	5.17E-137	L-arabinose ABC transporter periplasmic binding protein
araE	139.81	7.13	1.39E-131	L-arabinose:H^+ symporter
araG	126.13	6.98	1.42E-132	L-arabinose ABC transporter ATP binding subunit
araD	98.47	6.62	2.16E-109	L-ribulose-5-phosphate 4-epimerase
araA	59.08	5.88	4.47E-107	L-arabinose isomerase
lysR	46.53	5.54	6.61E-89	DNA-binding transcriptional dual regulator
Z0417	43.15	5.43	6.77E-75	Putative ATP-binding component of transport system
araB	42.78	5.42	3.31E-98	L-ribulokinase
Z0415	26.51	4.73	1.55E-79	Putative periplasmic binding protein
ninG	21.44	4.42	0.04	Unknown protein encoded by prophage CP-933K
araJ	12.75	3.67	1.19E-38	Putative transport protein
rutE	10.91	3.45	2.01E-32	Putative malonic semialdehyde reductase
rutF	9.22	3.20	4.27E-29	FMN reductase
ddpB	9.04	3.18	5.00E-14	Putative D,D-dipeptide ABC transporter membrane subunit
rutA	8.32	3.06	4.40E-30	Pyrimidine monooxygenase
nac	8.1	3.02	9.50E-30	DNA-binding transcriptional dual regulator
cbl	8.08	3.01	2.51E-29	DNA-binding transcriptional activator
rutB	8.05	3.01	5.29E-28	Ureidoacrylate amidohydrolase
rutD	7.64	2.93	9.66E-27	Putative aminoacrylate hydrolase
rutC	7.53	2.91	1.21E-25	3-aminoacrylate deaminase
glnK	7.4	2.89	6.52E-27	Nitrogen regulatory protein PII-2
yegT	7.35	2.88	2.31E-15	Putative nucleoside permease protein
rutG	7.31	2.87	8.10E-26	Pyrimidine:H^+ symporter
yegU	7.09	2.83	1.13E-15	Putative aminoacrylate hydrolase
Z0418	7	2.81	1.77E-19	Putative permease component of transport system
yegV	6.87	2.78	1.03E-16	Putative sugar kinase
epd	6.82	2.77	2.46E-24	D-erythrose-4-phosphate dehydrogenase
yjfF	6.66	2.74	6.61E-25	Galactofuranose ABC transporter putative membrane subunit
amtB	6.58	2.72	1.21E-24	Ammonium transporter
Z4629	6.35	2.67	4.68E-21	Putative periplasmic binding transport protein
glnA	6.23	2.64	1.74E-23	Glutamine synthetase
sgrT	5.95	2.57	1.50E-17	Glucose uptake inhibitor
yrbN	5.86	2.55	4.18E-03	Hypothetical protein
Z3167	5.36	2.42	0.04	CP4-44 prophage; putative uncharacterised protein
yneE	5.24	2.39	3.30E-19	PF01062 family inner membrane protein
ytfT	5.2	2.38	2.13E-19	Galactofuranose ABC transporter putative membrane subunit
yidA	5.12	2.36	3.93E-16	Sugar phosphatase
galP	5.01	2.32	2.81E-16	Galactose:H^+ symporter
cheB	5.01	2.32	0.02	Protein-glutamate methylesterase/protein glutamine deamidase
ddpX	4.9	2.29	1.78E-15	D-alanyl-D-alanine dipeptidase
ytfR	4.87	2.28	4.10E-18	Galactofuranose ABC transporter putative ATP binding subunit
ydiH	4.29	2.10	1.43E-13	Hypothetical protein
malE	4.23	2.08	4.20E-14	Maltose ABC transporter periplasmic binding protein
hdeB	4.23	2.08	2.25E-13	Periplasmic acid stress chaperone
ompC_1	4.14	2.05	7.98E-15	Outer membrane porin C
malF	3.9	1.96	3.45E-12	Maltose ABC transporter membrane subunit
gltJ	3.9	1.96	1.31E-11	Glutamate/aspartate ABC transporter membrane subunit
malK	3.83	1.94	6.34E-11	Maltose ABC transporter ATP binding subunit
fbp	3.79	1.92	4.05E-13	Fructose-1,6-bisphosphatase 1
Z0414	3.7	1.89	1.45E-08	Hypothetical protein
yedL	3.68	1.88	6.34E-11	Putative acetyltransferase
tam	3.63	1.86	2.00E-12	*Trans*-aconitate 2-methyltransferase
xylF	3.63	1.86	3.54E-11	Xylose ABC transporter periplasmic binding protein

Z2983	3.62	1.86	2.40E-10	Putative tail fibre assembly protein of prophage CP-933T
nudK	3.61	1.85	1.61E-07	GDP-mannose hydrolase
fhuD	3.61	1.85	4.62E-07	Iron(III) hydroxamate ABC transporter periplasmic binding protein
gadC	3.59	1.84	3.36E-12	L-glutamate:4-aminobutyrate antiporter
tusC	3.58	1.84	8.12E-08	Sulphur transferase complex subunit
adhE	3.54	1.82	4.76E-12	Fused acetaldehyde-CoA dehydrogenase
glnL	3.53	1.82	1.54E-10	Protein histidine kinase
glnH	3.46	1.79	2.76E-11	L-glutamine ABC transporter periplasmic binding protein
Z2981	3.41	1.77	2.74E-06	IS629 transposase encoded within prophage CP-933T
Z2377	3.41	1.77	0.04	Unknown protein encoded within prophage CP-933R
bioF	3.37	1.75	4.45E-10	8-amino-7-oxononanoate synthase
purA	3.27	1.71	1.65E-10	Adenylosuccinate synthetase
Z1539	3.25	1.70	6.57E-08	Hypothetical protein
glnP	3.24	1.70	2.99E-09	L-glutamine ABC transporter membrane subunit
lsrG	3.23	1.69	1.69E-10	(4*S*)-4-hydroxy-5-phosphonooxypentane-2,3-dione isomerase
avtA	3.21	1.68	9.09E-09	Valine-pyruvate aminotransferase
ruvX	3.21	1.68	5.77E-06	Holliday junction resolvase
rimL	3.17	1.66	3.00E-09	50S ribosomal protein L7/L12-serine acetyltransferase
Z2976	3.16	1.66	3.02E-07	Unknown protein encoded by prophage CP-933T
lsrF	3.12	1.64	5.56E-10	3-hydroxy-2,4-pentadione 5-phosphate thiolase
lamB	3.1	1.63	3.93E-09	Maltose outer membrane channel / phage lambda receptor protein
Z5892	3.09	1.63	3.04E-07	Hypothetical protein
galS	3.06	1.61	1.63E-09	DNA-binding transcriptional dual regulator
fruB	3.05	1.61	4.05E-09	Fructose-specific PTS multiphosphoryl transfer protein
gltI	3.04	1.60	1.64E-09	Glutamate/aspartate ABC transporter periplasmic binding protein
bioC	2.98	1.58	1.71E-06	Malonyl-acyl carrier protein methyltransferase
ddpA	2.97	1.57	2.29E-08	Putative D,D-dipeptide ABC transporter periplasmic binding protein
nfo	2.97	1.57	3.91E-06	Endonuclease IV
yobD	2.95	1.56	3.45E-03	DUF986 domain-containing inner membrane protein
ptsP	2.94	1.56	4.65E-08	Phosphoenolpyruvate-protein phosphotransferase
glnG	2.92	1.55	3.04E-08	DNA-binding transcriptional dual regulator
cspG	2.91	1.54	6.09E-07	Cold shock protein
cytR	2.9	1.54	5.33E-08	DNA-binding transcriptional repressor
galT	2.89	1.53	1.38E-08	Galactose-1-phosphate uridylyltransferase
aroG	2.87	1.52	3.20E-08	3-deoxy-7-phosphoheptulonate synthase
nlpI	2.86	1.52	1.44E-08	Lipoprotein
wcaF	2.86	1.52	0.02	Colanic acid biosynthesis acetyltransferase
ydiQ	2.86	1.52	0.05	Putative electron transfer flavoprotein subunit
lsrK	2.83	1.50	1.91E-08	Autoinducer-2 kinase
galK	2.83	1.50	2.60E-08	Galactokinase
trmL	2.82	1.50	5.13E-05	tRNA (cytidine/uridine-2'-*O*)-ribose methyltransferase
gmm	2.81	1.49	0.03	GDP-mannose mannosyl hydrolase
yijF	2.8	1.49	0.01	DUF1287 domain-containing protein
hutW	2.75	1.46	4.77E-07	Putative oxygen independent coproporphyrinogen III oxidase
dmsD	2.74	1.45	4.96E-04	Redox enzyme maturation protein
gltL	2.73	1.45	7.78E-07	Glutamate/aspartate ABC transporter ATP binding subunit
hda	2.69	1.43	1.27E-06	Inhibitor of reinitiation of DNA replication
lsrR	2.68	1.42	1.20E-07	DNA-binding transcriptional repressor
cydD	2.67	1.42	3.42E-06	Glutathione/L-cysteine ABC exporter subunit
phrB	2.66	1.41	7.59E-07	Deoxyribodipyrimidine photo-lyase
fhuC	2.66	1.41	1.10E-05	Iron(III) hydroxamate ABC transporter ATP binding subunit
nei	2.66	1.41	1.05E-04	Endonuclease VIII
Z2972	2.64	1.40	2.17E-07	Unknown protein encoded by prophage CP-933T
bglJ	2.64	1.40	5.16E-06	DNA-binding transcriptional regulator
rstA	2.64	1.40	2.39E-04	DNA-binding transcriptional regulator
ygaH	2.64	1.40	1.31E-03	L-valine exporter
Z_RS32400	2.64	1.40	2.79E-03	Hypothetical protein
Z4628	2.64	1.40	0.01	Membrane protein
bioD	2.62	1.39	1.06E-04	Dethiobiotin synthetase
fau	2.61	1.38	1.32E-05	Putative 5-formyltetrahydrofolate cyclo-ligase
ytfQ	2.6	1.38	2.85E-07	Galactofuranose ABC transporter periplasmic binding protein
ydfH	2.6	1.38	1.09E-05	DNA-binding transcriptional repressor
btsR	2.6	1.38	1.02E-04	DNA-binding transcriptional dual regulator
mglB	2.56	1.36	4.64E-07	D-galactose/methyl-galactoside ABC transporter periplasmic binding

yoaG	2.56	1.36	4.48E-03	DUF1869 domain-containing protein
ybaQ	2.53	1.34	2.62E-06	Hypothetical protein
malG	2.53	1.34	1.01E-05	Maltose ABC transporter membrane subunit
yigA	2.53	1.34	1.33E-04	DUF484 domain-containing protein
ydhX	2.53	1.34	0.02	Putative 4Fe-4*S* ferredoxin-like protein
glnQ	2.52	1.33	1.00E-05	L-glutamine ABC transporter ATP binding subunit
galE	2.51	1.33	1.03E-06	UDP-glucose 4-epimerase
ptsN	2.51	1.33	1.77E-05	Phosphotransferase system enzyme IIA
yfiB	2.5	1.32	3.00E-04	Lipoprotein
ymcF	2.5	1.32	3.29E-04	Hypothetical protein
yggU	2.49	1.32	0.01	DUF167 domain-containing protein
Z1538	2.48	1.31	5.12E-06	Putative pilin
lipB	2.48	1.31	8.08E-05	Lipoate biosynthesis protein
yfaE	2.47	1.30	0.02	Ferredoxin-like diferric-tyrosyl radical cofactor maintenance protein
nhaR	2.45	1.29	1.08E-05	DNA-binding transcriptional activator
Z2970	2.44	1.29	2.14E-06	Putative regulator for prophage CP-933T
kduD	2.44	1.29	1.13E-05	Putative 2-keto-3-deoxy-D-gluconate dehydrogenase
fkpB	2.44	1.29	1.16E-03	Peptidyl-prolyl *cis-trans* isomerase
argO	2.44	1.29	7.57E-03	L-arginine exporter
bioA	2.43	1.28	4.24E-06	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
ddpC	2.43	1.28	9.88E-03	Putative D,D-dipeptide ABC transporter membrane subunit DdpC
Z1217	2.43	1.28	0.04	CP4-44 prophage; RadC-like JAB domain-containing protein
emrR	2.42	1.28	2.88E-03	DNA-binding transcriptional regulator
uhpB	2.42	1.28	4.86E-03	Sensory histidine kinase
lsrB	2.41	1.27	2.56E-06	Autoinducer-2 ABC transporter periplasmic binding protein
Z_RS14610	2.4	1.26	7.57E-04	Hypothetical protein
galM	2.39	1.26	4.40E-06	Galactose-1-epimerase
gadB	2.38	1.25	3.75E-06	Glutamate decarboxylase B
gadA	2.38	1.25	4.21E-06	Glutamate decarboxylase A
Z0325	2.38	1.25	4.39E-04	Unknown protein encoded in prophage CP-933I
Z2970	2.37	1.24	4.56E-06	Putative regulator for prophage CP-933T
cydC	2.36	1.24	5.35E-05	Glutathione/L-cysteine ABC exporter subunit
guaC	2.36	1.24	2.71E-04	GMP reductase
pgsA	2.34	1.23	5.45E-05	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
pgaA	2.32	1.21	2.42E-05	Poly-β-1,6-*N*-acetyl-D-glucosamine export outer membrane porin
folD	2.32	1.21	1.95E-04	Bifunctional methylenetetrahydrofolate dehydrogenase
nrdF	2.31	1.21	2.54E-05	Ribonucleoside-diphosphate reductase 2 subunit β
tadA	2.3	1.20	9.29E-04	tRNA adenosine34 deaminase
gltK	2.28	1.19	8.10E-05	Glutamate/aspartate ABC transporter membrane subunit
yeiI	2.28	1.19	1.20E-03	Putative sugar kinase
Z5882	2.28	1.19	2.28E-03	Hypothetical protein
yhdX	2.27	1.18	6.75E-04	Putative ABC transporter membrane subunit
hybG	2.27	1.18	0.04	Hydrogenase maturation factor
exbD	2.26	1.18	5.73E-05	Ton complex subunit
rarA	2.26	1.18	1.36E-04	Replication-associated recombination protein A
bioP	2.26	1.18	2.28E-04	Biotin transporter
hdeA	2.25	1.17	1.82E-05	Periplasmic acid stress chaperone
aceE	2.25	1.17	2.70E-05	Pyruvate dehydrogenase E1 component
alaA	2.25	1.17	3.89E-05	Glutamate-pyruvate aminotransferase
appC	2.25	1.17	2.00E-03	Cytochrome *bd*-II subunit 1
xerC	2.24	1.16	4.86E-04	Site-specific tyrosine recombinase
yciA	2.24	1.16	4.87E-04	Acyl-CoA thioesterase
nleF	2.23	1.16	1.18E-04	T3SS effector
yciY	2.23	1.16	1.51E-04	Hypothetical protein
nudE	2.23	1.16	1.60E-04	ADP-sugar diphosphatase
trmJ	2.22	1.15	1.33E-04	tRNA Cm32/Um32 methyltransferase
recQ	2.22	1.15	1.28E-03	ATP-dependent DNA helicase
dnaX	2.21	1.14	1.17E-04	DNA polymerase III, tau, gamma subunits; DNA elongation factor III
rapZ	2.21	1.14	3.37E-04	RNase adapter protein
Z4273	2.21	1.14	4.20E-03	P-loop NTPase domain-containing protein
lolD	2.21	1.14	0.02	Lipoprotein release complex - ATP binding subunit
folE	2.2	1.14	9.11E-05	GTP cyclohydrolase 1
dgcP	2.2	1.14	1.33E-04	Diguanylate cyclase
rhtC	2.2	1.14	5.27E-04	Threonine export protein

dbpA	2.19	1.13	4.67E-04	ATP-dependent RNA helicase
yigB	2.19	1.13	4.52E-03	5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase
gmd	2.19	1.13	4.84E-03	GDP-mannose 4,6-dehydratase
xylA	2.17	1.12	7.20E-05	D-xylose isomerase
fruK	2.17	1.12	8.60E-05	1-phosphofructokinase
pbpG	2.17	1.12	4.78E-04	Peptidoglycan DD-endopeptidase
Z2389	2.17	1.12	8.61E-04	Putative DNA modification methyltransferase prophage CP-933R
hemP	2.17	1.12	1.92E-03	Hemin uptake protein
csgC	2.17	1.12	0.05	Putative curli production protein
lnt	2.16	1.11	1.69E-04	Apolipoprotein *N*-acyltransferase
iap	2.16	1.11	3.42E-04	Alkaline phosphatase isozyme conversion protein
pgaB	2.16	1.11	5.54E-04	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase, β-1,6 glycoside hydrolase
dpaA	2.15	1.10	3.95E-04	Peptidoglycan meso-diaminopimelic acid protein amidase A
ysaA	2.15	1.10	1.57E-03	Putative electron transport protein
tsaA	2.15	1.10	0.02	tRNA m^6t^6A37 methyltransferase
thiH	2.15	1.10	0.04	2-iminoacetate synthase
cspA	2.14	1.10	1.15E-04	Cold shock protein
pdxH	2.14	1.10	8.07E-04	Pyridoxine/pyridoxamine 5'-phosphate oxidase
lgt	2.14	1.10	2.54E-03	Phosphatidylglycerol-prolipoprotein diacylglyceryl transferase
mnmH	2.14	1.10	0.01	tRNA 2-selenouridine synthase
Z2971	2.13	1.09	7.40E-05	Unknown protein encoded by prophage CP-933T
Z4317	2.13	1.09	2.31E-04	Unknown protein encoded by ISEc8
efeO	2.13	1.09	6.06E-04	Ferrous iron transport system protein
yfbR	2.13	1.09	2.22E-03	dCMP phosphohydrolase
yeeO	2.12	1.08	1.55E-04	Hypothetical protein
rfaF	2.12	1.08	5.60E-04	ADP-heptose-LPS heptosyltransferase
ycdY	2.12	1.08	1.34E-03	Chaperone protein
fhuF	2.11	1.08	1.13E-04	Ferric-siderophore reductase
hutX	2.11	1.08	8.98E-04	Heme utilization cystosolic carrier protein
Z5881	2.11	1.08	2.22E-03	Hypothetical protein
ybaP	2.1	1.07	1.59E-04	TraB family protein
aceF	2.1	1.07	1.59E-04	Pyruvate dehydrogenase, E2 subunit
Z_RS08590	2.1	1.07	5.35E-03	Hypothetical protein
yhhJ	2.09	1.06	1.16E-03	ABC transporter family protein
yeaR	2.09	1.06	1.95E-03	Hypothetical protein
map	2.09	1.06	4.60E-03	T3SS effector
tccP	2.08	1.06	5.16E-03	T3SS effector
entF	2.07	1.05	1.62E-04	Apo-serine activating enzyme
Z2974	2.07	1.05	1.82E-04	Unknown protein encoded by prophage CP-933T
fbpC	2.07	1.05	3.11E-03	CP4-6 prophage; ABC transporter ATP-binding protein
yhiD	2.07	1.05	5.96E-03	Inner membrane protein
menH	2.07	1.05	0.03	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
csgA	2.06	1.04	1.90E-04	Curlin, major subunit
Z2991	2.06	1.04	5.88E-04	Putative tail sheath protein of prophage CP-933T
infA	2.06	1.04	7.02E-04	Translation initiation factor IF-1
copD	2.06	1.04	2.88E-03	CopD family protein
wecC	2.06	1.04	9.28E-03	UDP-N-acetyl-D-mannosamine dehydrogenase
mltF	2.06	1.04	0.01	Membrane-bound lytic murein transglycosylase F
Z1341	2.05	1.04	3.66E-04	Unknown protein encoded by cryptic prophage CP-933M
Z2979	2.05	1.04	4.96E-04	Putative stability/partitioning protein prophage CP-933T
yigL	2.05	1.04	1.84E-03	Hypothetical protein
escT	2.05	1.04	0.05	T3SS biogenesis protein
pdhR	2.04	1.03	1.28E-03	DNA-binding transcriptional dual regulator
mtgA	2.04	1.03	1.33E-03	Peptidoglycan glycosyltransferase
ispA	2.04	1.03	1.45E-03	Geranyl diphosphate/farnesyl diphosphate synthase
wcaD	2.04	1.03	0.03	Colanic acid polymerase
yfaV	2.03	1.02	0.03	Putative transporter
pstA	2.03	1.02	0.03	Phosphate ABC transporter membrane subunit
Z5002	2.02	1.01	4.16E-03	Putative glycoside hydrolase 127 protein
yjcO	2.01	1.01	1.50E-03	Sel1 repeat-containing protein
ubiI	2.01	1.01	1.95E-03	2-octaprenylphenol 6-hydroxylase
Z2387	2.01	1.01	4.48E-03	Unknown protein encoded within prophage CP-933R
hisP	2.01	1.01	5.51E-03	Lysine/arginine/ornithine ABC transporter / histidine ABC transporter
btuF	2.01	1.01	0.02	Vitamin B12 ABC transporter periplasmic binding protein

sbcB	2	1.00	9.02E-04	Exodeoxyribonuclease I
gsk	2	1.00	1.16E-03	Inosine/guanosine kinase
metK	2	1.00	2.39E-03	Methionine adenosyltransferase
Z_RS09815	2	1.00	0.04	DNA-binding protein
serS	1.99	0.99	4.82E-04	Serine-tRNA ligase
tonB	1.98	0.99	6.38E-04	Ton complex subunit
Z_RS13965	1.98	0.99	6.69E-04	Hypothetical protein
minC	1.97	0.98	8.99E-04	Z-ring positioning protein
ybjX	1.97	0.98	2.50E-03	DUF535 domain-containing protein
entD	1.97	0.98	9.07E-03	Phosphopantetheinyl transferase
Z_RS18390	1.97	0.98	0.01	Hypothetical protein
Z3664	1.97	0.98	0.01	Putative RNA-guided DNA endonuclease
aspC	1.96	0.97	6.67E-04	Aspartate aminotransferase
gabT	1.96	0.97	7.71E-04	4-aminobutyrate aminotransferase
hemB	1.96	0.97	2.91E-03	Porphobilinogen synthase
nfi	1.96	0.97	0.02	Endonuclease V
ppc	1.95	0.96	1.01E-03	Phosphoenolpyruvate carboxylase
shuT	1.95	0.96	1.42E-03	Putative periplasmic binding protein
mdtF	1.95	0.96	1.95E-03	Multidrug efflux pump RND permease
ypfG	1.95	0.96	2.89E-03	DUF1176 domain-containing protein
potI	1.95	0.96	4.15E-03	Putrescine ABC transporter membrane subunit
hemG	1.95	0.96	5.41E-03	Protoporphyrinogen oxidase
rnhB	1.95	0.96	0.04	RNase HII
yfiR	1.93	0.95	2.17E-03	DUF4154 domain-containing protein
menC	1.93	0.95	4.45E-03	*o*-succinylbenzoate synthase
Z2980	1.93	0.95	8.78E-03	Putative stability/partitioning protein encoded within CP-933T
murE	1.92	0.94	2.01E-03	UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase
modE	1.92	0.94	9.61E-03	DNA-binding transcriptional dual regulator
znuC	1.91	0.93	1.23E-03	Zn^{2+} ABC transporter ATP binding subunit
yojI	1.91	0.93	3.65E-03	ABC transporter family protein / microcin J25 efflux protein
Z1533	1.91	0.93	3.71E-03	Putative oxidoreductase
Z2992	1.91	0.93	4.09E-03	Putative tail assembly protein of prophage CP-933T
pflC	1.91	0.93	0.04	Putative pyruvate formate-lyase 2 activating enzyme
tkt_1	1.9	0.93	1.35E-03	Transketolase 1
zapC	1.9	0.93	1.52E-03	Cell division protein
znuB	1.9	0.93	1.92E-03	Zn^{2+} ABC transporter membrane subunit
lpxB	1.9	0.93	7.93E-03	Lipid A disaccharide synthase
nadD	1.9	0.93	0.01	Nicotinate-nucleotide adenylyltransferase
cutA	1.9	0.93	0.02	Copper binding protein
aroL	1.9	0.93	0.03	Shikimate kinase 2
pyrF	1.9	0.93	0.04	Orotidine-5'-phosphate decarboxylase
lsrD	1.89	0.92	1.29E-03	Autoinducer-2 ABC transporter membrane subunit
bioB	1.89	0.92	1.66E-03	Biotin synthase
Z_RS13970	1.89	0.92	1.82E-03	Transcriptional regulator
yehS	1.89	0.92	0.01	DUF1456 domain-containing protein
yeiG	1.88	0.91	2.02E-03	*S*-formylglutathione hydrolase / *S*-lactoylglutathione hydrolase
tatB	1.88	0.91	2.78E-03	Sec-independent protein translocase protein
ribD	1.88	0.91	3.12E-03	Diaminohydroxyphosphoribosylaminopyrimidine deaminase
tusD	1.88	0.91	5.00E-03	Sulphurtransferase complex subunit
lplA	1.88	0.91	0.02	Outer membrane lipoprotein carrier protein
pstC	1.88	0.91	0.04	Phosphate ABC transporter membrane subunit
hscA	1.87	0.90	2.18E-03	Iron-sulphur cluster biosynthesis chaperone
lapB	1.87	0.90	2.63E-03	Lipopolysaccharide assembly protein
ratA	1.87	0.90	4.75E-03	Ribosome association toxin
coaBC	1.87	0.90	7.29E-03	4'-phosphopantothenoylcysteine decarboxylase
malQ	1.87	0.90	8.41E-03	4-α-glucanotransferase
tyrB	1.87	0.90	9.75E-03	Tyrosine aminotransferase
pgaC	1.87	0.90	0.01	Poly-*N*-acetyl-D-glucosamine synthase subunit
torD	1.87	0.90	0.01	Trimethylamine-*N*-oxide reductase-specific chaperone
dedA	1.87	0.90	0.01	Hypothetical protein
yjfZ	1.87	0.90	0.04	DUF2686 domain-containing protein
oxyR	1.86	0.90	2.23E-03	DNA-binding transcriptional regulator
ygfX	1.86	0.90	7.43E-03	Hypothetical protein
Z2977	1.86	0.90	0.02	Unknown protein encoded by prophage CP-933T

yrbL	1.85	0.89	3.65E-03	Protein kinase-like domain-containing protein
rfaL	1.85	0.89	0.01	*O*-antigen ligase
Z2989	1.85	0.89	0.02	Unknown protein encoded by prophage CP-933T
apt	1.85	0.89	0.02	Adenine phosphoribosyltransferase
yfjF	1.85	0.89	0.02	Putative component of the Rsx system
gabD	1.84	0.88	2.50E-03	Succinate-semialdehyde dehydrogenase
yagU	1.84	0.88	5.71E-03	DUF1440 domain-containing inner membrane protein
Z_RS29390	1.84	0.88	6.12E-03	Hypothetical protein
birA	1.84	0.88	0.01	DNA-binding transcriptional repressor/biotin-[acetyl-CoA-carboxylase] ligase
mukE	1.84	0.88	0.01	Chromosome partitioning protein
aroA	1.84	0.88	0.02	3-phosphoshikimate 1-carboxyvinyltransferase
pgaD	1.84	0.88	0.03	Poly-*N*-acetyl-D-glucosamine synthase subunit
tusA	1.84	0.88	0.03	Sulphur transfer protein
lpxA	1.83	0.87	4.75E-03	Acyl-[acyl-carrier-protein]—UDP-*N*-acetylglucosamine *O*-acyltransferase
malP	1.83	0.87	7.43E-03	Maltodextrin phosphorylase
rfaC	1.83	0.87	0.01	Lipopolysaccharide heptosyltransferase
dam	1.83	0.87	0.01	DNA adenine methyltransferase
frlD	1.83	0.87	0.03	Fructoselysine 6-kinase
hscB	1.82	0.86	6.04E-03	[Fe-S] cluster biosynthesis co-chaperone
yjaH	1.82	0.86	0.01	DUF1481 domain-containing protein
Z_RS31060	1.82	0.86	0.02	Transcriptional regulator
Z3269	1.82	0.86	0.02	Hypothetical protein
yddE	1.82	0.86	0.02	PF02567 family protein
yacC	1.82	0.86	0.02	Hypothetical protein
rlmH	1.82	0.86	0.03	23S rRNA $m^3\Psi1915$ methyltransferase
udk	1.82	0.86	0.03	Uridine/cytidine kinase
dnaC	1.82	0.86	0.03	DNA replication protein
tmaR	1.81	0.86	3.83E-03	Putative alpha helix protein
rppH	1.81	0.86	4.73E-03	RNA pyrophosphohydrolase
uvrC	1.81	0.86	6.17E-03	UvrABC excision nuclease subunit
thrA	1.81	0.86	7.06E-03	Fused aspartate kinase/homoserine dehydrogenase 1
mukF	1.81	0.86	8.16E-03	Chromosome partitioning protein
ybjT	1.81	0.86	0.01	Putative NAD(P)-binding protein
spoT	1.8	0.85	5.99E-03	Bifunctional (p)ppGpp synthase/hydrolase
moaC	1.8	0.85	0.01	Cyclic pyranopterin monophosphate synthase
mdtE	1.8	0.85	0.01	Multidrug efflux pump membrane fusion protein
rsmB	1.8	0.85	0.01	16S rRNA m^5C967 methyltransferase
epmB	1.8	0.85	0.02	Lysine 2,3-aminomutase
espL	1.8	0.85	0.03	T3SS biogenesis protein
metB	1.8	0.85	0.03	*O*-succinylhomoserine(thiol)-lyase
yjjA	1.8	0.85	0.04	DUF2501 domain-containing protein
glnD	1.79	0.84	6.03E-03	Protein-PII uridylyltransferase/uridylyl-removing enzyme
gss	1.79	0.84	6.13E-03	Fused glutathionylspermidine amidase / glutathionylspermidine synthetase
recB	1.79	0.84	8.26E-03	Exodeoxyribonuclease V subunit
Z0957	1.79	0.84	0.02	Unknown protein encoded by prophage CP-933K
recO	1.79	0.84	0.02	Recombination mediator protein
mutH	1.79	0.84	0.03	DNA mismatch repair protein
fre	1.78	0.83	8.29E-03	NAD(P)H-flavin reductase
dgcN	1.78	0.83	0.01	Diguanylate cyclase
glyQ	1.78	0.83	0.01	Glycine-tRNA ligase subunit α
efeB	1.78	0.83	0.01	Heme-containing peroxidase/deferrochelatase
yeeN	1.78	0.83	0.02	Putative transcriptional regulator
Z4852	1.78	0.83	0.04	Putative phospholipid biosynthesis acyltransferase
yfeO	1.77	0.82	8.35E-03	Ion channel protein
yjjV	1.77	0.82	0.05	Hypothetical protein
hisS	1.76	0.82	7.77E-03	Histidine tRNA ligase
ubiH	1.76	0.82	8.29E-03	2-octaprenyl-6-methoxyphenol 4-hydroxylase
rapA	1.76	0.82	0.01	RNA polymerase-binding ATPase and RNAP recycling factor
viaA	1.76	0.82	0.01	Hypothetical protein
yijE	1.76	0.82	0.02	Cysteine exporter
ydgC	1.76	0.82	0.02	Hypothetical protein
yjcB	1.76	0.82	0.02	Hypothetical protein
shuU	1.76	0.82	0.02	Heme ABC transporter permease
fumE	1.76	0.82	0.03	Fumarase E

bcsQ	1.75	0.81	0.02	Cellulose biosynthesis protein
yibL	1.75	0.81	0.02	DUF2810 domain-containing protein
fsa	1.75	0.81	0.04	Fructose-6-phosphate aldolase
mglA	1.74	0.80	7.56E-03	D-galactose/methyl-galactoside ABC transporter ATP binding subunit
Z4850	1.74	0.80	0.02	Putative *O*-methyltransferase
yceB	1.74	0.80	0.02	Putative lipid-binding lipoprotein
yshB	1.74	0.80	0.02	Small membrane protein
yhhS	1.74	0.80	0.02	Putative transporter
Z2519	1.74	0.80	0.04	UPF0509 family protein
ppiD	1.73	0.79	9.42E-03	Periplasmic folding chaperone
Z0309	1.73	0.79	9.70E-03	Putative cI repressor protein for prophage CP-933H
glyS	1.73	0.79	0.01	Glycine-tRNA ligase subunit β
topB	1.73	0.79	0.01	DNA topoisomerase III
atpH	1.73	0.79	0.02	ATP synthase F1 complex subunit δ
csdA	1.73	0.79	0.02	Cysteine sulfinate desulphinase
thiQ	1.73	0.79	0.04	Thiamine ABC transporter ATP binding subunit
Z0308	1.73	0.79	0.05	Unknown protein from prophage CP-933H
abrB	1.73	0.79	0.05	Putative regulator
lhgO	1.72	0.78	8.78E-03	L-2-hydroxyglutarate dehydrogenase
pldB	1.72	0.78	0.01	Lysophospholipase L2
trxB	1.71	0.77	0.01	Thioredoxin reductase
gfcB	1.71	0.77	0.01	Lipoprotein
ybdZ	1.71	0.77	0.02	Enterobactin biosynthesis protein
rplK	1.7	0.77	0.01	50S ribosomal subunit protein L11
fruA	1.7	0.77	0.01	Fructose-specific PTS multiphosphoryl transfer protein
ubiD	1.7	0.77	0.01	3-octaprenyl-4-hydroxybenzoate decarboxylase
glk	1.7	0.77	0.01	Glucokinase
bisC	1.7	0.77	0.02	Biotin sulfoxide reductase
tilS	1.7	0.77	0.02	tRNAIle-lysidine synthetase
Z2975	1.7	0.77	0.02	Unknown protein encoded by prophage CP-933T
gfcC	1.7	0.77	0.03	Capsule biosynthesis GfcC family protein
clsB	1.7	0.77	0.03	Cardiolipin synthase B
ygfI	1.7	0.77	0.04	Transcriptional regulator
mgtA	1.7	0.77	0.04	Mg^{2+} importing P-type ATPase
Z2978	1.69	0.76	0.01	Putative replication protein for prophage CP-933T
relA	1.69	0.76	0.01	GDP/GTP pyrophosphokinase
rlmD	1.69	0.76	0.02	23S rRNA m^5U1939 methyltransferase
alaC	1.69	0.76	0.02	Glutamate-pyruvate aminotransferase
rng	1.69	0.76	0.02	RNase G
dcm	1.69	0.76	0.03	DNA-cytosine methyltransferase
wzxE	1.69	0.76	0.03	Lipid IIIECA flippase
bfd	1.69	0.76	0.03	Bacterioferritin-associated ferredoxin
mdoC	1.69	0.76	0.04	Osmoregulated periplasmic glucans (OPG) biosynthesis protein C
hrpB	1.68	0.75	0.02	RNA-dependent NTPase
Z3931	1.68	0.75	0.04	Unknown protein encoded by prophage CP-933Y
glrR	1.68	0.75	0.05	DNA-binding transcriptional activator
iscR	1.67	0.74	0.01	DNA-binding transcriptional dual regulator
ptsP	1.67	0.74	0.03	Phosphoenolpyruvate-protein phosphotransferase
ygaZ	1.67	0.74	0.03	L-valine exporter
Z2990	1.67	0.74	0.04	Putative tail fibre component of prophage CP-933T
hisM	1.67	0.74	0.04	Lysine/arginine/ornithine ABC transporter / histidine ABC transporter
Z_RS32395	1.66	0.73	0.01	Hypothetical protein
ycbX	1.66	0.73	0.02	6-N-hydroxylaminopurine resistance protein
ddlA	1.66	0.73	0.02	D-alanine-D-alanine ligase A
can	1.66	0.73	0.02	Carbonic anhydrase 2
exbB	1.66	0.73	0.02	Ton complex subunit
acrA	1.66	0.73	0.03	Multidrug efflux pump membrane fusion lipoprotein
yqiA	1.66	0.73	0.03	Esterase
polB	1.65	0.72	0.02	DNA polymerase II
rpsF	1.65	0.72	0.02	30S ribosomal subunit protein S6
yjbH	1.65	0.72	0.03	Putative lipoprotein
murF	1.65	0.72	0.03	Dalanyl-D-alanine-adding enzyme
rlmB	1.65	0.72	0.03	23S rRNA 2'-*O*-ribose G2251 methyltransferase
dapF	1.65	0.72	0.03	Diaminopimelate epimerase

mnmC	1.65	0.72	0.03	tRNA 5-aminomethyl-2-thiouridylate methyltransferase
lolA	1.65	0.72	0.03	Outer membrane lipoprotein carrier protein
mnmE	1.65	0.72	0.04	5-carboxymethylaminomethyluridine-tRNA synthase GTPase subunit
Z4385	1.65	0.72	0.04	Putative ATP-binding protein of ABC transporter family
pxpA	1.65	0.72	0.04	5-oxoprolinase component A
rplD	1.64	0.71	0.02	50S ribosomal subunit protein L4
fbaA	1.64	0.71	0.02	Fructose-bisphosphate aldolase class II
yfgM	1.64	0.71	0.03	Ancillary SecYEG translocon subunit
hemE	1.64	0.71	0.03	Uroporphyrinogen decarboxylase
mdtH	1.64	0.71	0.04	Hypothetical protein
priB	1.64	0.71	0.04	Primosomal replication protein N
mukB	1.63	0.70	0.02	Chromosome partitioning protein
ubiB	1.63	0.70	0.03	Ubiquinone biosynthesis protein
Z2985	1.63	0.70	0.04	Putative tail fibre protein of prophage CP-933T
Z1519	1.63	0.70	0.04	Hypothetical protein
dnaA	1.62	0.70	0.02	Chromosomal replication initiator protein
lsrC	1.62	0.70	0.02	Autoinducer-2 ABC transporter membrane subunit
ribF	1.62	0.70	0.03	Bifunctional riboflavin kinase
gpmM	1.62	0.70	0.03	2,3-bisphosphoglycerate-independent phosphoglycerate mutase
dapA	1.62	0.70	0.03	4-hydroxy-tetrahydrodipicolinate synthase
yqhD	1.61	0.69	0.03	NADPH-dependent aldehyde reductase
pepD	1.61	0.69	0.03	Peptidase D
rplA	1.61	0.69	0.03	50S ribosomal subunit protein L1
zinT	1.61	0.69	0.03	Metal-binding protein
espG	1.61	0.69	0.04	T3SS effector
rbbA	1.61	0.69	0.04	Ribosome-associated ATPase
iscS	1.6	0.68	0.03	Cysteine desulphurase
cyaA	1.6	0.68	0.03	Adenylate cyclase
nanR	1.6	0.68	0.04	Transcriptional regulator
galR	1.6	0.68	0.04	DNA-binding transcriptional dual regulator
cpxA	1.59	0.67	0.03	Sensor histidine kinase
iscU	1.59	0.67	0.04	Scaffold protein for iron-sulphur cluster assembly
cvrA	1.58	0.66	0.04	Potassium/proton antiporter
rpsK	1.58	0.66	0.04	30S ribosomal subunit protein S11
cydA	1.58	0.66	0.04	Cytochrome *bd*-I subunit 1
maeA	1.58	0.66	0.05	Malate dehydrogenase
pdxA	1.58	0.66	0.05	4-hydroxythreonine-4-phosphate dehydrogenase
rpsS	1.58	0.66	0.05	30S ribosomal subunit protein S19
sthA	1.57	0.65	0.03	Soluble pyridine nucleotide transhydrogenase
rplC	1.57	0.65	0.04	50S ribosomal subunit protein L3
nuoC	1.57	0.65	0.04	NADH:quinone oxidoreductase subunit
hpf	1.57	0.65	0.04	Ribosome hibernation-promoting factor
hrpA	1.57	0.65	0.04	ATP-dependent 3'→5' RNA helicase
prfB	1.57	0.65	0.05	Peptide chain release factor RF2
Z2987	1.57	0.65	0.05	Putative tail fibre component of prophage CP-933T
rplB	1.56	0.64	0.04	50S ribosomal subunit protein L2
nrdE	1.56	0.64	0.04	Ribonucleoside-diphosphate reductase 2 subunit α
mpl	1.56	0.64	0.05	DP-N-acetylmuramate-L-alanyl-γ-D-glutamyl-meso-2,6-diaminoheptanedioate ligase
minD	1.55	0.63	0.05	Z-ring positioning protein
dctA	1.54	0.62	0.04	C4 dicarboxylate/orotate:H^+ symporter
bamA	1.54	0.62	0.05	Outer membrane protein assembly factor

Table 8.2. Significantly downregulated genes in TUV93-0 following growth in MEM-HEPES supplemented with L-arabinose.

Feature ID	Fold change	Log_2 FC	FDR *P* value	Product
garP	-14.7	-3.88	3.98E-32	Galactarate/D-glucarate transporter
mgtS	-12.91	-3.69	1.95E-43	Hypothetical protein
yjfN	-12.2	-3.61	0	Protease inhibitor
fadD	-11.32	-3.50	0	Long-chain fatty acid CoA ligase
mntS	-11	-3.46	0	Manganese accumulation protein
idlP	-10.43	-3.38	0.04	Leader peptide
bsmA	-9.67	-3.27	0	Biofilm peroxide resistance protein
murQ	-8.8	-3.14	5.02E-29	*N*-acetylmuramic acid 6-phosphate etherase
putA	-8.39	-3.07	9.24E-30	Proline dehydrogenase
Z5589	-8.06	-3.01	3.10E-30	23S ribosomal RNA
Z_RS12265	-7.53	-2.91	6.29E-11	Hypothetical protein
Z4637	-7.45	-2.90	6.05E-24	23S ribosomal RNA
garL	-7.37	-2.88	3.85E-16	α-dehydro-β-deoxy-*D*-glucarate aldolase
Z3875	-6.83	-2.77	1.21E-19	23S ribosomal RNA
mcbA	-6.78	-2.76	9.52E-25	DUF1471 family periplasmic protein
fadE	-6.73	-2.75	0	Acyl-CoA dehydrogenase
gtdA	-6.49	-2.70	3.18E-21	Putative 1,2-dioxygenase
pspG	-6.47	-2.69	0	Phage shock protein G
fadL	-6.36	-2.67	0	Long-chain fatty acid outer membrane channel
Z5259	-6.34	-2.66	4.07E-04	16S ribosomal RNA
ybeL	-6.28	-2.65	0	DUF1451 domain-containing protein
ynfM	-6.22	-2.64	0	Putative transport protein
cysD	-6.19	-2.63	9.18E-21	Sulphate adenylyl transferase subunit 2
Z3392	-6.07	-2.60	7.00E-20	Putative isomerase-decarboxylase
ivbL	-6.01	-2.59	0.04	Leader peptide
Z5379	-6.01	-2.59	3.85E-15	23S ribosomal RNA
Z0219	-5.96	-2.58	0	23S ribosomal RNA
ugpB	-5.94	-2.57	0	*sn*-glycerol 3-phosphate ABC transporter periplasmic binding protein
bolA	-5.93	-2.57	0	Putative regulator of murein genes
grcA	-5.85	-2.55	0	Stress-induced alternate pyruvate formate-lyase subunit
ydjF	-5.84	-2.55	4.91E-21	Putative DNA-binding transcriptional regulator
fadI	-5.83	-2.54	0	3-ketoacyl-CoA thiolase
yodC	-5.77	-2.53	3.73E-21	Hypothetical protein
ydcJ	-5.68	-2.51	0	VOC family protein
maiA	-5.64	-2.50	2.48E-17	Putative glutathione-*S*-transferase
yejG	-5.57	-2.48	0	Hypothetical protein
Z5534	-5.56	-2.48	4.20E-20	23S ribosomal RNA
cutC	-5.54	-2.47	0	Copper homeostasis protein
yohP	-5.53	-2.47	2.88E-03	Small membrane protein
ychH	-5.29	-2.40	0	Stress-induced protein
nrdD	-5.27	-2.40	0	Anaerobic ribonucleoside-triphosphate reductase
ydfA	-5.24	-2.39	0.05	Putative sulphatase
murP	-5.22	-2.38	1.01E-17	*N*-acetylmuramic acid-specific PTS enzyme IICB component
actP	-5.09	-2.35	2.52E-18	Acetate/glycolate cation symporter
cysP	-5.05	-2.34	4.10E-18	Thiosulfate/sulphate ABC transporter periplasmic binding protein
dhaK	-5.05	-2.34	0	Dihydroxyacetone kinase subunit K
acs	-5.03	-2.33	0	Acetyl-CoA synthetase
Z4018	-4.95	-2.31	0	Putative flavodoxin
fadB	-4.94	-2.30	0	Multifunctional enoyl-CoA hydratase
fadH	-4.89	-2.29	1.50E-17	2,4-dienoyl-CoA reductase
yfeK	-4.86	-2.28	1.29E-11	Hypothetical protein
pbp4b	-4.77	-2.25	4.72E-16	Penicillin binding protein 4B
pspC	-4.76	-2.25	0	Phage shock protein C
pspB	-4.74	-2.24	0	Phage shock protein B
pspE	-4.74	-2.24	0	Phage shock protein E
dkgA	-4.68	-2.23	0	2,5-didehydrogluconate reductase
YibT	-4.67	-2.22	0	Putative RNAse adapter protein
sseA	-4.65	-2.22	0	3-mercaptopyruvate sulphur transferase

ldtD	-4.62	-2.21	0	L,D-transpeptidase
adrA	-4.56	-2.19	1.44E-15	Diguanylate cyclase
yacL	-4.55	-2.19	0	Hypothetical protein
yjgR	-4.54	-2.18	0	DUF853 domain-containing protein
Z4090	-4.47	-2.16	0.04	Hypothetical protein
ybhQ	-4.46	-2.16	0	Inner membrane protein
aceA	-4.41	-2.14	1.04E-14	Isocitrate lyase
dhaL	-4.4	-2.14	4.05E-15	Dihydroxyacetone kinase subunit L
pspA	-4.38	-2.13	0	Phage shock protein A
pspD	-4.31	-2.11	0	Phage shock protein D
ybiI	-4.29	-2.10	0	Zinc finger domain-containing protein
norW	-4.16	-2.06	2.04E-14	NADH-flavorubredoxin reductase
yeaY	-4.13	-2.05	9.95E-14	Slp family lipoprotein
puuD	-4.06	-2.02	9.08E-12	γ-glutamyl-γ-aminobutyrate hydrolase
Z3394	-4.06	-2.02	3.61E-13	Putative transporter
ygjR	-4.05	-2.02	2.39E-14	Putative oxidoreductase
osmE	-4	-2.00	3.52E-14	Osmotically inducible lipoprotein
melR	-3.99	-2.00	3.52E-14	DNA-binding transcriptional dual regulator
Z4645	-3.98	-1.99	4.80E-03	16S ribosomal RNA
aceB	-3.92	-1.97	1.52E-12	Malate synthase
prpR	-3.88	-1.96	2.20E-13	DNA-binding dual transcriptional regulator
dhaM	-3.86	-1.95	2.38E-13	Dihydroxyacetone kinase subunit M
sbmC	-3.86	-1.95	1.79E-13	Putrescine ABC exporter membrane subunit
tcyJ	-3.86	-1.95	2.11E-13	Cysteine ABC transporter substrate-binding protein
Z3882	-3.86	-1.95	5.37E-10	16S ribosomal RNA
ycgB	-3.79	-1.92	3.21E-13	PF04293 family protein
hmpA	-3.76	-1.91	6.15E-13	Nitric oxide dioxygenase
bssR	-3.75	-1.91	4.38E-13	Regulator of biofilm formation
chpS	-3.75	-1.91	2.14E-11	ChpS antitoxin of the ChpB-ChpS toxin-antitoxin system
cstA	-3.74	-1.90	4.86E-13	Carbon starvation protein
ynhF	-3.72	-1.90	1.57E-11	Cytochrome bd-I accessory subunit
Z4353	-3.68	-1.88	1.93E-12	Putative enzyme
frdD	-3.67	-1.88	3.77E-12	Fumarate reductase membrane protein
yqaE	-3.66	-1.87	1.46E-12	Pmp3 family protein
ldtE	-3.65	-1.87	1.48E-12	L,D-transpeptidase
fadA	-3.63	-1.86	3.51E-12	3-ketoacyl-CoA thiolase
yobB	-3.62	-1.86	4.27E-12	Putative carbon-nitrogen hydrolase family protein
ybeQ	-3.6	-1.85	6.88E-12	Sel1 repeat-containing protein
fadJ	-3.56	-1.83	3.96E-12	Multifunctional 3-hydroxyacyl-CoA dehydrogenase
allR	-3.5	-1.81	8.45E-12	DNA-binding transcriptional repressor
yodD	-3.5	-1.81	7.17E-12	Stress-induced protein
dinJ	-3.49	-1.80	1.49E-11	Antitoxin/DNA-binding transcriptional repressor
folA	-3.49	-1.80	7.04E-11	Dihydrofolate reductase
ilvN	-3.48	-1.80	6.90E-07	Acetohydroxy acid synthase I subunit
mhpR	-3.48	-1.80	3.40E-11	DNA-binding transcriptional activator
yfdY	-3.48	-1.80	1.72E-11	DUF2545 domain-containing protein
htpX	-3.47	-1.79	1.18E-11	Protease
ynfD	-3.45	-1.79	1.48E-11	DUF1161 domain-containing protein
putP	-3.44	-1.78	3.03E-11	Major sodium/proline symporter
garR	-3.43	-1.78	1.45E-09	Tartronate semialdehyde reductase
Z_RS32125	-3.41	-1.77	2.11E-11	Hypothetical protein
ataR	-3.37	-1.75	4.00E-11	Hypothetical protein
cysN	-3.34	-1.74	2.55E-10	Sulphate adenylyltransferase subunit 1
kdpD	-3.33	-1.74	8.01E-11	Sensor histidine kinase
yajO	-3.33	-1.74	7.02E-11	1-deoxyxylulose-5-phosphate synthase
yfcH	-3.33	-1.74	7.32E-11	Epimerase family protein
Z5684	-3.33	-1.74	1.35E-10	Putative transcriptional regulator
yceH	-3.28	-1.71	2.62E-10	DUF480 domain-containing protein
ytfH	-3.28	-1.71	1.41E-09	Hypothetical protein
yhaH	-3.26	-1.70	1.30E-10	Putative inner membrane protein
aldB	-3.25	-1.70	1.65E-10	Aldehyde dehydrogenase B
yhhA	-3.25	-1.70	1.55E-10	DUF2756 domain-containing protein
ygaC	-3.22	-1.69	8.31E-10	DUF2002 domain-containing protein
yidQ	-3.22	-1.69	1.22E-09	Hypothetical protein

dnaK	-3.21	-1.68	2.02E-10	Chaperone Hsp70
yiaG	-3.21	-1.68	2.67E-10	Putative DNA-binding transcriptional regulator
csiE	-3.19	-1.67	2.58E-10	Stationary phase-inducible protein
yhfG	-3.19	-1.67	9.54E-09	DUF2559 domain-containing protein
yjiJ	-3.17	-1.66	1.06E-09	Putative transporter
Z1059	-3.16	-1.66	7.41E-10	Hypothetical protein
ygaV	-3.15	-1.66	2.13E-08	Putative DNA-binding transcriptional regulator
ytfJ	-3.14	-1.65	1.42E-09	PF09695 family protein
frdC	-3.13	-1.65	2.57E-09	Fumarate reductase membrane protein
qmcA	-3.13	-1.65	5.97E-10	PHB domain-containing protein
yaiA	-3.11	-1.64	2.12E-09	Hypothetical protein
ihfA	-3.1	-1.63	6.87E-10	Integration host factor subunit α
yraR	-3.1	-1.63	9.16E-10	Putative nucleoside-diphosphate-sugar epimerase
cpxP	-3.06	-1.61	1.27E-09	Periplasmic protein
yohC	-3.06	-1.61	1.22E-09	Putative inner membrane protein
folC	-3.05	-1.61	2.44E-09	Bifunctional folylpolyglutamate synthetase / dihydrofolate synthetase
dauA	-3.04	-1.60	1.85E-09	Aerobic C4-dicarboxylate transporte
agp	-3.03	-1.60	1.60E-09	glucose-1-phosphatase
cysJ	-3.03	-1.60	8.38E-08	Sulphite reductase, flavoprotein subunit
ssb1	-3.03	-1.60	2.39E-09	Single-stranded DNA-binding protein
yqjG	-3.03	-1.60	3.26E-09	Glutathionyl-hydroquinone reductase
srlR	-3.02	-1.59	1.37E-08	DNA-binding transcriptional repressor
ssrA	-3.02	-1.59	1.75E-09	10Sa RNA
yeaG	-3	-1.58	2.40E-09	Protein kinase
groS	-2.99	-1.58	4.59E-09	Co-chaperonin GroES
Z1060	-2.99	-1.58	3.41E-09	Hypothetical protein
ahr	-2.98	-1.58	1.41E-08	NADPH-dependent aldehyde reductase
sodC_2	-2.98	-1.58	3.46E-09	Superoxide dismutase family protein
btsT	-2.97	-1.57	9.22E-09	Pyruvate-H^+ symporter
uspE	-2.96	-1.57	4.05E-09	Universal stress protein E
tomB	-2.95	-1.56	5.02E-09	Hha toxicity modulator
puuA	-2.93	-1.55	4.04E-08	Putative glutamine synthetase
uspB	-2.93	-1.55	5.59E-09	Universal stress protein B
Z3624	-2.93	-1.55	2.42E-07	D-fructokinase
ribB	-2.92	-1.55	7.61E-09	3,4-dihydroxy-2-butanone-4-phosphate synthase
trxC	-2.92	-1.55	1.40E-08	Reduced thioredoxin 2
hcp	-2.9	-1.54	7.60E-06	*S*-nitrosylase
ybcK	-2.89	-1.53	4.07E-07	DLP12 prophage putative recombinase
ydhL	-2.89	-1.53	1.58E-08	DUF1289 domain-containing protein
yihS	-2.89	-1.53	8.55E-06	Sulfoquinovose isomerase
dsrB	-2.88	-1.53	8.17E-07	Hypothetical protein
ydjL	-2.88	-1.53	4.15E-08	Putative zinc-binding dehydrogenase
Z4789	-2.87	-1.52	4.59E-07	Hypothetical protein
galF	-2.86	-1.52	1.74E-08	UTP:glucose-1-phosphate uridylyltransferase
nrfD	-2.86	-1.52	0.03	Formate-dependent nitrate reductase complex
dosC	-2.85	-1.51	1.59E-08	DosC-DosP complex
wrbA	-2.85	-1.51	1.45E-08	NAD(P)H:quinone oxidoreductase
ascF	-2.84	-1.51	7.35E-08	β-glucoside specific PTS enzyme IIBC component
ytfK	-2.84	-1.51	1.73E-08	Hypothetical protein
Z4340	-2.84	-1.51	4.49E-07	Unknown protein encoded by ISEc8
Z5087	-2.84	-1.51	2.69E-08	Putative integrase for prophage 933L and LEE pathogenicity island
def	-2.83	-1.50	1.99E-08	Peptide deformylase
mokC	-2.83	-1.50	2.15E-08	Regulatory protein
ypfN	-2.82	-1.50	2.35E-08	PF13980 family protein
Z2148	-2.82	-1.50	6.62E-04	Unknown protein encoded within prophage CP-933O
rimJ	-2.81	-1.49	4.73E-08	Ribosomal-protein-S5-alanine *N*-acetyltransferase
ygaM	-2.81	-1.49	2.75E-08	DUF883 domain-containing protein
glpD	-2.8	-1.49	2.59E-08	Aerobic glycerol 3-phosphate dehydrogenase
bssS	-2.79	-1.48	2.77E-08	Regulator of biofilm formation
cspD	-2.79	-1.48	2.86E-08	DNA replication inhibitor
pgl	-2.79	-1.48	5.11E-08	6-phosphogluconolactonase
yccJ	-2.79	-1.48	3.21E-08	PF13993 family protein
ybhL	-2.78	-1.48	3.38E-08	Bax1-I family protein
Z5009	-2.78	-1.48	4.09E-08	Hypothetical protein

ymgE	-2.77	-1.47	9.14E-08	PF04226 family protein
csrA	-2.75	-1.46	4.22E-08	Carbon storage regulator
ydcK	-2.75	-1.46	1.32E-07	Putative acyltransferase
cysI	-2.74	-1.45	1.60E-06	Sulphite reductase, haemoprotein subunit
pdxI	-2.74	-1.45	7.93E-08	Pyridoxal reductase
ybaV	-2.74	-1.45	1.32E-05	Hypothetical protein
htpG	-2.72	-1.44	7.24E-08	Chaperone protein
Z0509	-2.71	-1.44	1.85E-07	Hypothetical protein
Z0974	-2.71	-1.44	2.46E-03	Putative tail component of prophage CP-933K
csqR	-2.7	-1.43	1.21E-06	DNA-binding transcriptional dual regulator
Z3620	-2.69	-1.43	9.71E-08	Hypothetical protein
nudL	-2.68	-1.42	1.13E-05	Putative NUDIX hydrolase
yobF	-2.68	-1.42	1.14E-07	DUF2527 domain-containing protein
chpB	-2.66	-1.41	4.49E-07	Endoribonuclease toxin
pat	-2.66	-1.41	1.30E-07	Protein lysine acetyltransferase
cysT	-2.64	-1.40	1.69E-06	tRNA-Cys
hokB	-2.64	-1.40	2.17E-07	Type I toxin-antitoxin system toxin
yjfY	-2.64	-1.40	6.47E-07	DUF1471 domain-containing protein
yniA	-2.63	-1.40	2.26E-07	Putative kinase
kdgR	-2.62	-1.39	2.78E-07	DNA-binding transcriptional repressor
dacC	-2.61	-1.38	3.02E-07	D-alanyl-D-alanine carboxypeptidase
iraD	-2.61	-1.38	1.39E-03	Anti-adapter protein
yfiL	-2.61	-1.38	2.66E-06	Hypothetical protein
yjgA	-2.61	-1.38	3.03E-07	Putative ribosome biogenesis factor
acnA	-2.6	-1.38	3.08E-07	Aconitate hydrase 1
hokD_3	-2.6	-1.38	2.91E-07	Type I toxin-antitoxin system toxin
ydcL	-2.6	-1.38	2.94E-07	DUF3313 domain-containing lipoprotein
yhhT	-2.6	-1.38	7.59E-07	Hypothetical protein
ytjA	-2.6	-1.38	4.05E-07	Hypothetical protein
rnpB	-2.59	-1.37	2.97E-07	RNase P catalytic RNA component
yoaC	-2.59	-1.37	1.00E-05	Hypothetical protein
Z5086	-2.59	-1.37	0.03	tRNA-Sec
zntB	-2.58	-1.37	6.01E-07	Zn^{2+}-H^{+} symporter
treA	-2.57	-1.36	9.17E-07	α , α-trehalase
ybdK	-2.57	-1.36	4.80E-07	Putative glutamate cysteine ligase 2
Z1924	-2.56	-1.36	1.05E-06	Stress-induced protein
Z_RS33015	-2.54	-1.34	6.04E-07	Hypothetical protein
glgS	-2.53	-1.34	7.27E-07	Surface composition regulator
yniB	-2.53	-1.34	7.06E-07	Uncharacterised protein
caiF	-2.52	-1.33	3.30E-05	Transcriptional regulator
panE	-2.52	-1.33	9.30E-07	2-dehydropantoate 2-reductase
psiF	-2.52	-1.33	1.57E-06	Phosphate starvation-inducible protein
rhaS	-2.52	-1.33	5.05E-05	Transcriptional activator
ygfZ	-2.52	-1.33	1.03E-06	Folate-binding protein
yibI	-2.52	-1.33	8.43E-06	DUF3302 domain-containing protein
Z5150	-2.52	-1.33	1.91E-06	Hypothetical protein
sdaA	-2.51	-1.33	8.40E-07	L-serine deaminase
ubiF	-2.51	-1.33	1.03E-06	3-demethoxyubiquinol 3-hydroxylase
yeaX	-2.51	-1.33	9.93E-06	Carnitine monooxygenase subunit
yihU	-2.51	-1.33	3.95E-03	3-sulfolactaldehyde reductase
yhgE	-2.5	-1.32	4.24E-06	Putative transporter
gatY	-2.48	-1.31	0.03	Tagatose-1,6-bisphosphate aldolase 2
groL	-2.48	-1.31	1.17E-06	Chaperonin GroEL
uhpT	-2.48	-1.31	2.10E-05	Hexose-6-phosphate:phosphate antiporter
yccU	-2.48	-1.31	1.87E-06	Putative HspQ acetyl donor
ydcS	-2.48	-1.31	1.75E-06	Putative ABC transporter periplasmic binding protein / polyhydroxybutyrate synthase
ygiW	-2.48	-1.31	1.31E-06	BOF family protein
xylG	-2.47	-1.30	2.56E-06	Xylose ABC transporter ATP binding subunit
Z0771	-2.47	-1.30	1.78E-06	Hypothetical protein
nimT	-2.46	-1.30	1.34E-05	2-nitroimidazole exporter
yeaH	-2.46	-1.30	1.51E-06	DUF444 domain-containing protein
yicJ	-2.46	-1.30	8.84E-06	Putative permease
rbsD	-2.45	-1.29	1.85E-06	D-ribose pyranase
asnC	-2.44	-1.29	1.37E-04	DNA-binding transcriptional dual regulator

osmB	-2.43	-1.28	2.31E-06	Osmotically inducible lipoprotein
uspG	-2.42	-1.28	2.67E-06	Universal stress protein G
hspQ	-2.41	-1.27	2.56E-06	Putative HspQ acetyl donor
kdpE	-2.41	-1.27	6.03E-06	DNA-binding transcriptional activator
rffT	-2.41	-1.27	1.57E-05	TDP-*N*-acetylfucosamine:lipid II *N*-acetylfucosaminyltransferase
Z_RS14675	-2.41	-1.27	3.91E-06	LexA family transcriptional regulator
cysH	-2.4	-1.26	1.05E-04	Phosphoadenosine phosphosulphate reductase
agaV	-2.39	-1.26	0.03	*N*-acetyl-D-galactosamine specific PTS
robA	-2.39	-1.26	3.82E-06	DNA-binding transcriptional dual regulator
yphG	-2.39	-1.26	1.62E-05	DUF5107 domain-containing protein
cspE	-2.38	-1.25	4.11E-06	Transcription anti-terminator and regulator of RNA stability
eamB	-2.38	-1.25	6.23E-05	Cysteine/O-acetylserine exporter
pdeR	-2.38	-1.25	4.58E-06	Cyclic di-GMP phosphodiesterase
fic	-2.37	-1.24	5.28E-06	Putative adenosine monophosphate—protein transferase
ybeD	-2.37	-1.24	1.67E-05	DUF493 domain-containing protein
aspA	-2.36	-1.24	4.77E-06	Aspartase
clpB	-2.36	-1.24	5.12E-06	Heat shock protein
lpp	-2.36	-1.24	4.66E-06	Murein lipoprotein
rpmE	-2.36	-1.24	5.44E-06	50S ribosomal subunit protein L31
yjcH	-2.36	-1.24	6.44E-04	DUF485 domain-containing inner membrane protein
frdA	-2.35	-1.23	6.03E-06	Fumarate reductase flavoprotein subunit
ybdD	-2.35	-1.23	2.12E-05	PF04328 family protein
yedK	-2.35	-1.23	1.33E-05	Genome maintenance protein
yjaB	-2.35	-1.23	2.50E-05	Peptidyl-lysine *N*-acetyltransferase
loiP	-2.34	-1.23	8.84E-06	Metalloprotease
ysaB	-2.34	-1.23	2.21E-03	Putative lipoprotein
YjjY	-2.33	-1.22	8.41E-06	Uncharacterised protein
Z4874	-2.33	-1.22	1.01E-05	Putative regulator
cysQ	-2.32	-1.21	8.21E-06	3'(2'),5'-bisphosphate nucleotidase
yjbE	-2.32	-1.21	9.06E-06	Hypothetical protein
Z5148	-2.32	-1.21	2.76E-04	Hypothetical protein
bioD	-2.31	-1.21	1.77E-05	Dethiobiotin synthetase
ucpA	-2.31	-1.21	9.94E-06	Oxidoreductase
ybeZ	-2.31	-1.21	9.35E-06	PhoH-like protein
yphA	-2.31	-1.21	5.16E-05	Hypothetical protein
ynjH	-2.3	-1.20	4.78E-04	DUF1496 domain-containing protein
Z_RS33335	-2.3	-1.20	9.56E-06	Hypothetical protein
ilvB	-2.29	-1.20	2.08E-05	Acetohydroxy acid synthase I subunit
mcbR	-2.29	-1.20	2.99E-05	DNA-binding transcriptional dual regulator
ssrS	-2.29	-1.20	1.14E-05	6S RNA
agaD	-2.28	-1.19	0.01	Galactosamine-specific PTS enzyme IID component
cecR	-2.28	-1.19	5.23E-05	DNA-binding transcriptional dual regulator
dadA	-2.28	-1.19	1.33E-05	D-amino acid dehydrogenase
nagE	-2.28	-1.19	1.73E-05	*N*-acetylglucosamine-specific PTS enzyme IIABC component
yjbJ	-2.28	-1.19	1.23E-05	Putative stress response protein
cnoX	-2.27	-1.18	2.28E-05	Chaperedoxin
gutQ	-2.27	-1.18	4.67E-05	D-arabinose 5-phosphate isomerase
rseA	-2.27	-1.18	1.53E-05	Anti-σ-E factor
ushA	-2.27	-1.18	1.78E-05	5'-nucleotidase / UDP-sugar hydrolase
Z2302	-2.27	-1.18	1.54E-04	Unknown protein encoded within prophage CP-933U
ffs	-2.26	-1.18	2.50E-05	4.5S RNA
nagB	-2.26	-1.18	2.47E-05	Glucosamine-6-phosphate deaminase
speA	-2.26	-1.18	1.84E-05	Biosynthetic arginine decarboxylase
Z2149	-2.26	-1.18	9.21E-04	Hypothetical protein
hslR	-2.25	-1.17	1.79E-04	Heat shock protein Hsp15
ycgN	-2.25	-1.17	5.04E-05	Putative metal-chelating domain-containing protein
yncL	-2.25	-1.17	2.19E-05	Stress response membrane protein
hcaR	-2.24	-1.16	4.61E-05	DNA-binding transcriptional dual regulator
rrf	-2.23	-1.16	2.54E-05	5S ribosomal RNA
ybaB	-2.23	-1.16	3.48E-05	Putative nucleoid-associated protein
ygeW	-2.23	-1.16	6.72E-03	Putative carbamoyltransferase
yihI	-2.23	-1.16	3.15E-05	Der GTPase-activating protein
yjfP	-2.23	-1.16	8.00E-05	Carboxylesterase
gpr	-2.22	-1.15	3.68E-05	L-glyceraldehyde 3-phosphate reductase

tas	-2.22	-1.15	4.91E-05	NADP(H)-dependent aldo-keto reductase
fdnI	-2.21	-1.14	2.09E-03	Formate dehydrogenase N subunit γ
frdB	-2.21	-1.14	4.16E-05	Fumarate reductase iron-sulphur protein
yncJ	-2.21	-1.14	7.05E-04	DUF2554 domain-containing protein
Z0609	-2.21	-1.14	3.03E-05	Hypothetical protein
cybC	-2.2	-1.14	6.51E-05	Cytochrome b (562)
osmF	-2.2	-1.14	3.65E-04	Glycine betaine ABC transporter periplasmic binding protein
ybaA	-2.2	-1.14	9.97E-05	DUF1428 domain-containing protein
yhcO	-2.2	-1.14	5.16E-05	Putative barnase inhibitor
Z5890	-2.2	-1.14	4.64E-05	Partial putative integrase
dapE	-2.19	-1.13	4.77E-05	Succinyl-diaminopimelate desuccinylase
decR	-2.19	-1.13	1.33E-04	DNA-binding transcriptional activator
nnr	-2.19	-1.13	5.74E-05	NAD(P)HX epimerase / NAD(P)HX dehydratase
yafN	-2.19	-1.13	4.07E-04	Antitoxin
yicH	-2.19	-1.13	5.06E-05	AsmA family protein
Z0510	-2.19	-1.13	5.16E-05	Hypothetical protein
Z0608	-2.19	-1.13	4.22E-05	Putative outer membrane export protein
yceM	-2.17	-1.12	7.88E-05	Putative oxidoreductase
Z6057	-2.17	-1.12	1.29E-03	tRNA-Arg
degP	-2.16	-1.11	7.05E-05	Periplasmic serine endoprotease
dmlR	-2.16	-1.11	1.59E-04	DNA-binding transcriptional regulator
ecnB	-2.16	-1.11	6.05E-05	Entericidin B
hokE_1	-2.16	-1.11	3.66E-04	Type I toxin-antitoxin system toxin
csrA	-2.15	-1.10	1.98E-04	Carbon storage regulator
macB	-2.15	-1.10	8.32E-05	ABC-type tripartite efflux pump ATP binding/membrane subunit
mtlA	-2.15	-1.10	6.59E-05	Mannitol-specific PTS enzyme IICBA component
pdeB	-2.15	-1.10	7.32E-05	c-di-GMP phosphodiesterase
rseD	-2.15	-1.10	1.31E-04	RpoE leader peptide
yajL	-2.15	-1.10	1.28E-04	Protein/nucleic acid deglycase 3
yedI	-2.15	-1.10	1.79E-04	DUF808 domain-containing inner membrane protein
yneJ	-2.15	-1.10	4.50E-04	Putative DNA-binding transcriptional regulator
Z1874	-2.15	-1.10	0.03	Putative anti-terminator Q of prophage CP-933X
mscK	-2.14	-1.10	9.69E-05	Potassium dependent, small conductance mechanosensitive channel
nadR	-2.14	-1.10	8.49E-05	DNA-binding transcriptional repressor/NMN adenylyltransferase
ygiN	-2.14	-1.10	9.88E-05	Hypothetical protein
hokD_5	-2.13	-1.09	1.04E-04	Type I toxin-antitoxin system toxin
tnaA	-2.13	-1.09	6.33E-04	Tryptophanase
lysP	-2.12	-1.08	1.45E-04	Lysine-H^+ symporter
gatZ	-2.11	-1.08	0.03	Putative tagatose-1,6-bisphosphate aldolase 2 chaperone
sfsA	-2.11	-1.08	1.01E-04	Sugar fermentation stimulation protein A
ygiB	-2.11	-1.08	1.22E-04	DUF1190 domain-containing protein
blc	-2.1	-1.07	1.40E-04	Outer membrane lipoprotein
dksA	-2.1	-1.07	1.23E-04	RNA polymerase-binding transcription factor
hslV	-2.1	-1.07	1.77E-04	Peptidase component of the HslVU protease
yahO	-2.1	-1.07	1.16E-04	DUF1471 domain-containing protein
adiY	-2.09	-1.06	1.16E-03	DNA-binding transcriptional activator
asr	-2.09	-1.06	3.65E-03	Periplasmic chaperone
slyB	-2.09	-1.06	1.38E-04	Outer membrane lipoprotein
uxuR	-2.09	-1.06	1.79E-04	DNA-binding transcriptional repressor
agaE	-2.08	-1.06	0.01	Putative phosphotransferase system enzyme subunit
cysK	-2.08	-1.06	1.41E-04	Cysteine synthase A
ypeB	-2.08	-1.06	6.90E-04	PF12843 family protein
Z_RS09840	-2.08	-1.06	3.24E-04	Hok/Gef family protein
marC	-2.07	-1.05	4.40E-04	Inner membrane protein
narK	-2.07	-1.05	4.09E-03	Nitrite extrusion protein
sufA	-2.07	-1.05	2.23E-04	Fe-S cluster assembly scaffold
yehP	-2.07	-1.05	0.03	VWA domain-containing protein
cdd	-2.06	-1.04	2.05E-04	Cytidine/deoxycytidine deaminase
eptA	-2.06	-1.04	1.85E-03	Phosphoethanolamine transferase
sixA	-2.06	-1.04	1.75E-04	Phosphohistidine phosphatase
yeaQ	-2.06	-1.04	2.09E-04	PF04226 family protein
potE	-2.05	-1.04	1.18E-03	Putrescine transport protein
rsd	-2.05	-1.04	2.63E-04	Regulator of σ D
ybbW	-2.05	-1.04	2.17E-03	Putative allantoin permease

Z4776	-2.05	-1.04	3.99E-04	Hypothetical protein
ldtA	-2.04	-1.03	2.38E-04	L,D-transpeptidase
mlaC	-2.04	-1.03	4.00E-04	Intermembrane phospholipid transport system - periplasmic binding protein
uspF	-2.04	-1.03	2.14E-04	Universal stress protein F
Z4882	-2.04	-1.03	3.00E-04	Type II toxin-antitoxin system HicB family antitoxin
Z6073	-2.04	-1.03	2.23E-04	Putative repressor protein encoded by cryptic prophage CP-933P
arnT	-2.03	-1.02	6.35E-04	Lipid IVA 4-amino-4-deoxy-L-arabinosyltransferase
frwB	-2.03	-1.02	2.75E-03	PTS system fructose-like IIB component 1
ldcC	-2.03	-1.02	2.97E-04	Lysine decarboxylase 2
nifJ	-2.03	-1.02	2.57E-04	Flavodoxin oxidoreductase
ybiB	-2.03	-1.02	3.01E-04	Non-specific DNA-binding protein
Z1383	-2.03	-1.02	2.84E-03	Unknown protein encoded by cryptic prophage CP-933M
Z3390	-2.03	-1.02	2.95E-04	Putative hydroxylase
yfcG	-2.02	-1.01	5.91E-04	Disulphide bond oxidoreductase
Z_RS29520	-2.02	-1.01	2.66E-03	KGG domain-containing protein
ghoS_2	-2.01	-1.01	2.90E-03	Antitoxin of the GhoTS toxin-antitoxin system
grpE	-2.01	-1.01	3.74E-04	Nucleotide exchange factor
rpoE	-2.01	-1.01	3.42E-04	RNA polymerase σ factor
rtcB	-2.01	-1.01	1.45E-03	RNA-splicing ligase
yghA	-2.01	-1.01	6.50E-04	$NADP^+$-dependent aldehyde reductase
ppnN	-2	-1.00	3.56E-04	Nucleotide 5'-monophosphate nucleosidase
yeeE	-2	-1.00	6.24E-04	Thiosulphate transporter
Z4324	-2	-1.00	3.76E-03	Putative transposase
cusR	-1.99	-0.99	2.83E-03	DNA-binding transcriptional activator
hokD_1	-1.99	-0.99	5.33E-04	Type I toxin-antitoxin system toxin
rnd	-1.99	-0.99	7.96E-04	RNase D
Z0347	-1.99	-0.99	8.10E-03	Hypothetical protein
Z5619	-1.99	-0.99	3.56E-03	Putative transcriptional regulator of sorbose uptake
fucA	-1.98	-0.99	2.54E-03	L-fuculose-1-phosphate aldolase
dkgB	-1.97	-0.98	1.77E-03	Methylglyoxal reductase
gloB	-1.97	-0.98	7.28E-04	Hydroxyacylglutathione hydrolase
ycfH	-1.97	-0.98	5.47E-04	Putative metal-dependent hydrolase
yciW	-1.97	-0.98	0.01	Putative peroxidase
ydiL	-1.97	-0.98	0.02	DUF1870 domain-containing protein
yfgG	-1.97	-0.98	4.96E-04	Nickel/cobalt stress response protein
yhjD	-1.97	-0.98	5.15E-04	Putative transporter
ompR	-1.96	-0.97	6.62E-04	DNA-binding dual transcriptional regulator
yajI	-1.96	-0.97	6.20E-03	Hypothetical protein
ycfP	-1.96	-0.97	6.57E-04	PF05728 family protein
yeaW	-1.96	-0.97	0.01	Carnitine monooxygenase subunit
arcA	-1.95	-0.96	6.55E-04	DNA-binding transcriptional dual regulator
ligA	-1.95	-0.96	1.35E-03	DNA ligase
trkA	-1.95	-0.96	1.15E-03	NAD-binding component of Trk potassium transporters
uspA	-1.95	-0.96	6.33E-04	Universal stress protein A
ytfE	-1.95	-0.96	6.22E-04	Iron-sulphur cluster repair protein
gcvT	-1.94	-0.96	5.09E-03	Aminomethyltransferase
hlyD	-1.94	-0.96	8.86E-04	Secretion protein
mdaB	-1.94	-0.96	3.58E-03	NADPH:quinone oxidoreductase
yibH	-1.94	-0.96	1.45E-03	Inner membrane protein
yjdC	-1.94	-0.96	9.26E-04	Putative DNA-binding transcriptional regulator
crr	-1.93	-0.95	8.07E-04	PTS system, glucose-specific IIA component
ihfB	-1.93	-0.95	8.99E-04	Integration host factor subunit β
melA	-1.93	-0.95	1.58E-03	Alpha-galactosidase
ydeN	-1.93	-0.95	2.07E-03	Putative sulfatase
yncG	-1.93	-0.95	0.03	Putative glutathione S-transferase
nagA	-1.92	-0.94	1.18E-03	*N*-acetylglucosamine-6-phosphate deacetylase
pheT	-1.92	-0.94	9.08E-04	Phenylalanine tRNA synthetase, beta-subunit
rssB	-1.92	-0.94	1.06E-03	Two-component system response regulator
ybaE	-1.92	-0.94	1.16E-03	Putative nucleoid-associated protein
yjbB	-1.92	-0.94	1.91E-03	Putative inorganic phosphate export protein
acpP	-1.91	-0.93	9.91E-04	Acyl carrier protein
emtA	-1.91	-0.93	2.09E-03	Murein transglycosylase E
higA	-1.91	-0.93	3.65E-03	Antitoxin/DNA-binding transcriptional repressor
nudC	-1.91	-0.93	2.46E-03	RNA decapping hydrolase

ypfJ	-1.91	-0.93	1.15E-03	Uncharacterised protein
zitB	-1.91	-0.93	1.77E-03	Zn^{2+}/Cd^{2+}/Ni^{2+}/Cu^{2+} exporter
aldA	-1.9	-0.93	1.15E-03	Aldehyde dehydrogenase
sstT	-1.9	-0.93	1.09E-03	Serine/threonine transporter
ybjQ	-1.9	-0.93	2.39E-03	Putative heavy metal binding protein
yegP	-1.9	-0.93	1.15E-03	Hypothetical protein
yifK	-1.9	-0.93	6.35E-03	Putative transporter
zupT	-1.9	-0.93	1.96E-03	Divalent metal ion transporter
cysA	-1.89	-0.92	3.34E-03	Sulphate/thiosulphate ABC transporter ATP binding subunit
rpoH	-1.89	-0.92	1.40E-03	RNA polymerase, σ(32) factor
ynfL	-1.89	-0.92	0.01	Putative DNA-binding transcriptional regulator
gmhB	-1.88	-0.91	1.64E-03	D-*glycero*-β-D-*manno*-heptose-1,7-bisphosphate 7-phosphatase
ibpA	-1.88	-0.91	2.16E-03	Small heat shock protein
msyB_1	-1.88	-0.91	1.53E-03	Acidic protein
queE	-1.88	-0.91	1.53E-03	Putative 7-carboxy-7-deazaguanine synthase
sbp	-1.88	-0.91	0.01	Periplasmic sulphate-binding protein
yggE	-1.88	-0.91	1.79E-03	DUF541 domain-containing protein
yidB	-1.88	-0.91	1.45E-03	DUF937 domain-containing protein
Z0972	-1.88	-0.91	8.68E-03	Putative tail component of prophage CP-933K
dnaJ	-1.87	-0.90	1.74E-03	Chaperone protein
katG	-1.87	-0.90	1.98E-03	Catalase
pflA	-1.87	-0.90	1.98E-03	Pyruvate formate lyase activating enzyme 1
ybbJ	-1.87	-0.90	1.77E-03	NfeD-like family protein
yfbV	-1.87	-0.90	1.64E-03	PF04217 family membrane protein
Z_RS29535	-1.87	-0.90	0.02	ACP S-malonyltransferase
kbp	-1.86	-0.90	1.94E-03	K^+ binding protein
mlc	-1.86	-0.90	3.07E-03	Putative NAGC-like transcriptional regulator
msyB_2	-1.86	-0.90	5.74E-03	Acidic protein
sdaC	-1.86	-0.90	3.50E-03	Probable serine transporter
ybjD	-1.86	-0.90	2.09E-03	DUF2813 domain-containing protein
yiiM	-1.86	-0.90	1.91E-03	6-hydroxyaminopurine reductase
hslO	-1.85	-0.89	3.16E-03	Molecular chaperone Hsp33
nrdG	-1.85	-0.89	4.26E-03	Anaerobic ribonucleotide reductase activating protein
ygaP	-1.85	-0.89	5.35E-03	Thiosulphate sulphurtransferase
yhbO	-1.85	-0.89	4.80E-03	Protein/nucleic acid deglycase 2
yijD	-1.85	-0.89	4.64E-03	DUF1422 domain-containing inner membrane protein
araC	-1.84	-0.88	2.42E-03	Transcriptional regulator
ldtC	-1.84	-0.88	5.28E-03	L,D-transpeptidase
napC	-1.84	-0.88	0.03	Cytochrome c-type protein
yihV	-1.84	-0.88	0.01	6-deoxy-6-sulfofructose kinase
yohF	-1.84	-0.88	4.71E-03	Putative oxidoreductase
Z3118	-1.84	-0.88	0.01	Unknown protein encoded within prophage CP-933U
cmk	-1.83	-0.87	3.70E-03	Cytidylate kinase
cof	-1.83	-0.87	7.58E-03	HMP-PP phosphatase
cycA	-1.83	-0.87	2.51E-03	D-serine/alanine/glycine/:H^+ symporter
dgcM	-1.83	-0.87	4.64E-03	Diguanylate cyclase
mlaA	-1.83	-0.87	3.80E-03	Outer membrane lipoprotein
mlaF	-1.83	-0.87	3.57E-03	Intermembrane phospholipid transport system, ATP binding subunit
mtlR	-1.83	-0.87	2.79E-03	Transcriptional repressor
prlF	-1.83	-0.87	4.48E-03	Antitoxin
ybhD	-1.83	-0.87	0.03	Putative DNA-binding transcriptional regulator
yiiR	-1.83	-0.87	0.02	DUF805 domain-containing protein
Z3123	-1.83	-0.87	4.09E-03	Unknown protein encoded within prophage CP-933U
chaB	-1.82	-0.86	5.25E-03	Putative cation transport regulator
dtpB	-1.82	-0.86	3.00E-03	Dipeptide/tripeptide:H^+ symporter
inaA	-1.82	-0.86	7.95E-03	Putative lipopolysaccharide kinase
otsB	-1.82	-0.86	3.77E-03	Trehalose-6-phosphate phosphatase
yihO	-1.82	-0.86	0.02	Putative sulfoquinovose transporter
yjgM	-1.82	-0.86	6.09E-03	GNAT family *N*-acetyltransferase
yphF	-1.82	-0.86	0.02	Putative ABC transporter periplasmic binding protein
fxsA	-1.81	-0.86	4.34E-03	Hypothetical protein
hns	-1.81	-0.86	2.80E-03	DNA-binding transcriptional dual regulator
rpnB	-1.81	-0.86	4.52E-03	Recombination-promoting nuclease
rraB	-1.81	-0.86	3.65E-03	Ribonuclease E inhibitor protein B

sufB	-1.81	-0.86	3.27E-03	Fe-S cluster assembly protein
tus	-1.81	-0.86	4.60E-03	DNA replication terminus site-binding protein
yaiY	-1.81	-0.86	8.78E-03	DUF2755 domain-containing inner membrane protein
ydhQ	-1.81	-0.86	3.93E-03	Putative adhesin-related protein
ysgA	-1.81	-0.86	3.40E-03	Putative dienelactone hydrolase
frr	-1.8	-0.85	3.66E-03	Ribosome recycling factor
tisB	-1.8	-0.85	0.02	Type I toxin-antitoxin system toxin
Z4787	-1.8	-0.85	0.02	Hypothetical protein
astC	-1.79	-0.84	3.68E-03	Succinylornithine transaminase
hcr	-1.79	-0.84	0.03	NADH oxidoreductase
pnuC	-1.79	-0.84	7.55E-03	Nicotinamide riboside transporter
rimI	-1.79	-0.84	0.02	*N*-acetyltransferase
sapA	-1.79	-0.84	4.35E-03	Putative periplasmic binding protein
tnaB	-1.79	-0.84	0.03	Low affinity tryptophan permease
yaeH	-1.79	-0.84	4.54E-03	DUF3461 domain-containing protein
yeaV	-1.79	-0.84	0.02	Putative transporter
adeD	-1.78	-0.83	7.19E-03	Adenine deaminase
bamB	-1.78	-0.83	4.60E-03	Outer membrane protein assembly factor
hokE_2	-1.78	-0.83	0.01	Type I toxin-antitoxin system toxin
holE	-1.78	-0.83	0.03	DNA polymerase III subunit θ
ppsA	-1.78	-0.83	4.09E-03	Phosphoenolpyruvate synthase
puuR	-1.78	-0.83	0.02	DNA-binding transcriptional repressor
yafV	-1.78	-0.83	8.83E-03	Antitoxin
yecN	-1.78	-0.83	7.22E-03	MAPEG family inner membrane protein
Z0056	-1.78	-0.83	7.45E-03	Putative antitoxin of gyrase inhibiting toxin-antitoxin system
zapA	-1.78	-0.83	4.59E-03	Cell division protein
iaaA	-1.77	-0.82	5.84E-03	β-aspartyl-peptidase
rpmI	-1.77	-0.82	4.93E-03	50S ribosomal subunit protein A
ydgD	-1.77	-0.82	4.50E-03	Putative serine protease
yegS	-1.77	-0.82	5.91E-03	Lipid kinase
nrdH	-1.76	-0.82	5.71E-03	Glutaredoxin-like protein
queG	-1.76	-0.82	8.18E-03	Epoxyqueuosine reductase
Z_RS08560	-1.76	-0.82	0.01	AlpA family transcriptional regulator
Z1769	-1.76	-0.82	6.45E-03	Unknown protein encoded by prophage CP-933N
Z6074	-1.76	-0.82	5.77E-03	Unknown protein encoded by cryptic prophage CP-933P
fmt	-1.75	-0.81	5.89E-03	10-formyltetrahydrofolate:L-methionyl-tRNA^{fMet} *N*-formyltransferase
fnr	-1.75	-0.81	6.09E-03	DNA-binding transcriptional dual regulator
fucO	-1.75	-0.81	6.01E-03	L-1,2-propanediol oxidoreductase
opgD	-1.75	-0.81	8.05E-03	Glucans biosynthesis protein D
Z1766	-1.75	-0.81	7.43E-03	Unknown protein encoded by prophage CP-933N
ackA	-1.74	-0.80	6.58E-03	Acetate kinase
ftsX	-1.74	-0.80	7.74E-03	Cell division membrane protein
uxaB	-1.74	-0.80	8.40E-03	Tagaturonate reductase
chbC	-1.73	-0.79	0.04	*N,N'*-diacetylchitobiose-specific PTS enzyme IIC component
espM1	-1.73	-0.79	7.43E-03	T3SS effector
sapB	-1.73	-0.79	9.79E-03	Putrescine ABC exporter membrane subunit
yahN	-1.73	-0.79	9.70E-03	2-oxoglutaramate amidase
ybhR	-1.73	-0.79	0.01	ABC exporter membrane subunit
ygdR	-1.73	-0.79	8.83E-03	DUF903 domain-containing lipoprotein
yqjA	-1.73	-0.79	8.12E-03	DedA family protein
Z1345	-1.73	-0.79	0.01	Qin prophage; putative antitermination protein Q
Z1768	-1.73	-0.79	7.02E-03	Unknown protein encoded by prophage CP-933N
Z4883	-1.73	-0.79	0.01	Type II toxin-antitoxin system HicA family toxin
nanA	-1.72	-0.78	0.02	*N*-acetylneuraminate lyase
yccT	-1.72	-0.78	8.29E-03	DUF2057 domain-containing protein
yebG	-1.72	-0.78	7.86E-03	DNA damage-inducible protein
yqgA	-1.72	-0.78	0.02	DUF554 domain-containing protein
Z_RS08615	-1.72	-0.78	0.02	PerC family transcriptional regulator
Z1781	-1.72	-0.78	8.30E-03	Unknown protein encoded by prophage CP-933N
Z4070	-1.72	-0.78	0.01	CRISPR-associated helicase/endonuclease Cas3
cysC	-1.71	-0.77	0.04	Adenylyl-sulphate kinase
idi	-1.71	-0.77	0.01	Isopentenyl-diphosphate Δ-isomerase
ygjG	-1.71	-0.77	8.78E-03	Putrescine aminotransferase
Z0967	-1.71	-0.77	0.03	Putative protease encoded in prophage CP-933K

Z1765	-1.71	-0.77	0.01	Putative excisionase for prophage CP-933N
allE	-1.7	-0.77	0.03	(S)-ureidoglycine aminohydrolase
anmK	-1.7	-0.77	0.02	Anhydro-*N*-acetylmuramic acid kinase
rna	-1.7	-0.77	0.01	RNase I
tdcG	-1.7	-0.77	0.05	L-serine ammonia-lyase
yebV	-1.7	-0.77	9.61E-03	DUF1480 domain-containing protein
yqhC	-1.7	-0.77	0.02	Putative AraC-type regulator protein
ytfB	-1.7	-0.77	0.01	Cell division protein
ascB	-1.69	-0.76	0.02	6-phospho-β-glucosidase
betI	-1.69	-0.76	0.02	DNA-binding transcriptional repressor
garD	-1.69	-0.76	0.02	Galactarate dehydratase
garK	-1.69	-0.76	0.02	Glycerate 2-kinase 1
malM	-1.69	-0.76	0.01	Maltose regulon periplasmic protein
sra	-1.69	-0.76	0.01	Ribosome-associated protein
ydcF	-1.69	-0.76	0.02	DUF218 domain-containing protein
ydhF	-1.69	-0.76	0.01	Hypothetical protein
yeeZ	-1.69	-0.76	0.01	Putative epimerase
yeiE	-1.69	-0.76	0.01	DNA-binding transcriptional dual regulator
yfdE	-1.69	-0.76	0.03	Hypothetical protein
yhbW	-1.69	-0.76	0.02	Putative luciferase-like monooxygenase
yhbX	-1.69	-0.76	0.02	Putative hydrolase
Z0266	-1.69	-0.76	0.02	Hypothetical protein
fur	-1.68	-0.75	0.01	Negative regulator
glnB	-1.68	-0.75	0.03	Nitrogen regulatory protein PII
hcaT	-1.68	-0.75	0.01	Putative 3-phenylpropionate transporter
hokD_2	-1.68	-0.75	0.01	Type I toxin-antitoxin system toxin
mocA	-1.68	-0.75	0.01	Molybdenum cofactor cytidylyltransferase
oppF	-1.68	-0.75	0.03	Murein tripeptide ABC transporter
pfkB	-1.68	-0.75	0.01	6-phosphofructokinase II
yhaV	-1.68	-0.75	0.01	Ribosome-dependent mRNA interferase toxin
Z_RS32620	-1.68	-0.75	0.02	NadS family protein
Z1772	-1.68	-0.75	0.01	Unknown protein encoded by prophage CP-933N
espZ	-1.67	-0.74	0.02	T3SS effector
hslU	-1.67	-0.74	0.01	ATPase component of the HslVU protease
pmbA	-1.67	-0.74	0.02	Metalloprotease subunit
recN	-1.67	-0.74	0.01	DNA repair protein
speB	-1.67	-0.74	0.02	Agmatinase
Z1776	-1.67	-0.74	0.01	Unknown protein encoded by prophage CP-933N
Z5151	-1.67	-0.74	0.02	Hypothetical protein
gntP	-1.66	-0.73	0.03	Fructuronate transporter
kdgT	-1.66	-0.73	0.04	2-dehydro-3-deoxy-D-gluconate:H^+symporter
metL	-1.66	-0.73	0.02	Aspartokinase II and homoserine dehydrogenase II
yaaU	-1.66	-0.73	0.02	Putative transporter
yccM	-1.66	-0.73	0.04	Putative electron transport protein
yjeH	-1.66	-0.73	0.02	L-methionine/branched chain amino acid exporter
yrdA	-1.66	-0.73	0.02	Putative transferase
zntR	-1.66	-0.73	0.02	DNA-binding transcriptional activator
ataT	-1.65	-0.72	0.02	Hypothetical protein
brnQ	-1.65	-0.72	0.02	Branched chain amino acid transporter
nanM	-1.65	-0.72	0.02	*N*-acetylneuraminate mutarotase
rimK	-1.65	-0.72	0.03	Ribosomal protein S6 modification protein
ybiU	-1.65	-0.72	0.03	DUF1479 domain-containing protein
ydcU	-1.65	-0.72	0.05	Putative ABC transporter membrane subunit
yjdN	-1.65	-0.72	0.02	PF06983 family protein
yqjC	-1.65	-0.72	0.02	DUF1090 domain-containing protein
ytfF	-1.65	-0.72	0.03	Inner membrane protein
Z0956	-1.65	-0.72	0.02	Putative anti-terminator Q protein of prophage CP-933K
Z1773	-1.65	-0.72	0.02	Unknown protein encoded by prophage CP-933N
flgK	-1.64	-0.71	0.04	Flagellar hook-filament junction protein 1
hokD_4	-1.64	-0.71	0.02	Type I toxin-antitoxin system toxin
pepT	-1.64	-0.71	0.02	Putative peptidase T
ydcH	-1.64	-0.71	0.02	Hypothetical protein
yeaO	-1.64	-0.71	0.03	DUF488 domain-containing protein
yeeA	-1.64	-0.71	0.02	Putative transporter

Z1778	-1.64	-0.71	0.02	Unknown protein encoded by prophage CP-933N
Z3230	-1.64	-0.71	0.03	Hypothetical protein
dcuA	-1.63	-0.70	0.02	C4-dicarboxylate transporter
ldcA	-1.63	-0.70	0.03	Murein tetrapeptide carboxypeptidase
usg	-1.63	-0.70	0.02	Putative semialdehyde dehydrogenase
ycgM	-1.63	-0.70	0.02	Putative isomerase/hydrolase
ycjX	-1.63	-0.70	0.03	DUF463 domain-containing protein
yfhM	-1.63	-0.70	0.02	α2-macroglobulin
Z_RS12145	-1.63	-0.70	0.02	Type II toxin-antitoxin system RelE/ParE family
Z1771	-1.63	-0.70	0.02	Rac prophage; putative DNA-binding transcriptional regulator
Z1774	-1.63	-0.70	0.02	Unknown protein encoded by prophage CP-933N
Z1775	-1.63	-0.70	0.02	Unknown protein encoded by prophage CP-933N
clpP	-1.62	-0.70	0.02	ATP-dependent Clp protease proteolytic subunit
curA	-1.62	-0.70	0.03	NADPH-dependent curcumin/dihydrocurcumin reductase
rplT	-1.62	-0.70	0.02	50S ribosomal subunit protein L20
rraA	-1.62	-0.70	0.03	Ribonuclease E inhibitor protein A
sapC	-1.62	-0.70	0.03	Peptide ABC transporter permease
yceK	-1.62	-0.70	0.02	DUF1375 domain-containing lipoprotein
miaA	-1.61	-0.69	0.02	Delta(2)-isopentenylpyrophosphate tRNA-adenosine transferase
yaeP	-1.61	-0.69	0.03	Hypothetical protein
ycbJ	-1.61	-0.69	0.04	Putative phosphotransferase
ychN	-1.61	-0.69	0.03	DsrE/F sulphur relay family protein
yjjI	-1.61	-0.69	0.03	DUF3029 domain-containing protein
Z_RS08520	-1.61	-0.69	0.04	T3SS effector
Z_RS31570	-1.61	-0.69	0.04	Hypothetical protein
Z1764	-1.61	-0.69	0.02	Partial integrase for prophage CP-933N
dprA	-1.6	-0.68	0.03	NAD(P)H-dependent nitroreductase
elaB	-1.6	-0.68	0.03	Tail-anchored inner membrane protein
msrA	-1.6	-0.68	0.03	Peptide methionine sulfoxide reductase
sapD	-1.6	-0.68	0.03	Putative ATP-binding protein of peptide transport system
treF	-1.6	-0.68	0.03	Cytoplasmic trehalase
astA	-1.59	-0.67	0.03	Arginine *N*-succinyltransferase
clpA	-1.59	-0.67	0.03	Serine-protease ATP-binding component
cyuA	-1.59	-0.67	0.03	Putative L-cysteine desulphidase
dnaG	-1.59	-0.67	0.03	DNA primase
espN	-1.59	-0.67	0.03	T3SS effector
hfq	-1.59	-0.67	0.03	RNA binding protein
kdgA	-1.59	-0.67	0.03	KHG/KDPG aldolase
metJ	-1.59	-0.67	0.05	DNA-binding transcriptional repressor
uxuA	-1.59	-0.67	0.03	D-mannonate dehydratase
yaeR	-1.59	-0.67	0.04	VOC domain-containing protein
dedD	-1.58	-0.66	0.03	Cell division membrane protein
lgoR	-1.58	-0.66	0.03	Hypothetical protein
tal	-1.58	-0.66	0.03	Transaldolase
yicC	-1.58	-0.66	0.05	Putative RNAse adapter protein
bcr	-1.57	-0.65	0.04	Multidrug efflux pump
lldP	-1.57	-0.65	0.04	Lactate/glycolate:H^+ symporter
sufC	-1.57	-0.65	0.04	Fe-S cluster assembly ATPase
ydcY	-1.57	-0.65	0.04	DUF2526 domain-containing protein
yohD	-1.57	-0.65	0.05	DedA family protein
Z1777	-1.57	-0.65	0.03	Unknown protein encoded by prophage CP-933N
amiD	-1.56	-0.64	0.04	*N*-acetylmuramoyl-L-alanine amidase D
espK	-1.56	-0.64	0.04	T3SS effector
gltS	-1.56	-0.64	0.04	Glutamate-sodium symporter
lpxC	-1.56	-0.64	0.04	UDP-3-O-acyl *N*-acetylglucosamine deacetylase
yidE	-1.56	-0.64	0.04	Putative transport protein
ybjP	-1.55	-0.63	0.04	DUF3828 domain-containing lipoprotein
ygeV	-1.55	-0.63	0.04	Putative σ54-dependent transcriptional regulator
ymdB	-1.55	-0.63	0.05	2'-*O*-acetyl-ADP-ribose deacetylase, regulator of RNase III activity
ada	-1.54	-0.62	0.05	DNA-binding transcriptional dual regulator
rlmE	-1.54	-0.62	0.04	23S rRNA 2'-*O*-ribose U2552 methyltransferase
ybeY	-1.54	-0.62	0.05	Endoribonuclease
yihY	-1.54	-0.62	0.05	PF03631 family membrane protein