IDENTIFICATION OF *ESCHERICHIA COLI* GENES REQUIRED FOR BACTERIAL SURVIVAL AND MORPHOLOGICAL PLASTICITY IN URINARY TRACT INFECTIONS

DANIEL GIUSEPPE MEDIATI

A thesis submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

The ithree institute and School of Life Sciences,
University of Technology Sydney

November 2018

[blank page]

CERTIFICATE OF AUTHORSHIP AND ORIGINALITY

I, Daniel G. Mediati, certify that the work in this PhD thesis, carried out between March 2014 to July 2018, has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledge within the text.

I also certify that this PhD thesis has been written by me. Any assistance that I have received in my research work and preparation of this thesis has been acknowledged. In addition, I certify that all information sources and literature used are indicated in this thesis.

This PhD was supported by an Australian Government Research Program Scholarship, formerly known as the Australian Postgraduate Award Scholarship until 2017.

Signature of candidate:

Production Note:
Signature removed prior to publication.

Daniel Giuseppe Mediati

Date: 30/11/2018

TO MUM AND DAD,

WHO INSTILLED IN ME THE GREAT VALUE OF

EDUCATION.

Contents in Brief

TABLE OF CONTENTS

List of Figures and Supplementary Data

List of Tables and Supplementary Data

The following Supplementary Data Tables can be accessed online with this thesis.

ABBREVIATIONS

A	Absorbance
Amp	Ampicillin
BLAST	Basic local alignment search tool
bp	Basepair
Cm	Chloramphenicol
Δ	Delta (change in)
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTI	Defined trypsin inhibitor
g	Centrifugal force (x unit gravitational field)
GFP	Green fluorescent protein
h	Hours
IBC	Intracellular bacterial community
kb	Kilobase
Km	Kanamycin
L	Liters
LB	Luria Broth
M	Molarity (mol/L)
m	Meters
μ	micro (10^{-6})
Min	The Min protein system
min	Minutes
mL	Mililiters
MOI	Multiplicity of infection
ms	Miliseconds
MW	Molecular weight
n	nano (10^{-9})
NO	Nucleoid occlusion system
OD	Optical density
ORF	Open reading frame
ρ	Density (Rho)
PAI	Pathogenicity island
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
QIR	Quiescent intracellular reservoir
RNA	Ribonucleic acid
RT	Room temperature
rpm	Revolutions per minute

TBE	Tris/borate/EDTA
TraDIS	Transposon-directed insertion-site sequencing
UPEC	Uropathogenic *Escherichia coli*
UTI	Urinary tract infections
V	Volts
v	Volume
w	Weight
WT	Wild-type

PUBLICATIONS

This doctoral thesis is a compilation of the work carried out as part of the project "Identification of *Escherichia coli* genes required for bacterial survival and morphological plasticity in urinary tract infections" which resulted in the following written works:

Mediati, D.G., Iosifidis, G. (2016) Lifestyle dynamics of uropathogenic *Escherichia coli* in the urinary tract. *ASM Syntrophy*, 17:3.

Mann, R., Mediati, D.G., Duggin, I.G., Harry, E.J., & Bottomley, A.L. (2017) Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. *Frontiers in Cellular and Infection Microbiology*, 7:241.

Mediati, D.G., Burke, C.M., Ansari, S., Harry, E.J., Duggin, I.G. (2018) High-throughput sequencing of sorted expression libraries reveals inhibitors of bacterial cell division. *BMC Genomics*, 19:781.

Mediati, D.G., Monahan, L.G., Charles, I.G., Duggin, I.G. (2018) Genome-wide comparison of *E. coli* genes required for growth in complex and minimal media. *Journal of Bacteriology*, (manuscript to be submitted).

CONFERENCE PROCEEDINGS

This doctoral thesis is a compilation of the work carried out as part of the project "Identification of *Escherichia coli* genes required for bacterial survival and morphological plasticity in urinary tract infections" which resulted in the following presentations:

East Coast Bacillus Meeting, Wollongong, NSW, Australia 2014
Oral Presentation
Authors: Mediati, D.G., Iosifidis, G., Duggin, I.G.

40th Annual Infection and Immunity, Lorne, VIC, Australia 2015
Poster Presentation
Authors: Mediati, D.G., Duggin, I.G.

13th Annual Bacterial Pathogenesis, Phillip Island, VIC, Australia 2015
Oral Presentation
Authors: Mediati, D.G., Burke, C., Ansari, S., Harry, L., Duggin, I.G.

University of Technology Sydney: 3-minute thesis competition 2015
Oral Presentation
Authors: Mediati, D.G., Duggin, I.G.

East Coast Bacillus Meeting, Wollongong, NSW, Australia 2015
Oral Presentation
Authors: Mediati, D.G., Burke, C., Ansari, S., Harry, L., Duggin, I.G.

Becton-Dickinson (BD) Award Final, Sydney, NSW, Australia 2016
Oral Presentation
Authors: Mediati, D.G., Burke, C., Ansari, S., Harry, E., Duggin, I.G.

East Coast Bacillus Meeting, Sydney, NSW, Australia 2016
Oral Presentation
Authors: Mediati, D.G., Monahan, L.G., Charles, I.G., Duggin, I.G.

14th Annual Bacterial Pathogenesis, Adelaide Hills, SA, Australia 2017
Poster Presentation
Authors: Mediati, D.G., Monahan, L.G., Charles, I.G., Duggin, I.G

ACKNOWLEDGEMENTS

First and foremost acknowledgement goes to Associate Professor Iain Duggin, for supervision of the work contributing to this PhD thesis. I cannot thank you enough for believing in me and giving me the opportunity to join your research team. You have been an excellent mentor and friend to learn from. Thank you for all your advice and support over the years.

I am also grateful for help from the following people: Leigh Monahan, for all your advice on the TraDIS technique and discussions on *E. coli* metabolism; Michael Liu, for all your assistance and guidance on sequencing and DNA library preparations; Greg Iosifidis, who taught me everything about mammalian tissue culturing and microscopy; Catherine Burke, who played a central role in the flow cytometry sorting and analysis results within Chapter 3; Elizabeth Harry, for all your insights into antimicrobial resistance and, together with the members of the Harry lab, who always provided great discussion and insight into bacterial cell division.

Finally, thank you to my parents and family. Mum and Dad, you have always encouraged and supported all of my academic endeavors. Without your guidance and love, I would have never had the courage to overcome the adversities I have faced in life. This PhD would not have been accomplished without the both of you.

PREFACE

Urinary tract infections (UTIs) are the second most common bacterial infectious disease affecting humans, after pneumonia. A range of pathogens have been implicated in causing UTIs, however strains of uropathogenic *Escherichia coli* (UPEC) are the predominant etiological agents. UPEC originate within the intestine, but have adapted the ability to disseminate and colonise the human urinary tract via a multi-stage intracellular infection cycle within the cells of the bladder. This infection cycle is a complex pathway involving epithelial cell attachment, invasion and intracellular biofilm-like proliferation, leading to the formation of a sub-population of filamentous bacteria. Bacterial filamentation occurs when rod-shaped cells grow without dividing. This filamentation accompanies bacterial dispersal and the rupture of the host bladder cell. UPEC filaments have the potential to revert to bacillary, rod-shaped morphology and can thereafter divide as normal bacterial cells, thereby initiating a new infection cycle. There is also the potential for UPEC to ascend to the kidneys and enter the bloodstream causing urosepsis.

The rapid emergence of antibiotic resistance has greatly influenced the severity of UTIs and financial burden on the health-care sector. This further complicates UTI therapies and highlights the urgent need to advance our understanding of the biological mechanisms and requirements underpinning UPEC survival and pathogenesis.

The work carried out for this PhD thesis aimed to expand our existing knowledge surrounding UPEC survival and morphological plasticity (cell shape changes). Each Chapter within this thesis focused on distinct, but related aspects of this overarching aim. Firstly, we developed and applied a high-throughput sequencing-based method for the genome-wide identification of genes and genomic DNA fragments that induce filamentation in *E. coli*. This revealed genes from several prophages, carbon metabolic pathways, as well as endogenous bacterial genes or loci that have known and novel roles in cell division or bacterial filamentation. A large number of short predicted peptides that trigger filamentation were also identified for the first time, which are not expected to have evolved with the purpose of causing filamentation, but could be used as synthetic, artificial inhibitors of cell division.

A transposon-insertion mutant library in the uropathogenic *E. coli* cystitis isolate UTI89 was constructed at a moderate scale. We combined this library with a modified transposon-directed insertion-site sequencing (TraDIS) technique to define the genes required for UPEC growth and survival in M9-glycerol minimal medium compared to a rich LB medium. We identified 60 mutants with a significant fitness defect and reduced capacity to survive in the M9-glycerol, the majority of which encode gluconeogenic and amino acid catabolism proteins. We also highlight novel differences and several apparent discrepancies in the metabolic requirements between uropathogenic and commensal *E. coli*. Several uncharacterised and UPEC-specific genes were identified that likely underlie metabolic capacities of UPEC strains during infection. Two of these genes, *neuC* and *hisF* were confirmed and verified as important during UPEC infection of human bladder epithelial cells *in vitro*.

Finally, we employed TraDIS to an up-scaled *in vitro* UTI model to identify the UTI89 genes required for survival at distinct phases of the multi-stage bladder infection cycle, focusing particularly on the later infection events, when bacterial physiology appears to most substantially change. This revealed a total of 143, 333 and 322 statistically significant genes required for the IBC, dispersal and recovery phases of bladder cell infection. We additionally characterised the extent and distribution of UPEC filamentation in the dispersal phase of our up-scaled model and implicate, through TraDIS, known and novel cell division regulators to potentially be involved in the UTI-filamentation response pathway. Further, the catalog of genes identified through our TraDIS experiments provides a foundation for further characterisation of UPEC factors needed for survival, both in laboratory culture conditions and during human bladder infection.

The extensive gene functional identification reported in this PhD thesis represents both a substantial resource and some of the first steps towards the understanding of UPEC gene function during infection. This is expected to facilitate the development of treatments for UTIs or other bacterial infections in a future faced with increasing resistance to current antibiotics.

CHAPTER 1

INTRODUCTION

DISCLOSURE

I acknowledge that a proportion of this chapter has been published in the *Frontiers of Cellular and Infection Microbiology* journal as a review article under the title "Metabolic adaptations of uropathogenic *E. coli* in the urinary tract" (refer to page *xiv* for list of publications). I certify that I carried out significant work presented in this manuscript.

- Riti Mann and Daniel G. Mediati wrote the first draft and edited all subsequent drafts of the review.
- Iain G. Duggin, Elizabeth J. Harry and Amy L. Bottomley revised and proof-read the manuscript.
- Amy L. Bottomley and Riti Mann conceived the project idea.

1.1 URINARY TRACT INFECTIONS

Urinary tract infections (UTIs) are among the most frequent and significant bacterial infections worldwide (*1*). Annually, in the developed world, approximately 150 million cases are reported (*2*). Men and women of all ages can develop a UTI, however women are significantly more at risk with 50% experiencing a UTI during their lifetime (*3, 4*). Furthermore, individuals with catheters installed, immuno-compromised and those with prior urological conditions are more susceptible to infection (*5*).

UTIs are classified as either complicated or uncomplicated on the basis of antimicrobial therapy given to the patient (*5*). Uncomplicated UTIs typically affect those who are otherwise healthy and have no structural urinary tract abnormalities (*6*). These are differentiated into upper and lower UTIs, most commonly pyelonephritis (kidney) and cystitis (bladder), respectively. On the other hand, complicated UTIs are associated with urinary tract abnormalities such as urinary obstruction, retention, immunosuppression and renal failure. These require a prolonged therapy and are associated with an increased risk of recurrent infections in the future (*7*). Together, these factors compromise the urinary tract and increase the risk of serious complications and treatment failure, leading to kidney disease and life threatening sepsis (*7*). Sepsis kills approx. 50,000 patients in the UK every year, and 25% of all sepsis cases are of UTI origin (*8*).

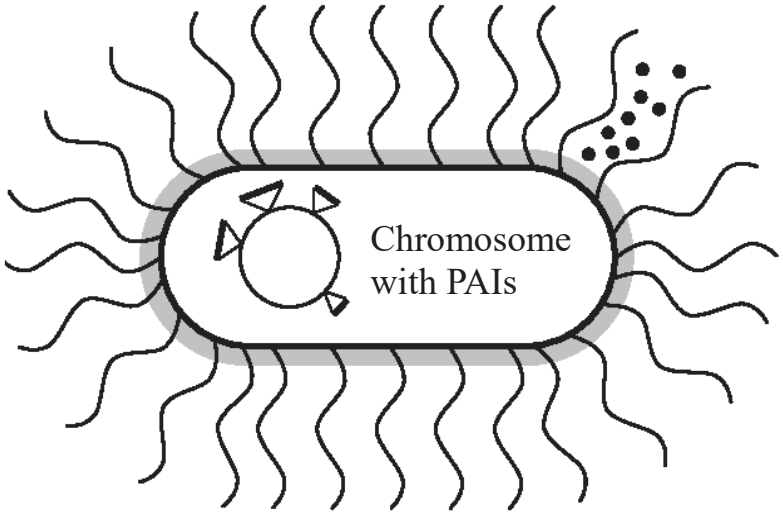
In recent times, the rapid emergence of antibiotic-resistant UTIs has greatly influenced the severity of infection and financial cost burden to the health-care system, costing up to USD$6 billion on treatment each year (*5, 9*). With the ongoing rise in antibiotic-resistant UTIs, the severity of infection and amounting burden on the health-care sector is set to escalate (*4*). This further complicates UTI therapies and highlights the urgent need to advance our understanding of the biological mechanisms underpinning bacterial survival and pathogenesis.

1.1.1 UROPATHOGENIC ESCHERICHIA COLI

In 1885, the Austrian bacteriologist Dr. Theodor Escherich identified the agent causing diarrhea and gastroenteritis in children as the colon bacillus bacteria, now known as the Gram-negative bacterial genus *Escherichia* (*10*). Within this genus, *Escherichia coli* (*E. coli*), as well as a range of other pathogens have been implicated in causing UTIs, including both Gram-positive and Gram-negative bacteria such as *Proteus mirabilis* and *Staphylococcus saprophyticus* (*11-13*). However *E. coli*, or more specific strains known as uropathogenic *E. coli* (UPEC), are the predominant etiological agent accountable for up to 90% of all UTIs (*14*). UPEC are a pathotype of extraintestinal pathogenic *E. coli* and originate from the intestinal microbiome (*5*). Within the intestine, UPEC rarely cause complications and exist in a beneficial symbiotic relationship within the microbial flora (*15*). However, UPEC have adapted the ability to disseminate and colonise other host environments such as the urinary tract, kidneys and bloodstream in a pathogenic nature (*16*). This is carried out via an arsenal of virulence factors such as toxins, invasins and adhesins, which play specific roles in a multi-stage infection cycle enabling host colonisation (*17*) (Fig. 1.1). Majority of these virulence factors are concentrated at particular genetic regions termed pathogenicity-associated islands (PAIs), which most strains of commensal *E. coli* (*18*). Furthermore, physiological factors that do not directly damage the host but nevertheless are required for UPEC survival in the urinary tract are now being appreciated for their role in pathogenesis (*19, 20*). The capacity to utilise diverse nutrients in nutritionally-distinctive environments such as the intestines, urine, bladder, kidney and bloodstream clearly plays a significant role in UPEC pathogenesis (*21*). This PhD thesis concentrates on the genetic responses that allow UPEC to survive and cause infection.

Adhesins
- Type-1 pili
- P-fimbriae

Toxins
- α-hemolysin
- Cytotoxic necrotizing factor 1 (*cnf1*)



Immune evasion
- LPS
- Capsular antigens
- Yersiniabactin

Metabolic
- Iron acquisition
- Salmochelin

Other
- Antigen 43
- Flagella

Figure 1.1 Well characterised virulence factors of uropathogenic E. coli

Type-1 pili and P-fimbriae adhesins (which are discussed below) are present on the cell surface and help UPEC establish a stronghold in the bladder. Toxins, immune evasion factors and flagella play specific roles during UPEC proliferation resulting in intracellular bacterial communities (IBCs) within the bladder cell cytoplasm. Within the bladder cell, UPEC have great metabolic flexibility, and are able to scavenge iron and other metabolites from the host in order to cause infection. These virulence factors are concentrated at particular genetic regions termed pathogenicity-associated islands (PAIs). Adapted from Svanborg et al. (2006) Curr Opin Microbiol (*22*).

1.2 UPEC Lifestyle from the Intestine to the Urinary Tract

The gastrointestinal tract is considered the primary reservoir of UPEC in humans (*15*). For successful colonisation in the colon, UPEC must first survive passage through the acidic conditions of the stomach and penetrate the epithelial cell mucus layer (*23*). This intestinal mucus layer provides UPEC with major resources of nutrients – mucins, polysaccharides and glycoproteins (*24*). Additionally, UPEC circumvent host defense mechanisms and outcompete other members of mixed microbiota for the acquisition of nutrients (*21*). Metabolic and genetic (gene expression) responses, which are detailed further below, are altered accordingly when shifting to the urinary tract to accommodate different nutrient availability. This allows UPEC to effectively utilise multiple types of available metabolites under such competitive and fluctuating conditions.

UTIs are initiated when UPEC, from the gastrointestinal tract, contaminate the periurethral region and colonises the urethra (*5*). Subsequently, UPEC migrate into the bladder lumen where they invade the bladder epithelium. Previously, UPEC was viewed strictly as an extracellular pathogen that merely attached to the bladder epithelium or grew freely in urine (*25*). However, UPEC strains are now known, due to the last 25 years of research, to be highly invasive intracellular pathogens that undergo a multi-stage infection cycle (*17, 26, 27*). This infection cycle is a complex pathway involving epithelial cell attachment, invasion and intracellular proliferation, leading to the eventual rupture of the bladder epithelial cell and UPEC dissemination, including to neighbouring host cells (*17, 26*). In recurrent UTIs, UPEC persist in a dormant reservoir state in the underlying epithelial cells, termed quiescent intracellular reservoirs (QIR), awaiting as yet unknown cellular cues that appear to reinvigorate UPEC into a pathogenic growth state (*25*). As mentioned previously, in complicated UTIs the potential exists for UPEC to progress to the kidneys and enter the bloodstream causing sepsis (*16*), as shown in Figure 1.2.

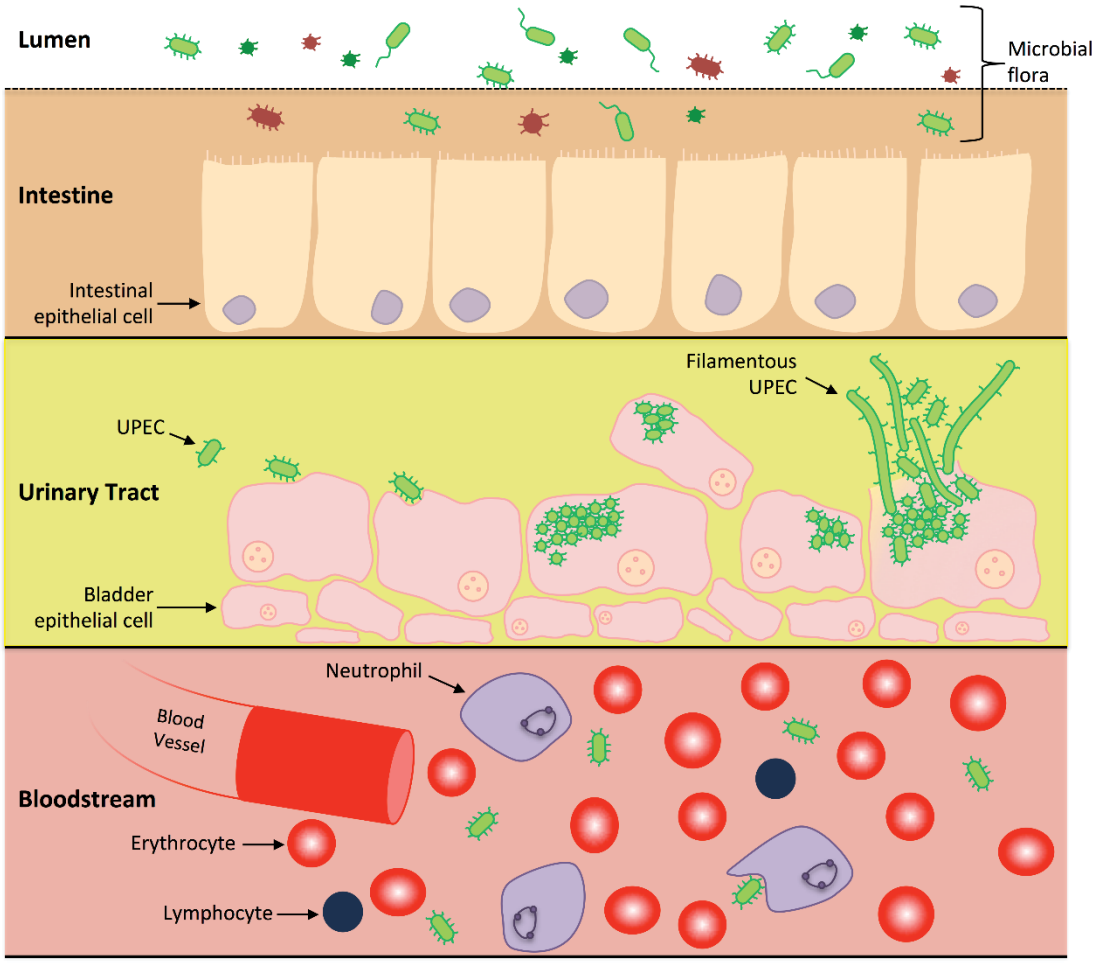


Figure 1.2 UPEC pathogenesis from the intestine to the urinary tract and bloodstream

UPEC grow in the human intestine as part of the microbiome. Within this environment, UPEC interact with the intestinal epithelial cells in a symbiotic relationship typical of commensal E. coli strains, however there is competition for nutrients between other microorganisms (A). UPEC has also adapted the ability to cause urinary tract infections and urosepsis by transitioning to a pathogenic lifecycle in the urinary tract (B) and bloodstream (C). To gain a stronghold within the urinary tract, UPEC express numerous pili systems to facilitate attachment to the superficial bladder epithelial cell layer. Invasion into the host cell initiates replication immediately to form intracellular bacterial communities (IBCs) and a subpopulation subsequently undergoes a dispersal phase involving host cell rupture and release of motile bacteria and extensive cell elongation (filamentation). UPEC can then infect neighboring host epithelia to continue the infective cycle. Adapted from Mann et al. (2017) Front Cell Infect Microbiol (28).

1.3 Infection Cycle Events in the Bladder

The intracellular infection cycle of UPEC *in vivo* was first visualised in infected mouse bladder explants using time-lapse microscopy (*29*). Analysis of urine samples taken from patients (*30*) and a human flow chamber-based cell culture model (*26*) support the infection cycle observed in mice. During the cycle, UPEC undergo various morphological changes resulting in an elongated filamentous phenotype and subsequent reversal to bacillary rod-shape, which aid in their survival and pathogenicity of the urinary tract (*31*) (Fig 1.2).

1.3.1 Bacterial Attachment and Invasion of Host Cells

Many studies have focused on the virulence factors that play a role in mediating UPEC colonisation of the bladder, most notably the initial attachment of bacterial adhesive structures to the luminal surface of the epithelium (Fig. 1.3A). This step of infection has been well characterized, since 1976 by Eden *et al.* (*32*) and has been a focus of attention in the development of therapies to treat UTIs (*33*).

Type-1 pili and the adhesin structural protein FimH, recognise mannosylated glycoprotein receptors present on the host epithelium, mediating UPEC binding (*27, 34*) (Fig. 1.3B). The adhesin, FimH, positioned on the pilus tip (Fig. 1.3C-D) can be found on the surface of a range of *E. coli* pathogens capable of host epithelial attachment. FimH has also been documented to interact with α3/β1 subunits of epithelial cell integrins (*35*), fibronectins (*36*) and uroplakins (*37*). In fact, disruption of the β1 subunit of the integrin gene, and blocking of both the α3/β1 subunits by integrin-specific antibodies results in an attenuated invasion capacity of UPEC (*35*).

Along with Type-1 pili, UPEC express a variety of additional fimbriae on their surface, which facilitate bacterial attachment (*38*). Genetic screening analysis revealed the presence of ten additional pili systems found on the surface of UPEC (*39*). This large inventory is suspected to give UPEC multiple binding specificities and the capacity to colonise both the kidneys and bladder. As an example, the P-fimbriae have an essential role in the adhesion

to kidney mucosa during ascending UTIs (*40*). P-fimbria were among the first factors associated with UPEC virulence (encoded by the *pap* operon) and can also interact and bind to Gal(α1-4)Galβ moieties present on red blood cells and bladder epithelial cells (*40, 41*).

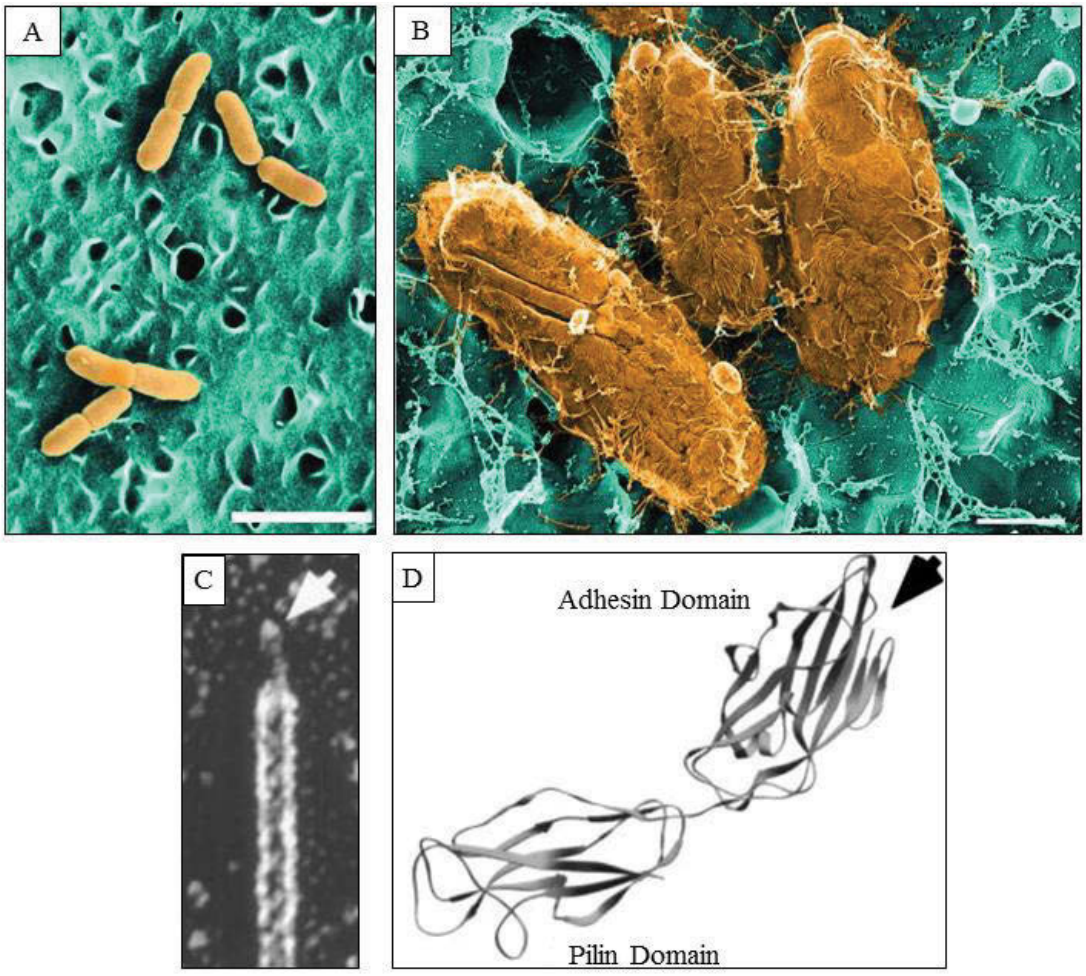


Figure 1.3 Type-1 pilus and the adhesin protein FimH-mediated binding

(A-B) Scanning electron microscopy demonstrates the attachment of Type-1 pili-containing UPEC (yellow) to the bladder epithelium of the mouse model of UTI (false-colour, green). (C-D) Structure and architecture of the Type-1 pilus of UPEC. Electron microscopy of the tip of the pilus (white arrowhead) (C), and the structural fold of the pilin domain and the adhesin domain, of the FimH protein (D). The binding region (black arrowhead) of the adhesin domain is capable of interacting with epithelial cell integrins, fibronectins and D-mannose. (Bars = 3µm (A), 5µm (B)) Adapted from Mulvey et al. (2000) PNAS (42).

Once bound, UPEC become internalised via endocytosis (*43*); activation of the Rho-family proteins, such as RhoA, Cdc42 and Rac1 modulate host actin dynamics and result in the rearrangement of the host cell membrane. This rearrangement leads to the host membrane undergoing a zipper-like mechanism, engulfing the attached bacteria (*43*). This internalisation process requires additional host membrane, for example lysosomes, vesicles and endosomes that are transported to the host cell surface to accommodate the zipper-like mechanism (*44 389*). The additional membrane is transported to the host cell surface via small GTP-binding proteins, in particular Rab27b (*45*). Internalised UPEC initially localise within these membrane-bound compartments or vacuoles, which are similar to late endosomes (*27*) (Fig. 1.4). However, growth is restricted due to the overwhelming host actin filament, which surround the vacuole and limit nutrient trafficking and uptake (*46*). Over time, damage to the actin cytoskeleton and endosomal membrane of some of the vacuoles results in the release of UPEC to the host cytoplasm where rapid growth can occur (*47*). However, the complete mechanism of escape into the host cytoplasm is not clear (*48*). Those UPEC-containing vacuoles that have a more dense assembly of host actin filament can become encased within and enter a quiescent-like state for long periods of time. These vacuoles do however offer protection from the host immune response and give UPEC a safe haven (Fig. 1.4) from the challenges associated with the extracellular environment, including the shear flow of urine and exposure to antibacterial molecules (*49*).

When inside the bladder cell, virulence gene expression needs to be coordinated. The *fnr* regulator, which contains a DNA-binding domain, controls gene expression within the bladder environment during the attachment and invasion stages of infection (*50*). The deletion of *fnr* within UPEC has shown a decreased adherence and subsequent invasion capacity within the epithelial cells of the kidneys and bladder when compared to wild-type (*50*). This is due to multiple virulence factor modifications, such as a reduced expression of Type-1 pili and P-fimbriae (*50*). In addition to this, the *fnr* homolog within Salmonella is upregulated during flagella synthesis and motility in intracellular infection (*51*). This highlights a possible regulatory mechanism in UPEC during bladder cell attachment and internalisation.

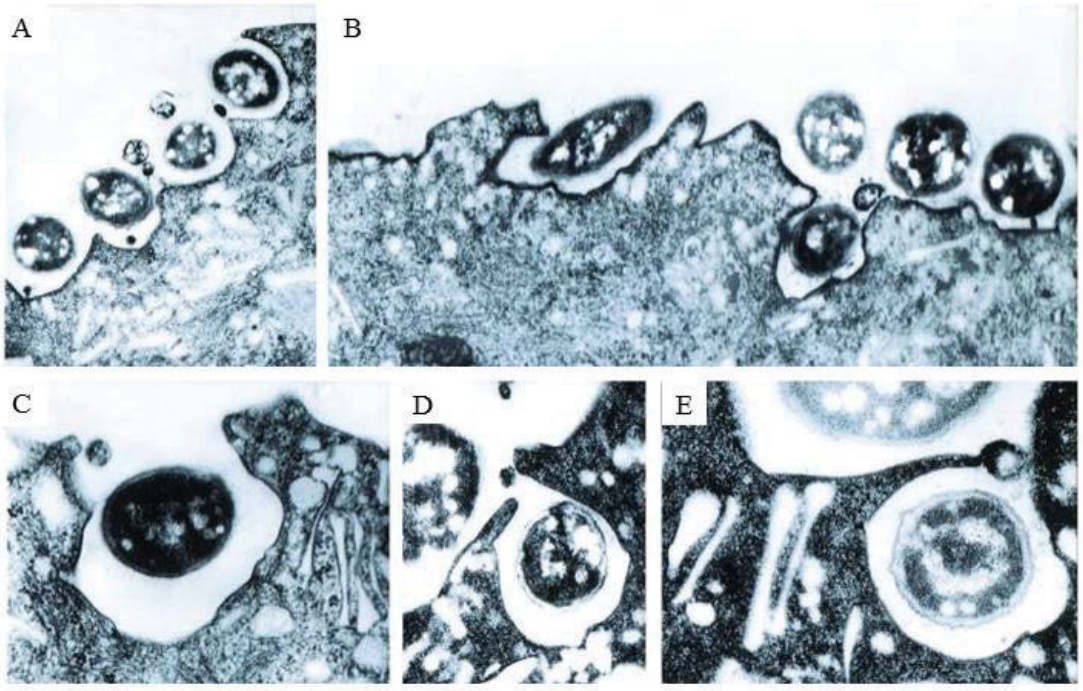


Figure 1.4 Mouse bladder epithelial cell endocytosis of UPEC

(A-E) Transmission electron microscopy of UPEC being internalised by mouse bladder epithelial cells by enveloping the bacteria via endocytosis. The infected mouse bladders were recovered 1 h post-infection. Adapted from Mulvey et al. (2000) PNAS (42).

1.3.2 INTRACELLULAR BACTERIAL COMMUNITIES

Within the cytoplasm, UPEC grow and divide, initially appearing as several individual bacteria, but soon form dense clusters encased within a polysaccharide-rich matrix, that excludes the host cell cytoskeleton (*52*). These biofilm-like clusters, known as intracellular bacterial communities (IBCs) have been observed in the murine model (*52*) (Fig. 1.5) and the human flow chamber-based cell culture model of UTI (*26*). In the murine model, over 500 individual IBCs have been observed within a single mouse bladder 12 h post-infection (*53*). The IBC begins as rod-shaped bacteria loosing organised within the host cytosol that multiply over time, producing small coccoid daughter cells that are tightly assembled (*17, 52*). This tightly assembled community enables higher loads of bacteria to be packed within each bladder cell (*17, 52*).

Both Justice *et al.* (*17*) and Anderson *et al.* (*52*) demonstrated a synchronous maturation process of the IBC that was divided into 3 stages (early, middle and late) that eventually form large bacterial "pods" (Fig. 1.5A-B**)**. Anderson *et al.* (*52*) demonstrated a lack of IBCs and protruding pods in the non-pathogenic *E. coli* strain MG1655 (*54*) when used to infect mouse bladders *in vivo* (Fig. 1.5C), showing that while capable of binding and entering host bladder epithelial cells, the non-pathogenic *E. coli* K-12 strain is unable to proliferate as substantially in host cells compared to UPEC strains.

Genes involved in the biogenesis of Type-1 and P-fimbriae are down-regulated in biofilms of UPEC pyelonephritis isolate CFT073 (*39*) grown in human urine, indicating that these fimbriae are not essential for biofilm formation, at least during growth in urine (*55*). Additionally, iron binding proteins such as the outer membrane receptors ChuA and Hma, as well as iron sequestering siderophores, synthesised by proteins IroB and IroE were also upregulated within the IBC (*56*). This highlights the necessity of iron uptake and transport systems within an IBC (see *Section 1.4.4* for more detail). Additionally, Antigen 43 (Ag43), an autotransporter protein on the bacterial surface that promotes autoaggregation, is highly expressed within the IBC (*52*). Toxins such as α-hemolysin and cytotoxic necrotizing factor 1 (*cnf1*) are secreted by UPEC during this time. Studies have demonstrated that secretion

of *cnf1* limits phagocytosis of UPEC by polymorphonuclear leukocytes, further promoting a resilient bacterial community within the bladder (*57*). Additionally, *in vivo* studies have shown that at least seven different functional classes of antibiotics are ineffective at eradicating UPEC from within the bladder cytoplasm once they have established an IBC (*9*).

Within the IBC, flagella facilitate weak surface contact inducing organised growth and subsequent motility (*17*). At some advanced point, the coccoid bacteria on the outer edge of the IBC differentiate into highly-motile and filamentous cells, which dissociate from the IBC (*17*). These highly-motile and filamentous cells were observed within the bladder cell cytoplasm and dispersing out from the host cells into the bladder lumen (*17*).

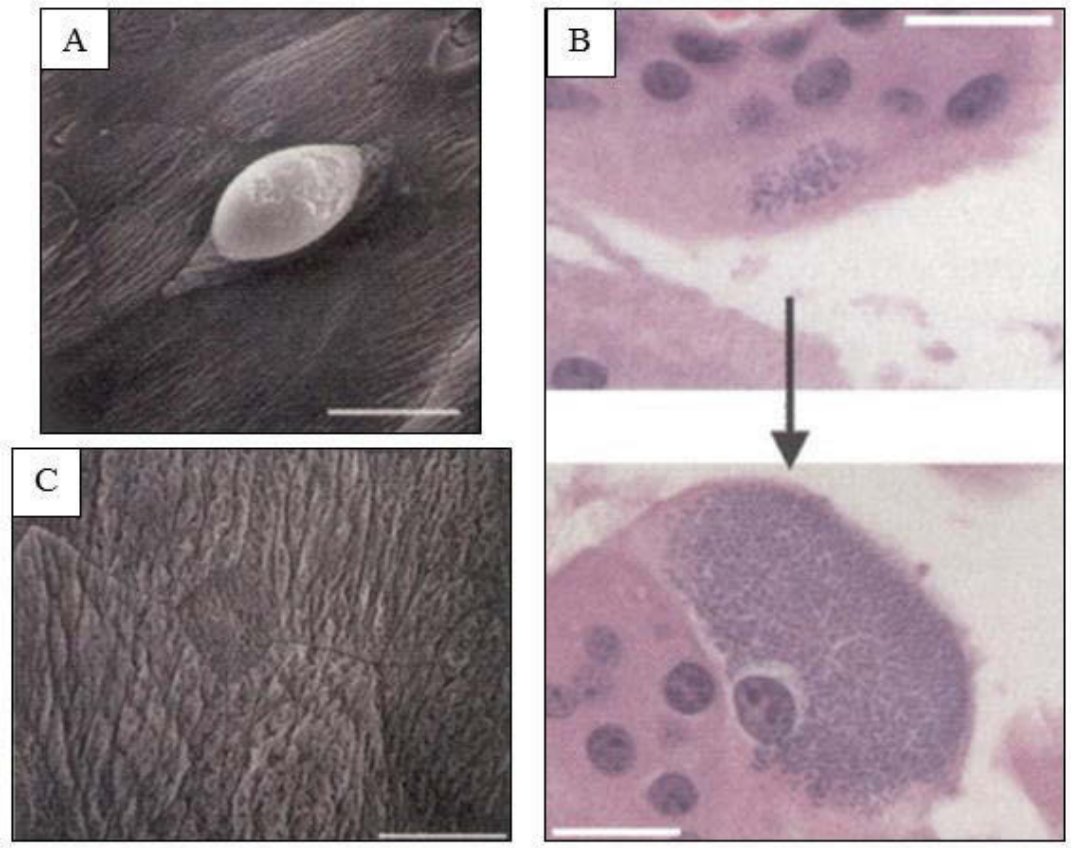


Figure 1.5 Intracellular bacterial communities of UPEC within the mouse bladder

(A) Electron microcopy of mouse bladders infected with cystitis isolate UTI89 (29) in vitro for 24 hr. A mature IBC is visible within the host cell protruding out into the bladder lumen as a pod-like structure. (B) Haematoxylin and Eosin stained mouse bladder infected with UTI89, demonstrating an early IBC after 6 h post-infection (top panel) and a late IBC at 24 h post-infection (bottom panel). (C) Electron microscopy of mouse bladders infected with non-pathogenic E. coli MG1655 reveal no biofilm-like or pod structures protruding out from the host cell. (Scale bars = 50µm (A), 20µm (B), 50µm (C)) Adapted from Anderson et al. (2003) Science (52).

1.3.3 UPEC DISPERSAL AND FILAMENTATION

After approximately 24 h following invasion, the IBC develops and overwhelms the host epithelial cell. The host cell ruptures spontaneously, and subpopulations of highly-motile and filamentous bacteria are released (*17*) (Fig. 1.6A-C). Filamentous bacteria are formed by continued growth and cell elongation with an arrest of normal cell division (*17, 26*). Through membrane staining of these filaments, Andersen *et al*. (*26*) saw minor invaginations along the filaments that could represent potential division-sites, at spacing that corresponded to approximately one normal cell length in the filaments (Fig. 1.6D). This suggests that cell division is blocked at an early stage of division, before full invagination of the cell envelope. Some of the proteins involved in the early stages of cell division, including FtsZ, FtsA, ZipA, ZapA and ZapB, may possibly be assembled and ready to divide the filaments again into rod-shape cells, which is necessary for long-term survival of *E. coli* (ongoing growth of filaments is eventually lethal). The cell division machinery will be summarised further below.

The development of filamentous bacteria accompanies host cell rupture and dissemination of UPEC to neighbouring host cells (*29*) (Fig. 1.6). UPEC filamentation is also known to assist in providing greater attachment to urinary tract surfaces and avoid consumption by the host immune response (*58*). *In vivo* studies undertaken on murine mouse models show that filamentous bacteria can subvert host phagocytic leukocytes due to their overwhelming size (*58*). This highlights the vital role filamentation can play in the progression of a UTI.

UPEC filaments have the potential to revert to bacillary rod-shaped morphology and can thereafter divide as normal bacterial cells, thereby reinitiating a new infection cycle (*17*). It has been suggested that filaments are capable of reverting to rod-shape with the purpose of regaining their invasive abilities in the bladder. Justice *et al.* (*17*) has observed reversal of filamentous UPEC in infected mouse bladders and Andersen *et al*. (*26*) observed reversal *in vitro* by harvesting filamentous bacteria from a human flow chamber-based cell culture model and recovering them in liquid LB media (Fig. 1.6E). This reversal, *in vitro,* is observed within 30 minutes of recovery and filaments greater than 100 μm still retained

their ability to revert, highlighting the sheer number of viable single cells produced by one filament (*26*).

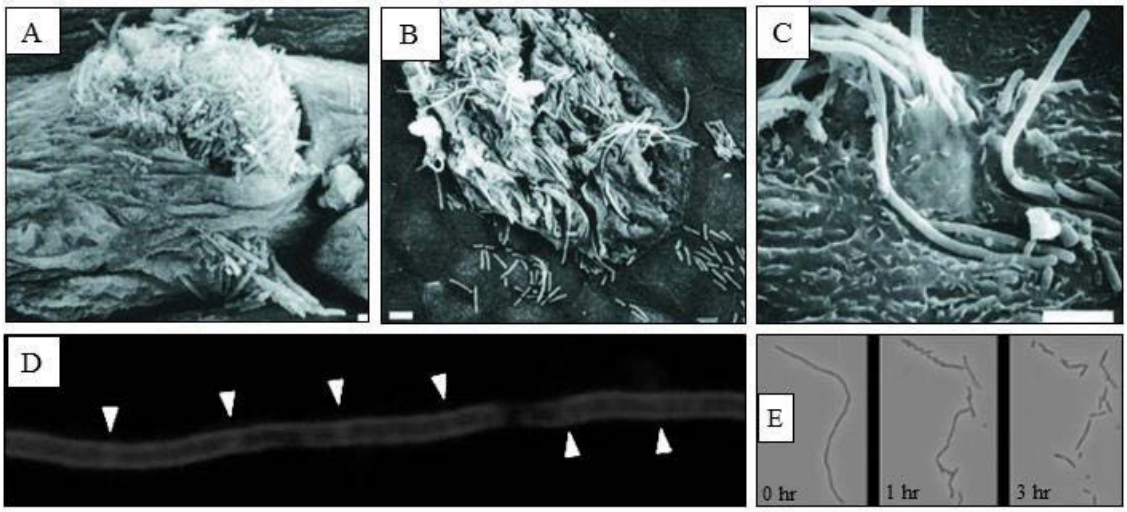


Figure 1.6 Bacterial filamentation, egress and dispersal from the bladder cell

(A-C) Scanning electron microscopy (SEM) images of infected mouse bladder epithelium, showing examples of IBCs erupting from within bladder epithelial cells leading dispersal of motile and filamentous E. coli UTI89. (D) A membrane stained filamentous cell (UTI89) obtained from UTI in vitro, demonstrating membrane cross-structures evenly spaced along the length of the filament (arrowheads), potentially consistent with the existence of partly-assembled divisomes that have been arrested to allow filament formation. (E) UTI89 filamentous cell harvested from an in vitro infection model undergoing reversal to rod-shape within the timeframe of 3 hr. (Scale bar = 5µm (A-C)). Adapted from Mulvey et al (2001) Infect. Immun (29) and Andersen et al. (2012) Infect. Immun (26).

1.3.3.1 E. coli cell division and filamentation

The filamentation process in UPEC involves ongoing bacterial cell growth and elongation, but a block to the normal cell division process. The exact mechanism of how this is regulated is yet to be understood. However, it must involve regulation of the cell division process, which will be summarised here, followed by a discussion of the known regulators of cell division during normal growth (see *Section 1.3.3.2*), and conditional filamentation during stress (*Section 1.3.3.3*).

There are over 30 essential proteins involved in the cell division process of *E. coli*, the most conserved and first to arrive is FtsZ, a homolog of the eukaryotic cytoskeletal tubulin (*59*). Polymerisation of FtsZ initially forms a scaffold at mid-cell, essential for the assembly of a complex protein structure, the divisome, which eventually constricts the cell envelope to divide the cell (*60*) (Fig. 1.7A). The basic organisation of FtsZ was first recognized as a ring structure around mid-cell (now referred to as the Z-ring) in *E. coli* using immunoelectron microscopy (*61*). Later, the use of GFP-labeled FtsZ revealed that the Z-ring is compiled gradually with FtsZ being loosely organised at mid-cell before condensing into the ring structure (*59*). Phospholipids (*62*) and recently, metabolic proteins (*63*) have also been reported to play a role in controlling Z-ring formation in *E. coli* and *B. subtilis*. The details of these regulators still remain largely unknown.

There are several other proteins involved in Z-ring control, stabilisation and membrane attachment that are considered early stage assembly proteins and most likely to be involved in arresting cell division. One such example, the essential cell division protein, FtsA, part of the actin family of ATPases, localises with FtsZ to the *E. coli* Z-ring (*64*). Overexpression of FtsA is known to cause cell division arrest and the formation of filamentous *E. coli*, which is the hallmark of a defect in or block to cell division (*65*). Together with the ZipA protein, FtsA supports Z-ring stability and both proteins are required for the recruitment of downstream division proteins, such as FtsE, FtsX and FtsK (*66*). ZipA is also an essential cell division protein (*67*) that localises to the Z-ring via

interactions between the *ftsZ-zipA* C-terminal domains, which lead to the organisation of FtsZ protofilaments (*68*).

Following ZipA, the non-essential cell division protein, ZapA, binds to FtsZ polymers and enables stability of the Z-ring by cross-linking the polymer molecules (*69, 70*). ZapA is required for recruiting ZapB to the Z-ring (*71*). Both the *zapA* and *zapB* deletion mutants contain irregular division septum and disordered arrangement of FtsZ (*72*). Overexpression of both proteins is known to cause cell division arrest and the formation of filamentous cells in *E. coli* (*73*).

The last stage of cell division involves cell wall peptidoglycan remodeling; preexisting peptidoglycan is degraded and the new synthesis of septal peptidoglycan results in cell wall invagination and the formation of two healthy daughter cells (*59*). It is expected that many more genes involved in the mechanism and regulation of cell division in *E. coli* remain to be discovered (*60, 74*).

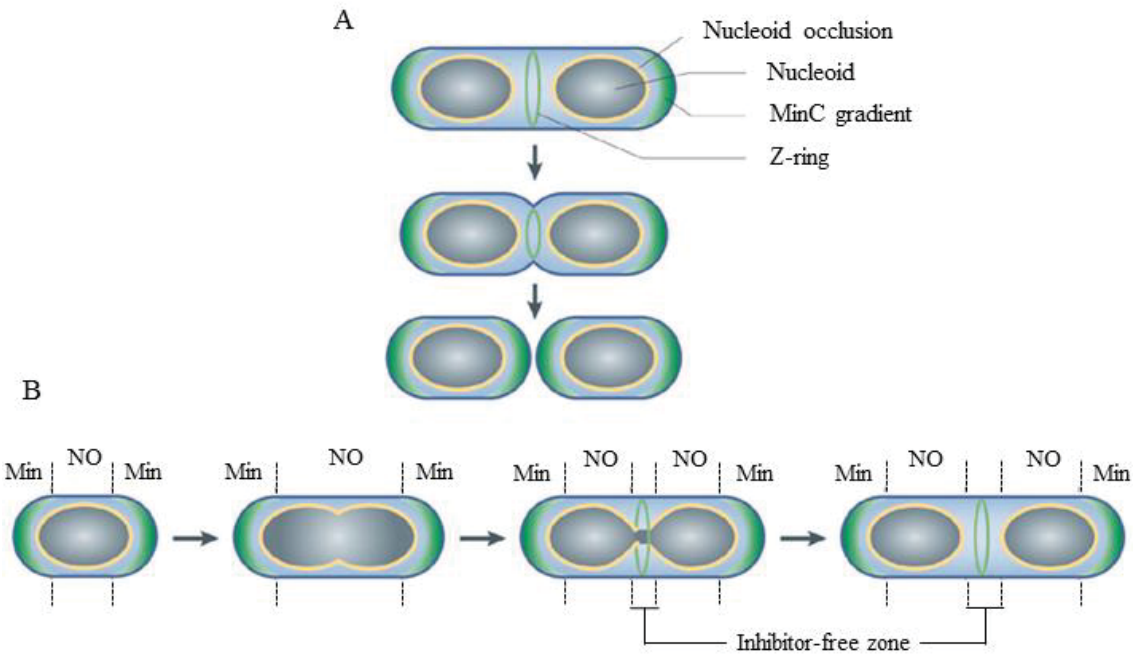


Figure 1.7 Simplified overview of E. coli cell division and the main known negative regulators of Z-ring formation

(A) Following chromosome replication, the Z-ring at mid-cell constricts to divide the cell. New synthesis of the cell wall continues towards the direction of the constricting Z-ring. The nucleoid occlusion (NO) and the Min system inhibit the Z-ring from forming at inappropriate regions within the cell. (B) The NO system functions through the activity of the nucleoid-binding protein, SlmA, which physically binds to FtsZ polymers to depolymerise the Z-ring. The Min system, involving the MinC, MinD and MinE proteins functions by concentrating the MinCD FtsZ-inhibitor towards the poles of the cells, which prevents division there. With the combined action of the NO and Min systems, the inhibitor-free zone is the most suitable placement for cell division to occur. These inhibitors of FtsZ assembly are positioned away from the cell center to effectively prevent Z-ring formation in their vicinity. Adapted from Adams & Errington (2009) Nat Rev Microbiol (60).

1.3.3.2 Negative Regulatory Systems of Cell Division in E. coli

To enable successful cell division, the appropriate mid-cell site must be selected for Z-ring formation (*59*). In *E. coli*, at least two negative regulatory processes govern the correct positioning of the Z-ring and prevent division at inappropriate sites by acting on FtsZ; the nucleoid occlusion system (NO) and the Min protein system.

The NO is responsible for inhibiting cell division from occurring over the nucleoids (*75*) (Fig. 1.7B). This process involves the nucleoid-binding protein, SlmA, which is an inhibitor of FtsZ polymerisation, preventing the Z-ring from forming when in close association to the chromosomes (*76*). The NO mechanism was first visualised when temperature-sensitive *E. coli* mutants undergoing nucleoid replication could influence the position of the divisome (*77*).

The Min protein system assists in determining the correct position of the Z-ring by inhibiting division at the poles of the cell and concentrating Z-ring formation at mid-cell (*78*). This involves interactions between the MinC, MinD and MinE proteins. Functionally, MinD binds to the cell membrane in an ATP dependent manner, recruiting MinC, which negatively regulates FtsZ polymerization (*79*). MinE accelerates the hydrolysis of ATP (bound to MinD) to ADP, causing the subsequent release of MinCD from the membrane (*79, 80*). The dynamics of the binding and unbinding cycle are sufficient to give rise to the oscillation of the MinCD complex from pole to pole, with a node of minimum MinCD concentration at mid-cell (*81*). This effectively concentrates MinCD towards the poles of the cell to avoid formation of the Z-ring at these regions. A depletion of the min system leads to cell division solely at the poles of the cell and results in the formation of mini-cells (*82*). It has been demonstrated that overproduction of MinCD prevents a node at mid-cell for FtsZ localisation, resulting in the arrest of cell division and formation of filamentous bacteria (*83*).

The NO and Min protein systems are relatively well characterised and fundamental to the placement and function of FtsZ. Their combined action ensures that FtsZ only assembles

at mid-cell (*59*) (Fig. 1.7B**)**. It is possible that these inhibitory systems are adapted, under certain conditions, to inhibit cell division at mid-cell and induce filamentation in *E. coli*.

1.3.3.3 The SOS Response and Involvement of SulA

An inducible filamentation mechanism already evident in *E. coli* is the bacterial "SOS" response and induction of SulA (*84*) (Fig. 1.8). The FtsZ polymerisation inhibitor SulA is induced as part of the "SOS" response to certain stresses, notably DNA damage. In this case, the block to cell division gives the cell time to repair DNA damage and segregate the chromosomes before resuming division (*84, 85*). In the classical SOS response, damaged DNA activates a sensor protein, RecA, which signals the degradation of the SOS repressor (LexA), resulting in the expression of a suite of genes needed for DNA repair, one of which is *sulA* (*83*). It has been suggested that SulA induces UPEC filamentation during UTIs (*58*).

Research into the formation of UPEC filamentation during UTIs and its potential association with the "SOS" response has been controversial. In 2006, Justice *et al.* (*58*) investigated the function of SulA on filamentation in the murine mouse model of UTI. A disrupted allele of *sulA* was transduced in the UPEC cystitis isolate UTI89 (*29*). Mice were then inoculated with the UTI89-Δ*sulA* mutant and compared to mice inoculated with wild-type (WT). Mice bladders and urinary tracts were examined at 6, 16, 24 and 48 h post-infection (*58*). The ability of UTI89-Δ*sulA* and WT to undergo filamentation was assessed by fluorescence microscopy. It was found that no filamentation was evident in the UTI89-Δ*sulA* mutant and thus, Justice *et al*. (*58*) concluded that SulA must be directly related to UPEC filamentation during UTIs.

In contrast, a more recent study undertaken in 2012 by Andersen *et al.* (*26*) found that filamentation in a human flow chamber-based cell culture model is independent of the bacterial SOS response and does not require SulA; they reported that filamentation was evident in the UTI89-Δ*sulA* mutant strain to a similar degree as WT.

It was noted in the report by Justice *et al.* (*58*) that the general bacterial load in mice infected with the UTI89-Δ*sulA* mutant strain was considerably lower than those mice

infected with WT. The generally low bacterial count of mice infected with UTI89-Δ*sulA* could hinder the detection of filaments at the fixed time-points and thus, may have resulted in filamentation being missed by Justice *et al*., (*58*) thereby misrepresenting filamentation as being completely lacking. Additionally, UTI89 filamentation is first noticeable 20 hours post-infection (*17, 26*), yet Justice *et al.* (*58*) reported observing no filamentation at 6 and 16 hours post-infection, concluding the absence of filamentation at these time points directly due to *sulA* being deleted in the mutant strain. In this case, the timing of the experimental sample could have accounted for the absence of filamentation in the UTI89-Δ*sulA* mutant.

Andersen *et al.* (*26*) concluded that UTI89 filamentation can be induced via alternative genetic pathways independent of *sulA*. Moreover, studies undertaken on filamentous *Salmonella typhimurium* have identified genes independent of *sulA* and the SOS response aiding in the filamentation of these bacterial cells within host macrophages (*86*). Global genomic expression analysis of UPEC filaments harvested during the filamentation stage in the human flow chamber-based cell culture model demonstrates that *sulA* and the SOS response are not activated (*56*). However, one of the highly upregulated genes from this study, *damX* (6-fold change of expression), had previously been implicated in cell division and its regulation (*56*), prompting further investigation into its function during UTI.

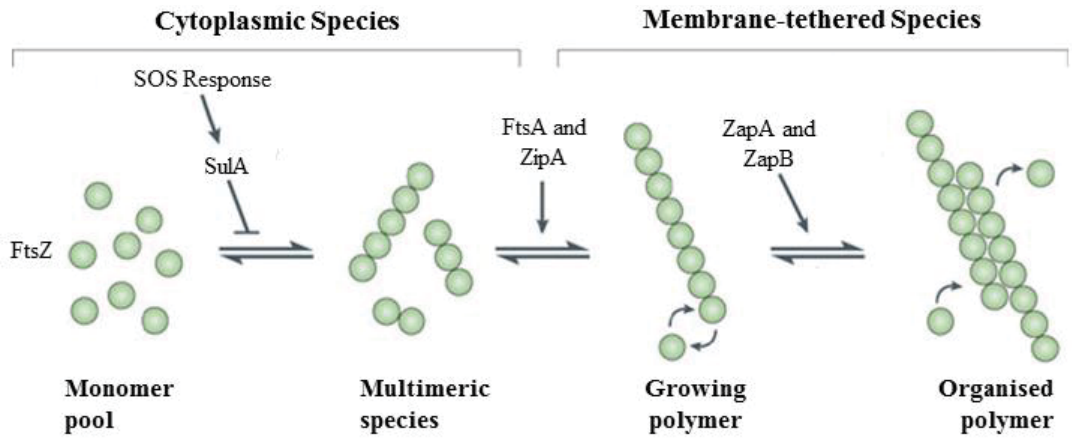


Figure 1.8 FtsZ assembly during cell division and the involvement of SulA

In the presence of DNA damage, the sensor protein RecA is activated which signals the degradation of the SOS repressor, LexA, resulting in the expression of SulA and various other genes needed for DNA repair. When induced, SulA is able to inhibit cell division by binding to FtsZ and blocking polymerisation. This block to cell division gives the cell time to mend the damaged DNA and segregate the chromosomes before restarting the division process. Adapted from Adams & Errington (2009) Nat Rev Microbiol (60).

1.3.3.4 DamX and reversible filamentation during UTIs

The SPOR-domain containing protein, *damX*, is recruited to the septum during cell division and is involved in binding septal peptidoglycan (*87*). Within UTI89, *damX* is positioned in a functionally diverse operon containing *aroB*, *aroK*, *dam*, *rpe*, *gph* and *trpS*. It has been demonstrated that overexpression of *damX* from an inducible promotor in *E. coli* results in cell division arrest and the production of filamentous cells (*88*). Recently, Khandige *et al.* (*56*) demonstrated a lack of filamentation in the *damX* deletion mutant of UTI89 (UTI89Δ*damX*) during the dispersal phase of the human flow chamber-based cell culture model and in the murine mouse model of UTI, implicating *damX* as a requirement for UPEC filamentation during infection. The *damX* mutant demonstrated a comparable growth rate to WT when grown in LB and concentrated urine (urine $\rho > 1.025$ kg/m^3) (*56*). Additionally, the mutant demonstrated equal adherence and invasion capabilities, as well as similar efficiency to proliferate and form IBCs *in vitro* (4-5 generations of doubling times) when compared to WT.

The microarray expression analysis from this study of UPEC harvested during the dispersal phase of infection also revealed the upregulation of *ftsZ*, *ftsA, zipA* and *zapB* – all of which were downregulated in the preceding IBC phase of infection (*56*) (this is discussed further in *Section 1.6.3*). Thus, *damX* has demonstrated to play an important role in UPEC filamentation, however the complete suite of factors that are involved in the filamentation response pathway (i.e. sensing external stimuli and transducing the signal to effector molecules internally to arrest division) still remain largely unknown and could involve an array of other known and novel cell division regulators (this hypothesis is focused upon in *Chapter 3* and *Chapter 5*).

1.4 UPEC Metabolic Responses During Pathogenesis

UPEC must rapidly adapt to new environments as they transition between the intestinal lumen and urinary tract (*19*). Considering the different nutritional standings of the intestine and urinary tract and the variability within each, metabolites within the urine and bladder are most likely to activate metabolic adaptive strategies in UPEC in order to survive and cause infection inside the urinary tract. Since many of these metabolic strategies are not pathogen-specific, their roles in pathogenesis and fitness during infection remains underappreciated.

1.4.1 THE PRIMARY MEDIUM: NUTRITIONAL ASPECTS OF URINE

In the urinary tract and bladder, urine is one of the major sources of nutrients encountered by UPEC during extracellular infection. Understanding the complete metabolic profile of urine that is utilised by UPEC is challenging mainly due to the compositional complexity and variability of urine. Nonetheless, advances in techniques such as metabolomics allow large-scale simultaneous analysis of compounds in complex biological mixtures and have greatly expanded our knowledge of the metabolic profile of urine (*89*).

Urine contains relatively high concentrations of urea, creatinine, amino acids, organic acids, and inorganic ions such as ammonia, sodium and potassium, all of which must be either utilised or tolerated by UPEC in the urinary tract (*90*). Urine contains 5-10 times more compounds and exhibits 2-3 times more chemical diversity compared to other biological fluids like saliva (*91*) and cerebral spinal fluid (*92*). The major metabolites of urine are amino acids and carbohydrates (*93*). D-serine is the most abundant amino acid in mammalian urine, whilst organic metabolites such as hydroxyl acid, citric acid, ammonia, creatinine and hippuric acid are also present in significant quantities (*94*). Despite containing high concentrations of many metabolites that may be utilised by UPEC, urine is classified as a fairly nutrient-limited growth medium, because arginine, methionine, uracil, adenine, isoleucine, glucose, iron and hormones such as oxytocin and melatonin are

in limited supply (*90, 93, 95*). UPEC would largely rely on its capacity for *de novo* biosynthesis of these and many other essential metabolites.

1.4.2 CENTRAL CARBON METABOLISM

During intracellular UPEC colonisation of the bladder, aerobic respiration and carbon metabolism are essential for survival (*20*). Pyelonephritis isolate CFT073 (*39*) defective in *sdhB*, the succinate dehydrogenase enzyme involved in the conversion of fumarate to succinate in the tricarboxylic acid cycle, is significantly outcompeted by the isogenic parental strain in mouse bladder UTI models (*20*) (Fig. 1.9). Similarly, disruption of gluconeogenesis by deletion of *pckA*, encoding the phosphoenol-pyruvate carboxykinase which catabolises oxaloacetate to phosphoenol-pyruvate, results in a significant reduction in fitness to colonise the bladder *in vivo* when compared to WT (*20*). However, mutants defective in gluconate catabolism of the Entner-Doudoroff pathway, by deletion of *edd*, do not affect *in vivo* fitness in a mouse model of UTI, suggesting this pathway is dispensable in UPEC pathogenesis (*20*), and indicates a specific metabolic adaptation by UPEC during bladder infection that utilises gluconeogenesis and the tricarboxylic acid cycle (Fig. 1.9). This highlights the importance of aerobic respiration, enabling growth at a faster rate. It should be noted that the infection model used in this study only demonstrates the fitness defect in the mutant of these enzymes when compared to WT and does not rule out the importance of other metabolic adaptations during UPEC infection. Interestingly, in a kidney mouse model of UTI, deletion of *tpiA*, the triosephosphate isomerase involved in glycolysis, was found to have attenuation and a colonisation defect in the mouse kidneys, but not in the bladder (*20*). Thus, it appears that the glycolysis pathway is dispensable during the initial establishment of infection within the bladder, but is critical for colonisation of the kidneys. This evidence suggests that gluconeogenesis and the tricarboxylic acid cycle must come to completion for UPEC pathogenesis during intracellular infection of the bladder, whilst glycolysis is important in kidney colonisation.

β-galactosides and sorbitol appear to provide a major source of carbon during intracellular infection in the urinary tract as both the sorbitol transporter, encoded by *srlA*, and the

galactoside catabolism gene, *lacZ* (encoding β-galactosides that cleaves the glycosidic bond in galactosides to release galactose), were found to be significantly upregulated during bladder cell colonisation (*96*) (Fig. 1.9). This is further supported by *galK* mutants (the gene product involved in conversion of galactose to glucose) producing significantly smaller IBCs compared to the WT in a mouse UTI model (*96*). Positive X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining of UPEC during the IBC phase of a mouse UTI model also implies strong β-galactosidase activity (*97*). Together, the evidence suggests glucose availability within intracellular infection is limiting and UPEC are adapted to utilise galactose and sorbitol as their prime carbon sources. This is in accordance with urine metabolome studies, which show that glucose is rarely abundant in the urinary tract unless in a diabetic situation (*93*).

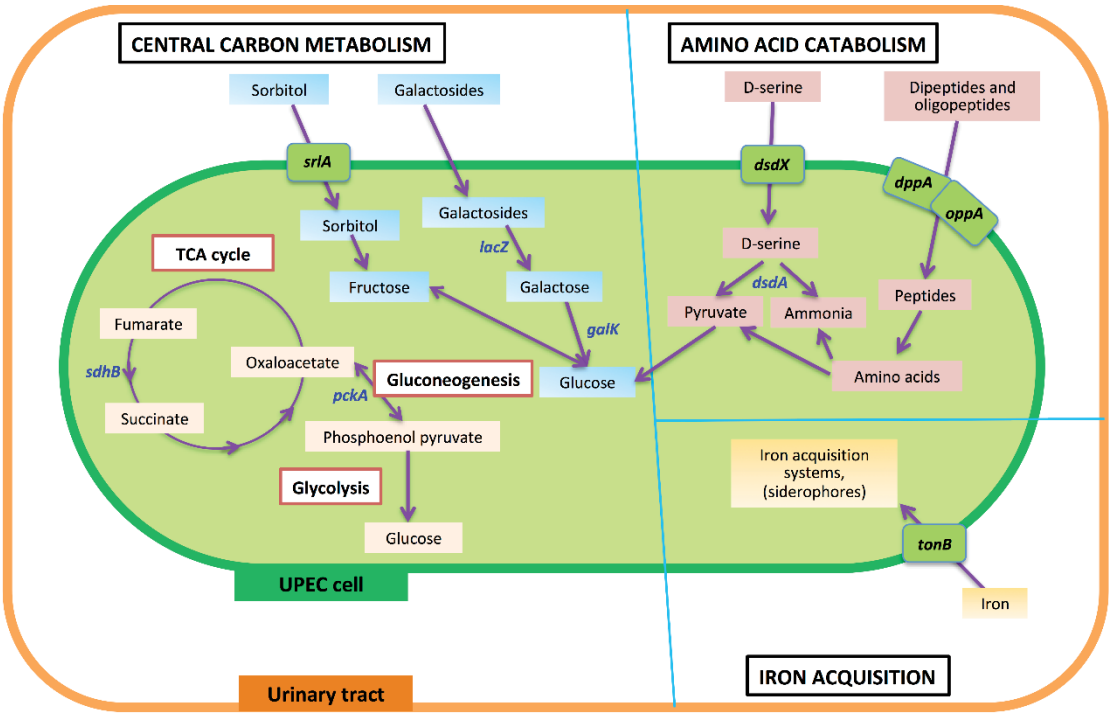


Figure 1.9 Some important metabolic pathways of UPEC during urinary tract infection

The main metabolic responses of UPEC during urinary tract infection, as detailed in text under the Section 1.4 "UPEC metabolic responses during pathogenesis". In comparison to the intestine, UPEC in the urinary tract displays a number of metabolic adaptations in central carbon metabolism, amino catabolism and other pathways, to cause infection. Genes identified to play a role in UPEC pathogenesis are shown. Genes in blue denote those that play a role in UPEC fitness in the urinary tract, or result in attenuation in a mouse or in vitro model. Transporters are shown in green boxes over the cell membrane. Adapted from Mann et al. (2017) Front Cell Infect Microbiol (28).

1.4.3 AMINO ACID CATABOLISM

The ability of UPEC to exploit the diverse amino acids and complex small peptides present within human urine has long been considered an adaptation in the urinary tract (*98*). The utilisation of these resources by UPEC is supported by the upregulation of genes involved in amino acid catabolism during growth in human urine, such as *dsdA* (D-serine deaminase) (*99*). Microarray studies have demonstrated the upregulation of arginine, serine and histidine transporters in UPEC during urine growth when compared to Luria-Bertani (LB) medium (*96*). Both serine and arginine auxotrophs do not exhibit any major fitness defects in a mouse UTI model (*20*), indicating that these amino acids are effectively utilised from the growth milieu. Small peptide uptake is also important for UPEC infection, as *dppA* (periplasmic dipeptide transport protein) and *oppA* (periplasmic oligopeptide-binding protein) deletion mutants defective in peptide transport in CFT073, were significantly out-competed by WT in the mouse bladder model of UTI (*20*) (Fig. 1.9). Interestingly, a subset of genes involved in amino acid utilisation and synthesis, most notably tryptophan, were downregulated during the IBC phase of infection (*96*), and strains that are auxotrophic for the *de novo* synthesis of this amino acids have the ability to grow to a high density in urine (*98*). This suggests that selective amino acid uptake and catabolism are utilised for growth in human urine and correlates with the high concentrations of amino acids reported in the metabolomics analysis of human urine.

1.4.4 IRON UPTAKE AND TRANSPORT SYSTEMS

UPEC express a variety of iron uptake and transport systems such as iron-chelating siderophores and hemophores during intracellular infection (*100*). Iron siderophores are low molecular weight metal chelators involved in ferric iron scavenging and play a major role in binding available iron to promote UPEC intracellular colonisation within the iron-limiting urine and bladder epithelial cells (*5*). Transcriptome studies have identified various genes involved in siderophore synthesis to be dramatically upregulated during mouse UTI infection, most notably salmochelin and yersiniabactin (*96*). Interestingly, the ferric iron uptake mediator TonB, which interacts with bacterial outer membrane receptors to

facilitate iron uptake, was shown to be important in the mouse kidney model of UTI, with a deletion of *tonB* resulting in a reduction in UPEC colonisation at 48 h post-infection when compared to WT (*101*) (Fig. 1.9). Additionally, three iron-scavenging genes were identified by Chen *et al*. (*102*) to be under strong positive selection within UPEC strains; *fhuA* (ferrichrome-iron receptor), *entD* (enterochelin synthase D) and *entF* (enterochelin synthase F) (*102*). This highlights the importance of iron acquisition systems in UPEC pathogenesis.

UPEC not only requires upregulation of genes needed for iron-acquisition under iron-limited conditions of the urinary tract, but also needs to precisely maintain a delicate balance of internal iron levels for survival (*103*). The sRNA RyhB and the global ferric uptake regulator Fur have been implicated for their role in iron acquisition and controlling the storage of iron in *E. coli* (*103*). Fur and RyhB regulate the activity of carbon metabolic processes as they directly target the tricarboxylic acid cycle (*104*). RyhB is important for colonisation of the urinary tract by CFT073 via the production of iron-acquisition systems, with a *ryhB* deletion mutant producing lower levels of the siderophores aerobactin, salmochelin and enterobactin (*105*).

The withholding of free iron by the host is usually mediated by iron-binding proteins like lactoferrin, ferritin, transferrin, ovalbumin and siderocalin (also known as lipocalin-2) (*106*). In an effort to counteract bacterial iron-scavenging proteins, siderocalin, introduced into the urinary tract by leukocytes and uroepithelial cells, can restrict the accessibility of iron to UPEC by scavenging ferric iron complexes (*107*). Siderocalin shows very high affinity to the bacterial siderophore enterobactin, which is found ubiquitously in *E. coli* strains. However, the lower binding affinity of siderocalin to other siderophores such as aerobactin, salmochelin, and yersinibactin, which are more often associated with UPEC strains (*100, 108*), may provide an evolutionary advantage for UPEC during host iron scavenging. Indeed, purified siderocalin is capable of effectively inhibiting *E. coli* K-12 that produces only enterobactin. However, no effect was observed on the growth of *E. coli* that produces all siderophores, such as UPEC (*109*).

1.4.5 OTHER METABOLIC REGULATORS OF UPEC

As well as the metabolic adaptation to survive within the above environments, nutritional sensing is vital for UPEC to efficiently respond to nutrient-fluctuating conditions. At least two regulatory systems are reported to be important for UPEC virulence, as described below.

qseBC: The two-component regulatory pathway system QseBC, consists of QseC which functions as a sensory kinase to the environment stimuli and host signal (specifically to epinephrine and norepinephrine), resulting in phosphorylation of the transcription factor QseB, acting as the response regulator in UPEC. QseC creates a positive feedback loop constantly expressing QseB (*110*). Furthermore, deletion of *qseC* in UTI89 leads to attenuated virulence and altered expression of 36 genes that modulate at least 18 metabolic pathways (*110*). These include an increase in pyrimidine production and utilisation, whilst purine biosynthetic genes, *purK*, *purM* and *purT*, as well as arginine synthesis genes, *argF*, *argG* and *argR* show downregulation (*110*). Bypassing purine production by the addition of guanosine or adenine monophosphate was found to restore growth in a *qseC* mutant, highlighting the importance of purine metabolism for UPEC survival. This is interesting as *de novo* purine biosynthesis has been recently shown to be essential for proliferation of UPEC within the bladder during the IBC phase, with Δ*purF* and Δ*cupA-purF* operon displaying impaired fitness at 3, 6 and 16 h post-infection *in vivo* (*111*).

Pst: The loss of the phosphate-specific transport system (Pst), encoded by the operon *pstSCAB-phoU*, results in the constitutive activation of the Pho regulon, involved in phosphate transport and metabolism. This constitutive activation has a negative effect on UPEC pathogenicity, with a Δ*pst* CFT073 mutant displaying a reduced fitness in a murine UTI model (*112*). This effect on virulence is thought to be due to the decreased production of Type 1 fimbriae in a Δ*pst* mutant, resulting in decreased invasion of bladder epithelial cells, and demonstrates a link between nutrient sensing and UPEC pathogenicity (*112*).

1.5 Host Immune Responses to UPEC Infection

There are multiple defense mechanisms that are implemented by the host to combat UPEC infection of the bladder. The flow of urine works as a preventative measure by washing away any unbound or weakly bound bacteria within the lumen (*5*). Urine also has certain properties that effectively decrease the survival of UPEC including the presence of urea, high salts and a mild acidity pH (~5.5) (*93*). Urine also contains various macro-molecules and anti-microbial peptides that can prevent bacterial attachment to the host cells. For example, the Tamm-Horsfall glycoprotein binds to Type 1 pili to block the surface binding capacity of UPEC (*113*). There are also multiple pathways that activate the host innate immune system: one of the most studied is the TLR-4 pathway. TLR-4 on the surface of host cells responds to the lipopolysaccharides on Gram-negative bacteria, which in turn activates the NF-κB pathway (*114*). NF-κB induces the expression of pro-inflammatory molecules, such as IL-6 (*114*). Alternatively, TLR-4 can also induce expression of pro-inflammatory molecules in a pathway independent of NF-κB by implementing secondary messengers such as, calcium and cAMP response element binding (CREB) protein (*115*).

1.5.1 UPEC EVASION OF HOST DEFENCES

The innate immune system uses pattern recognition receptors (PRR) to recognise pathogen-associated molecular patterns (PAMPs) that are unique to invasive pathogens (*116*). It is evident that UPEC avoid recognition by PRR early in infection by inhibiting components of immune pathways, such as TLR-4 and the subsequent production of pro-inflammatory molecules (*42*). This may grant them sufficient time to invade the bladder epithelial cells before neutrophils and macrophages are recruited to the infected area. This is a clear demonstration that UPEC are capable of slowing down the immune response. As mentioned previously, UPEC filamentation is also known to protect UPEC from engulfment by neutrophils (*58*), which may assist UPEC in maintaining ongoing infection.

1.6 'OMIC SCREENING APPROACHES TO FURTHER UNDERSTAND UTIS

The application of comparative bioinformatics and various 'omic approaches, including genomics and transcriptomics towards the study of UPEC has recently provided a large amount of data in regards to the virulence factors and biological mechanisms that are involved during infection. This thesis expands on these studies by use of several functional genomic approaches to identify and understand the UPEC nutritional and growth requirements in various environments. The relevant previous approaches are detailed below.

1.6.1 COMPARATIVE BIOINFORMATICS

In order to identify virulence genes within UPEC strains, numerous research groups have focused on identifying those genes conserved uniquely within UPEC strains. Type-1 pili (*34*), P-fimbriae (*40*), α-hemolysin (*117*), *cnf1* (*117*) and polysaccharide surface structures (O antigen and capsule) (*118, 119*) were among the virulence factors identified. Additionally, Chen *et al*. (*102*) identified genes under strong positive adaptive selection within UPEC strains, which resulted in the identification of genes involved in cell wall and membrane biogenesis, cell division regulation and nutrient acquisition. To determine positive adaptive selection, this study utilised six pathogenic strains of *E. coli*: UTI89, CFT073, EDL933, SAKAI, 2457T, and 301 (*102*). Each genomic sequence was aligned using the multiple alignment program ClustalW and genes under positive selection were identified using the *PAML* software (*102*). The findings showed that the *ompF* and *ompC* outer membrane porins were under strong positive selection in UPEC. Interestingly, both genes are controlled by the EnvR/OmpR and CpxA/CpxR two-component systems that regulate nutrient uptake. *ompC* was also found to be under strong positive selection in *Shigella flexneri* which, similar to UPEC, is able to invade and replicate within host epithelial cells (*102*). Additionally, the genes *cedA* (cell division activator) and *amiA* (N-acetylmuramyl-L-alanine amidase), which have roles in cell division regulation and septal cleavage, respectively, were also found under positive selection (*102*). Overexpression of CedA in *E. coli* activates cell division and prevents a filamentous phenotype (*120*).

Recently, the deletion of *cedA* in *E. coli* has been shown to exhibit increased susceptibility to the bactericidal nanoparticle, LAL-32, which is known to inhibit cell division (*121*). Changes in UPEC morphology during infection (filamentation and reversal) must require regulatory mechanisms in cell division. In summary, the genes under positive selection involved within cell wall and membrane biogenesis, septal cleavage and cell division regulation may play a key role in UPEC pathogenesis.

As mentioned previously, iron acquisition is pivotal to UPEC pathogenesis within the host bladder. Three iron scavenging genes were identified by Chen *et al*. (*102*) to be under strong positive selection within UPEC; *fhuA* (ferrichrome-iron receptor), *entD* (enterochelin synthase D) and *entF* (enterochelin synthase F) (*102*).

Further, a bioinformatics analysis of CFT073, EDL933 and MG1655 identified unique pathogenic clusters in CFT073, suggesting that various alterations in the UPEC genome have conferred UPEC the ability to be highly pathogenic in nature (*39*). These unique clusters contain mostly putative autotransporter proteins, iron-sequestration systems and adhesion fimbriae systems (*39*). Autotransporter proteins use the Sec system for translocation of a protein across the cytoplasmic (inner) membrane, which the autotransporter beta-barrel domain can transport the passenger domain across the outer membrane (*122*). One such secreted protein, Sat (a serine protease) is suspected to be a novel virulence factor for pyelonephritis isolates because Sat antibodies were strongly detected in the serum of mice following ascending UTI (*122*). Lateral gene transfer between pathogens was also demonstrated, suggesting evolutionary acquisition of pathogenicity within certain strains (*39*). However, further knowledge of these pathogenic mechanisms is needed and by elucidating the pathways that these genetic molecules work, in combination with comparative bioinformatics, the opportunity exists to improve and further develop our understanding of UPEC.

1.6.2 TRANSCRIPTOMICS

Microarray technologies initially produced large datasets of genome-wide expression profile in UPEC to understand the gene-expression changes that take place in the infection (*99*). Recently, transcriptomics was used to determine the genome-wide expression profile of UPEC transitioning through each individual phase of the bladder infection model. Khandige *et al*. (*56*) utilised microarray technology in combination with the human flow chamber-base cell culture model of UTI to determine the expression profile of UTI89 at the IBC, dispersal (filamentation) and reversal phases of infection (*56*). During the IBC phase, a large number of differentially expressed genes were involved in metabolic processes, stress response and iron homeostasis. For example, expression profiles of the *araA* (conversion of L-arabinose to L-ribulose), *idnD* (L-idonate degradation), *garL* (synthesises D-glycerate from galactarate) and *chuA* (outer membrane heme receptor) genes were all upregulated greater than 3.5-fold when compared to UTI89 grown in LB medium. The identification of a large number of genes involved in *de novo* metabolic synthesis pathways supports the evidence that UPEC utilise non-glucose carbon sources within the IBC and instead largely depend on biosynthesis of amino acids and gluconeogenesis for energy production (*96*).

During the dispersal and reversal phases, a large number of differentially expressed genes were involved in DNA replication, cell wall and membrane remodeling and cell division. Expression analysis revealed the upregulation of the essential cell division genes *zapB*, *ftsZ*, *ftsA* and *zipA* by more than 3-fold during the filamentation phase, when compared to the previous IBC phase and upcoming reversal phase (*56*). Additionally, the nucleoid occlusion protein, SlmA, involved in spatial regulation of cell division was upregulated 3.4-fold during the dispersal phase, suggesting a potential role for nucleoid occlusion (prevention of division over the chromosomes or nucleoids) in the arrest of UPEC cell division during infection. The most statistically differentially expressed gene was *damX*. Further experimentation showed that this gene, but not the other upregulated genes tested, was essential for UPEC filamentation during bladder infection (*56*) (see Section *1.3.3.5*). Expression analysis also revealed that genes involved in DNA replication and those

supporting the integrity of the outer membrane such as, *tolB*, *pal*, *ompA*, *lpp* and *nlpD* were all upregulated more than 5-fold when compared to the previous IBC phase and upcoming reversal phase, highlighting the potential regulation UPEC undergo to change shape (*56*).

1.6.3 FUNCTIONAL GENOMICS

The major advancements to genomics that have come about through the advent of high-throughput (or "next generation") DNA sequencing have enabled genetic analyses to be carried out in a high-throughput multiplexed manner. In this thesis, we utilise one of these very powerful DNA sequencing techniques, known as transposon-directed insertion-site sequencing (TraDIS), which maps the transposon insertion abundance within a pooled transposon-insertion mutant library. This reflects the abundance of the individual clones in the library. Bacterial transposon mutant libraries can then be subjected to various conditions and the abundance of surviving mutants can be compared to the pooled mutant library reference sample (*123*). This can provide a greater understanding of the genetic requirements and molecular pathways that underpin bacterial fitness and virulence through specific conditions.

TraDIS was first applied to UPEC to identify the genetic requirements for *in vitro* survival and growth of EC958 (an *E. coli* ST131 clonal lineage) in human serum (*124*). This study identified 56 genes that were required for survival in human serum, the majority of which were associated with lipopolysaccharide biosynthesis (the lipopolysaccharide outer core structure) and enterobacterial common antigen biosynthesis (carbohydrate-derived surface antigen) (*124*). Intriguingly, two genes associated with gluconeogenesis and carbohydrate biosynthesis, *pgi* (glucose-6-phosphate isomerase) and *fbp* (fructose-1-6-bisphosphatase), were identified as required for survival in human serum (*124*). These genes catalyse the production of D-fructose-6-phosphate, an essential step in the glycolysis and gluconeogenesis pathways. As well as their roles in carbohydrate metabolism, D-fructose-6-phosphate also is a precursor for the biosynthesis of uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) that is required for cell wall peptidoglycan, lipopolysaccharide and enterobacterial common antigen biosynthesis.

The molecular mechanisms that grant UPEC the ability to survive in serum may also represent vital traits that UPEC utilise globally to survive in diverse environments. Integrating TraDIS not only during UPEC infection-relevant conditions (i.e. growth in urine or serum), but during the UTI infection events *in vitro* and *in vivo* will provide greater insights into the virulence factors and metabolic response pathways associated with UPEC colonisation, IBC formation, filamentation and reversal.

1.7 Aims and Objectives

UTIs are one of the most prevalent infectious diseases, resulting in billions of dollars of expenditure every year, and very substantial morbidity and mortality globally (*5*). Of further concern is the rise of antibiotic resistance seen amongst pathogens causing UTIs (*5*). More than 90% of UTIs are caused by UPEC, which have the ability to invade and grow within the cells of the human urinary tract, undergoing a multi-stage infection cycle. As mentioned above, UPEC possess an arsenal of virulence factors, which play specific roles in enhancing host colonisation and overall result in damage to the host. In addition, broader metabolic flexibility and physiological adaptability has been shown to be critical for UPEC adaptation to specific host environments, such as the urinary tract.

The studies that have contributed to the knowledge of UPEC survival and infection of the bladder, in particular the later phases of infection (filamentation, dispersal and reversal) have provided only minimal mechanistic insight. Further identifying the genetic requirements and metabolic responses undertaken by UPEC, using genome-wide approaches, will help uncover the strategies and mechanisms underpinning survival within the bladder. This will facilitate development of new treatments and therapies for UTIs in future.

The work presented in this thesis aimed to explore the genetic and metabolic requirements of UPEC, which allow it to survive and cause infection in the human urinary tract. Some of the following research makes use of a human bladder cell culture line and an *in vitro* infection model. The UPEC cystitis isolate UTI89 (*29*) was used as the model organism for these studies.

The following chapters detail the methods and results obtained during the work carried out for the term of this PhD:

- Chapter 3: The filamentation response of UPEC is likely to involve a number of gene products involved in sensing the (unknown) external factors that induce

filamentation, transducing the signal to intracellular effector molecules, and ultimately inhibiting the cell division apparatus. None of these are currently confirmed. This chapter uses genome-wide high-throughput sequencing, in combination with flow cytometry sorting, to comprehensively screen and identify those genes in UTI89 that can cause filamentation under controlled laboratory culture conditions.

- Chapter 4: UPEC must adapt their metabolism according to the nutrients available inside the host in order to survive and cause infection, such as those of the bladder, kidneys and bloodstream. This metabolic capacity and flexibility of UPEC has been recognised to be important for pathogenesis in the urinary tract. This chapter uses the genome-wide screening technique termed TraDIS, which maps transposon insertions in a pooled collection of random mutants, to define the genes required for growth and survival of UPEC in minimal-media (sole glycerol carbon source) when compared to complex-media.

- Chapter 5: Knowledge into the UPEC infection cycle, particular the later stages of infection is lacking and studies have provided only minimal insight. As such, a genome-wide genetic screen, such as TraDIS is capable of providing comprehensive insights into the genetic requirements associated with UPEC pathogenesis. This chapter uses TraDIS to define the set of genes required for the IBC, dispersal and bacterial reversal/recovery phases of the UPEC infection cycle by utilising the human bladder cell culture line PD07i in an up-scaled *in vitro* static infection model incorporating concentrated human urine.

CHAPTER 2

GENERAL MATERIALS AND METHODS

2.1 Bacterial Strains and Plasmids

Table 2.1 and Table 2.2 list the general *E. coli* strains and plasmids used throughout this thesis, which were already available. Bacterial strains and plasmids produced in this work will be described where appropriate. All *E. coli* strains were stored at -80°C as stationary phase cultures with 16% (v/v) glycerol. Mammalian bladder cell line PDO7i (Moller-Jensen, J (*26*) and Mulvey, M (*43*)) was used in this thesis.

Table 2.1 General E. coli strains used in this thesis

***E. coli* Strain**	**Genotype**	**Source**
UTI89	Wild-type	Moller-Jensen, J.
K-12 BW25113	Wild-type	Baba *et al*. (*125*)
K-12 DH5α	*fhu*A2 Δ(*argF-lacZ*)U169 *phoA glnV44* Δ(*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsd*R17 F-*ompT hsd*SB(rB-, mB-) *gal dcm araB*::*T7RNAP-tetA*.	New England Biolabs (NEB).

Table 2.2 General plasmids used in this thesis

Plasmid	**Purpose (chapter used)**	**Selection**	**Source**
pBAD24	General cloning, overexpression of cloned genes (*Chapter 3*).	Amp	Moller-Jensen, J.
pKD4	γ-red recombination system for genetic knockout (*Chapters 4 and 5*).	Kan	Dasenko *et al*. (*126*).

2.2 Bacterial and Mammalian Growth Media

Bacterial growth media used in this thesis is listed in Table 2.3. The manufacturers of chemicals, reagents and solutions are listed in Table 2.4.

Table 2.3 Bacterial and mammalian growth media

Growth media	**Composition[a,b]**
Luria-Bertani (LB)	1% tryptone, 0.5% yeast extract, 1% NaCl.
Minimal-media (M9)	Composition as in Burke *et al.* (*88*). Where specified, the carbon source used was either 0.4% glucose and 0.1% casamino acids, or 1.0% glycerol.
SOC	2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 20mM glucose.
EpiLife	Composition as stated in EpiLife medium specifications sheet, Life Technologies, USA.

[a] *All percentages are given as (w/v) unless otherwise indicated.*
[b] *Solid LB contains 1.5% (w/v) agar.*

2.3 CHEMICALS, REAGENTS AND SOLUTIONS

General chemicals, reagents and solutions used in this thesis are described in Table 2.4, along with the companies they were obtained from. Aqueous buffers and solutions used in this thesis are described in Table 2.5.

Table 2.4 General chemicals, reagents and solutions used in this thesis

Chemical	**Source**
Agarose	Bioline, Australia
Ampicillin	Genlantis, USA
Arabinose	Sigma-Aldrich, USA
DMSO	New England Biolabs, UK
Defined Trypsin Inhibitor (DTI)	Gibco, Life Technologies, USA
FM4-64	Thermo-Fischer Scientific, USA
Gel-Red	Biotium
Gentamicin	Gibco, Life Technologies, USA
Glycerol	Sigma-Aldrich, USA
Human Keratin Growth Supplement (HKGS)	Gibco, Life Technologies, USA
Kanamycin	Sigma-Aldrich, USA
Triton X-100	Sigma-Aldrich, USA
Trypsin-EDTA	Gibco, Life Technologies, USA
Trypsin Inhibitor (DTI)	Gibco, Life Technologies, USA

Table 2.5 Aqueous buffers and solutions used in this thesis

Solution	**Composition**
PBS	150mM potassium phosphate, 50mM Tris-Cl, pH 7.4.
TAE	40mM Tris, 20mM acetate, 1mM EDTA, pH 8.6.
TBE	90mM Tris, 90mM boric acid, 2.5mM EDTA, pH 8.3
Lysis solution	0.5% trypsin-EDTA, 0.1% triton X-100.
Qubit working buffer	99.5% HS buffer, 0.5% HS dye.

2.4 General Growth Conditions of *E. coli*

To culture *E. coli*, scrapings from a frozen glycerol stock strain was used to inoculate solid LB (Table 2.3) (with appropriate antibiotics) and incubated for at least 18 h at 37°C. An individual colony was picked and used to inoculate liquid LB, M9 or SOC growth media (Table 2.3) where appropriate and further incubated for at least 18 h at 37°C with 180 rpm shaking.

2.4.1 PREPARATION AND TRANSFORMATION OF E. COLI

The *E. coli* strain was prepared as *above* and incubated for at least 18 h at 37°C in LB media with 180 rpm shaking. The culture was diluted into 100 mL of fresh LB media (no NaCl) and left to grow to mid-log phase (A_{600} = 0.4-0.6). The culture was washed three consecutive times by centrifuging at 4600 rpm at 4°C for 10 min in a bench top centrifuge and then resuspended into 25 mL, 5 mL and a final 1 mL of ice-cold 10% glycerol (Table 2.4). Aliquots were stored at -80°C.

A single aliquot was left to defrost on ice. 50 µL was mixed with 20-50 ng of DNA and added to an ice-cold 0.2 mm electroporation cuvette (BioRad, USA). The mixture was electroporated using a PowerPac Basic Electroporator (BioRad, USA) at 25 µF (capacitance), 2.5 kV (voltage) and 200 ohm (resistance). Immediately following electroporation, the cells were transferred into 1 mL pre-warmed liquid SOC media (Table 2.3) and allowed to recover for 1 h at 37°C. The cells were plated onto selective solid LB and incubated for ~24 h at 37°C. Negative controls containing competent *E. coli* cells but no DNA were treated identically.

2.5 General Mammalian Growth Conditions

The composition of the mammalian bladder cell growth media used throughout this thesis is outlined in Table 2.3. The growth factor supplement HKGS (Table 2.4) was always added to the EpiLife media, and appropriate antibiotics.

2.5.1 General Growth Conditions of PDO7i Bladder Cell Infection Model

Human bladder cell line PD07i was routinely cultured at 37°C with 5% CO_2 to confluency as previously described (*47*). Trypsin-EDTA (0.5%) followed by addition of 1x Defined Trypsin Inhibitor (DTI) (Table 2.4) was routinely used to release the bladder cells from base of flasks, as described by the manufacturer of these reagents. Bladder cells were grown to confluent and counted using a Coulter Counter (Beckman) to a minimum of $1x10^5$ cells/mL.

2.6 General Methods And Protocols Used in this Thesis

Various methods and protocols were used throughout this thesis and these are outlined below. Any specific changes from the manufacturer's instructions are noted where appropriate. A more detailed and specific methods sub-section can be found throughout each results chapter where that method was used.

2.6.1 PREPARATION AND EXTRACTION OF DNA FROM BACTERIA

Genomic and plasmid DNA was purified from *E. coli* using the Bioline Isolate II genomic DNA and plasmid DNA kits, respectively (Bioline, Australia). Elution was routinely performed at 70°C with a final volume of 30 µL. Concentrations of DNA samples were determined using either the Qubit 2.0 fluorometer or the NanoDrop 100 spectrophotometer.

2.6.2 POLYMERASE CHAIN REACTION (PCR)

PCR was used routinely throughout this thesis to amplify DNA for various purposes. In general, PCR was performed at a final volume of 50 µL and contained 5 – 20 ng of template DNA (prepared as per *Section 2.6.1*). Oligonucleotide primers were manufactured by Macrogen Inc (South Korea) or Integrated DNA Technologies (IDT) and dissolved in TE buffer to a concentration of 100 µM and stored at -20°C. For list of oligonucleotide primers, refer to the respective results chapter where that oligonucleotide primer was used. Each PCR reaction contained 1 mM primer pair, 10 mM dNTPs, 1% DMSO, Phusion HF reaction buffer, Phusion HF DNA polymerase and nuclease-free water. The reactions were performed according to Phusion instructions (New England Biolabs). Briefly, all reactions were subjected to an initial denaturation temperature of 98°C for 2 min and ended with a final extension at 72°C for at least 5 min. The annealing temperature and time (and 30x cycle) is specified in each results chapter where appropriate. The PCR was performed using a thermo-cycler (Eppendorf Mastercycler). All PCR products were analysed and visualised using agarose gel electrophoresis (see *Section 2.6.3*). PCR products were routinely purified

using the Bioline Isolate II PCR clean-up kit as per the manufacturer's instructions (Bioline, Australia).

2.6.3 AGAROSE GEL ELECTROPHORESIS

For most experiments, a 1% TBE agarose gel (Bioline, Australia) was prepared with the addition of 1 µL Gel-Red stock (Table 2.4). The percentage of TBE agarose gel varied depending on the size of the DNA, however the percentage never exceeded 2%. DNA loading buffer (Invitrogen, USA) to a final concentration of 1x was added to each DNA sample and loaded into the gel wells. For most experiments, 50–100 ng of DNA was loaded per well. Electrophoresis of the gel was routinely run at 120 V for 45 min and visualised via UV light (254 nm) using a transilluminator (Ingenius3: Syngene version 1.5.0) and photographed using a charge-coupled device camera (Synoptics CAM-FLXCM).

Alternatively, when performing agarose gel excision and purification of DNA, a 0.5% TAE (Table 2.5) agarose gel was prepared. Electrophoresis of the TAE agarose gel was routinely run at 60 V for 120 min. The DNA bands were visualised similarly and purified using the Bioline Isolate II PCR clean-up kit (Bioline, Australia). Concentrations of DNA samples were determined using either the Qubit 2.0 fluorimeter or the NanoDrop 100 spectrophotometer.

2.6.4 GENERAL MULTIPLEX DNA SEQUENCING PREPARATION

DNA was diluted to 5 ng using nuclease-free water and DNA sequencing library preparation was performed following Illumina instructions. Briefly, 5 ng of DNA was mixed with 10 µL of 2x Tagmentation Buffer and 5 µL of Tagmentation Enzyme (diluted 1/50 in TE buffer and 100% glycerol) (Illumina, USA) and allowed to incubate for 5 min at 55°C. A volume of 5 µL of 0.2 % SDS was immediately added to stop the tagmentation reaction and incubated for 5 min at 24°C. Following tagmentation, samples underwent 28 cycles of PCR enrichment at an annealing temperature of 55°C with KAPA HiFi Library Amplification Kit (KAPA) (see *Section 2.6.2*). Each DNA sample was amplified with

either Nextera unique-barcoded primers or custom-made transposon-specific primers, depending on the experiment. The PCR product was pooled to equimolar concentrations and underwent SPRI-select magnetic bead clean-up (Beckman Coulter) (bead ratios are detailed in the appropriate results chapters). Bead clean-up was routinely used to select for DNA of 200-800 bp. The sequencing libraries were quantified on a 2100 Agilent Bioanalyzer (Agilent Technologies), before dilution to 10 pM and sequenced on either an Illumina MiSeq or HiSeq sequencer. PhiX genomic adapter-ligated DNA control library (Illumina) was routinely included at 5%, unless otherwise stated.

2.6.5 GENERAL ANALYSIS OF DNA SEQUENCE DATA

Sequence reads from the Illumina MiSeq or HiSeq were routinely separated according to their unique index barcodes and underwent quality control analysis and filtering using the FastQC software tool (Andrew, S. 2010) version 0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), unless otherwise stated. Sequence read mapping to the *E. coli* genomes were done either using Bowtie2 software (*127*) or SMALT short read mapper software (*128*) (read mappings are detailed in the appropriate results chapters). The frequency of mapped reads and the analysis of data were determined with deepTools2 (*129*), SAMtools (*130*) and Artemis (*131*) software programs. Artemis was used to visualize the alignments (Artemis, version 13.0).

Chapter 3

High-throughput Sequencing of Sorted Expression Libraries Reveals Inhibitors of Bacterial Cell Division

DISCLOSURE

I acknowledge that this chapter has been published in the journal *BMC Genomics* under the title "High-throughput sequencing of sorted expression libraries reveals inhibitors of bacterial cell division" (refer to page *vi* for list of publications). I certify that I carried out significant work presented in this manuscript.

- Daniel G. Mediati and Iain G. Dugin conceived and designed the experiments.
- Daniel G. Mediati, Catherine M. Burke, Shirin Ansari and Iain G. Duggin performed the experiments and acquired the data.
- Daniel G. Mediati and Iain G. Duggin analysed and interpreted the data.
- Daniel G. Mediati and Iain G. Duggin wrote the paper.
- Daniel G. Mediati, Catherine M. Burke, Shirin Ansari, Elizabeth J. Harry and Iain G. Duggin revised and proof-read the manuscript.

3.1 INTRODUCTION

Bacterial cell division genes were first discovered in filamentous, temperature-sensitive (*fts*) mutants (*132*). Over 50 years later, it is expected that many more genes involved in the mechanism and regulation of division remain to be discovered (*60, 74*). In addition to identifying these, it is of great interest to identify inhibitors of cell division, as potential anti-microbial agents or research tools (*133, 134*). Filamentation is also observed in nature as a conditional response. For example, in UPEC, extensive, reversible filamentation occurs during the dispersal phase of bladder cell infection (*17*).

Bacterial cell division starts with the identification of an appropriate mid-cell site and the assembly of division proteins into a ring-shaped structure around the cell, at the inner surface of the cell envelope. This develops into a structure known as the divisome, which constricts the cell and builds new cell poles for the two daughter cells (*74*). Most of the known regulators act at the initial assembly stage, although subsequent cell constriction and closure by the mature divisome must also be coordinated with chromosome replication and segregation (*59*).

In the *E. coli* cell cycle, known regulators of division include the nucleoid occlusion, Min protein, and Ter-linkage systems (*135*). Nucleoid occlusion inhibits cell division in areas where the bacterial chromosomes (nucleoids) are located. This involves the binding of the SlmA protein to specific sites in the chromosome, where it inhibits polymerisation and assembly of the tubulin-like division protein, FtsZ (*136*). The MinCD proteins concentrate on the inner membrane towards the poles of the cell, inhibiting divisome assembly there (*137*). Overproduction of SlmA-DNA or MinCD results in filamentation (*83, 136*). The Ter-linkage system provides positive regulation of division. Interactions between MatP, bound specifically to sites in the chromosome replication terminus region, and the early-stage division protein ZapB, promote divisome assembly at mid-cell (*138*). This appears to help coordinate chromosome segregation and cell division (*138, 139*). It has been recognised that as yet unknown cell cycle regulators of division are also likely to be present in *E. coli* (*138, 140*).

When compared to the processes involved in housekeeping cell cycle regulation of division described above, fewer regulatory pathways that lead to arrest of cell division and filamentation under specific conditions have been characterised. The best-known case is the DNA damage (or "SOS") response (*141*). DNA damage is sensed by the RecA protein, which then triggers auto-degradation of the SOS repressor LexA, and expression of SOS genes. One of the SOS genes encodes the FtsZ assembly inhibitor, SulA. The resulting filamentation allows time for DNA repair; when the SOS response is switched back off, the filamentous cells can resume division to generate short rods with intact chromosomes (*83*).

Another important example of conditional filamentation is seen during urinary tract infections (UTI). During a late stage of the intracellular infection cycle, UPEC grow into extensive filaments that may be hundreds of microns long and erupt from the overwhelmed host cell (*17*). UPEC filaments are thought to help avoid phagocytosis by macrophages during infection and may improve bacterial dispersal and surface adhesion (*142*). It has been suggested that UPEC filamentation in UTI occurs via a SulA-dependent pathway (*58*). However, extensive SulA-independent filamentation by UPEC in experimental infection was recently observed, which instead required a known cell division protein, DamX (*56*). DamX normally contributes to cell envelope synthesis during division (*143*), and was previously noted to cause extensive filamentation when overproduced (*88, 144*). The exact roles of these proteins in UPEC need clarification, and the signaling pathway(s) that lead to filamentation during UTIs remain to be identified.

Here, we report the development and utilisation of a high-throughput DNA sequencing approach to identify genes, or genomic DNA fragments, that result in filamentation after induced over-expression from a plasmid-based DNA library. By applying this method to a UPEC strain capable of UTI-associated filamentation, we have identified new and known inhibitors or mediators of normal *E. coli* cell division as well as potential components of the UTI-associated or other filamentation responses.

3.2 RESULTS

3.2.1 A HIGH-THROUGHPUT SCREEN BASED ON FLOW CYTOMETRIC SORTING OF DNA EXPRESSION LIBRARIES

In order to identify genes or genomic regions that contain the potential to induce reversible filamentation, a plasmid-based expression library of UPEC genomic DNA fragments was first constructed. Genomic DNA fragments of 1-5 kb from UPEC cystitis isolate UTI89 (*29*) were ligated to the vector pBAD24, to place them under control of the arabinose-inducible promoter (P_{BAD}). Approximately 35,000 colonies arising from transformation of *E. coli* JW0941-1 (K-12 background, Δ*sulA*) were resuspended and pooled for storage. The Δ*sulA* host strain was selected in order to identify mediators of filamentation that are independent of the well-characterised SulA division inhibitor.

Transcription from P_{BAD} was induced by the addition of 0.2% (w/v) L-arabinose to a mid-log phase culture of the library, for 4 generations. These conditions were chosen to allow strong induction of transcription from $P_{BAD,}$ such that moderate or strong inhibitors of division can function, whilst providing sufficient time for these clones to generate filaments of lengths that were amenable to purification by flow cytometry (*88*). Flow cytometry cell sorting was then used to collect cells that had developed a filamentous morphology. Control cultures of JW0941-1/pBAD24 ("short" cells; Fig. 3.1A) and JW0941-1/pLau80 ("filamentous" cells, caused by over-production of FtsZ-YFP; Fig. 3.1B) were analysed too. Approximately 20,000 cells that displayed a filamentous phenotype were isolated from the induced expression library (Fig. 3.1C-D, "filamentous" gate) on the basis of the side-scatter (pulse width) parameter, and they were then transferred to LB medium including glucose—to suppress transcription from P_{BAD}. These cells were then incubated overnight at 37°C to allow reversal (division) of filaments and growth of viable bacteria (ongoing growth of filaments was expected to be eventually lethal). Plasmid DNA from this filamentous-sorted library was then purified and subjected to high-throughput sequencing, to identify the cloned UTI89 DNA that caused inducible filamentation. A reference sample from the same culture, treated in the same manner but not sorted, was also sequenced in

the same multiplexed sequencing run. The whole experiment was performed in duplicate, starting with two frozen aliquots of the library. The sequence data have been deposited in the European Nucleotide Archive (accession number PRJEB23331).

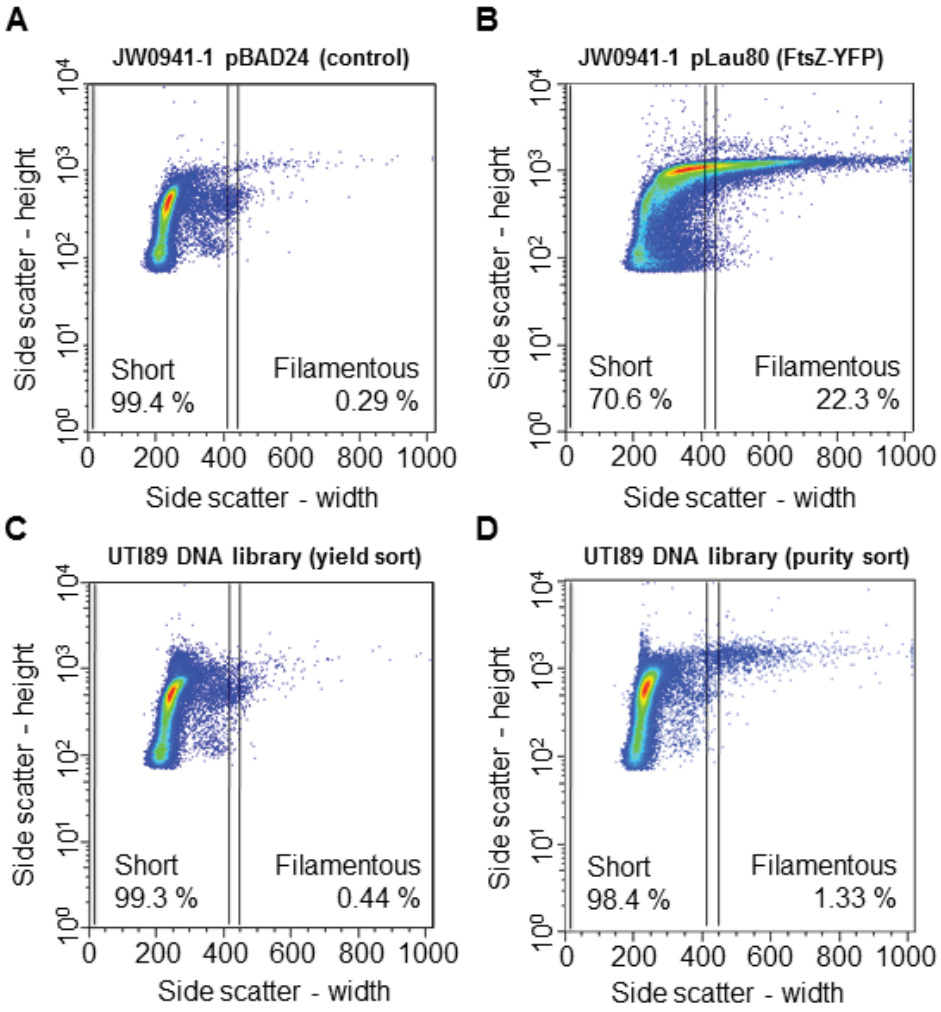


Figure 3.1 Analysis and purification of filamentous bacteria from the UTI89 DNA-expression library by flow cytometry.

Flow cytometry was carried out with the control populations of strains JW0941-1/pBAD24 (A) and JW0941-1/pLau80 (containing FtsZ-YFP, which induces strong filamentation) (B). The sorted populations of the UTI89 gDNA expression library in JW0941-1 are also shown: an initial high-speed, high-yield sort (C) and a subsequent lower speed "purity sort" to further enrich for filamentous cells (D). Dots represent single events plotted as side-scatter values of peak height and peak width, which correlates with filament length. Event density is colour coded with a heat-map of red (high density) to blue (displaying individual events). The gates for sorting are indicated with vertical lines, and the percentages of events contained within that gate are indicated. The "short" gate was determined to encompass greater than 99% of the control bacterial population (short cells), JW0941-1/pBAD24 (A).

The mapped sequence-read frequencies for each sample showed that the expected majority of reads mapped to the vector pBAD24, with most of the remainder matching the UTI89 chromosome or native plasmid, pUTI89 (Table 3.1). The reads matching pBAD24 were excluded from further analysis. A plot of the reference library's distribution of read depth per bp showed that the library covers 100% of the UTI89 chromosome by at least 1 read per bp, while <1% was covered by at least 30 reads per bp (reference-unsorted 2 data, Supp. Data Fig. S1A).

Table 3.1 Read frequencies from sequencing of plasmid libraries

Sample	No. of reads generated	No. of reads mapped to pBAD24 (%)	No. of reads mapped to UTI89 (%)
Filamentous-sorted 1	3,127,306	2,021,876 (64.6)	802,924 (25.7)
Reference-unsorted 1	2,093,673	1,414,249 (67.6)	470,937 (22.5)
Filamentous-sorted 2	4,462,451	2,279,041 (51.0)	1,223,720 (27.4)
Reference-unsorted 2	20,600,235	9,353,917 (45.4)	5,887,097 (28.6)

We then detected genomic regions that were significantly enriched in reads from the filamentous-sorted samples, compared to the respective reference-unsorted samples. A global analysis of the degree of enrichment of specific genomic regions is shown in Fig. 3.2A; the filamentous-sorted sample demonstrated a strong enrichment of reads in relatively fewer 10 kb genomic windows (or bins) compared to the reference-unsorted library. For example, 20% of the mapped reads from the filamentous-sorted sample were located in ~ 90% of the bins, while the remaining 80% of mapped reads were clustered in only 10% of the bins (dotted line, Fig. 3.2A). In contrast, 20% of mapped reads from the reference-unsorted sample were located in ~ 30% of bins, while remaining 80% of mapped reads were distributed throughout ~ 70% of bins. This comparison demonstrates the effectiveness of our combination of cell sorting and high-throughput sequencing of plasmid DNA in identifying enriched genomic regions. We characterised this enrichment further by plotting the number of mapped sequencing reads in 1 kb windows, to obtain an overview of the relative frequency of mapped reads along the entire chromosome. As may be seen in Fig. 3.2B, strong peaks of enrichment of read counts in the filamentous-sorted sample compared to the reference-unsorted sample were evident at loci distributed throughout the chromosome.

3.2.2 IDENTIFYING GENOMIC REGIONS THAT ENCODE MEDIATORS OF BACTERIAL FILAMENTATION

We identified the significantly enriched regions in the filamentous-sorted data by using the peak-detection software MACS (*145*). The input data were the combined paired-end reads aligned to the UTI89 genome, from the filamentous-sorted and reference-unsorted samples. The complete annotated output of enriched regions from each replicate, and a list of the 55 regions that were found common to both replicate screens, are provided in Supp. Data Table S1. We ranked these regions with respect to their significance score ($-\log_{10}$[P-value]), and found that the number of enriched regions identified by the two replicate screens corresponded strongly, above a significance score of ~70 **(**Supp. Data Fig. S1B). Furthermore, the fraction of enriched regions specifically in common between replicates was maintained near the maximum correspondence (~20%), at significance scores ≥70, but

reduced markedly at lower significance scores (Fig. 3.2C). By applying this high-stringency significance threshold (≥70), 23 regions were identified in common between both replicate screens (Table 3.2). These regions are therefore most strongly implicated in causing filamentation.

The list in Table 3.2 includes regions that contain genes with a known involvement in cell division and its regulation, such as *ZapB* (*146*) and *minCDE*. Fig. 3.2D shows the data for the *minCDE* locus, encoding the well-known spatial regulator of division. The strongest region of relative enrichment coincides with the *minC* gene, which is the FtsZ inhibitor of the system (*147*). Interestingly, the region of enrichment extended downstream of *minCDE*, including a conserved gene of unknown function (C1358), and a pseudogene C1357 (the homolog of *ycgI* in *E. coli* K-12), suggesting that these genes or fragments in this region might affect cell division too. Alternatively, they may be represented due to their linkage to *minCDE* in our library. Some of the strongest hits were located in prophages or prophage remnants that have previously-demonstrated roles in blocking cell division, including *kil1* and *kil2* (λ-prophages), *dicB/dicF* (Qin prophage), and the *ymfN-ymfK* region of the CP-933O (e14) prophage. There were also numerous genes encoding membrane-associated proteins, such as *ompA* and *rfaG* (*waaG*), and others involved in peptidoglycan cell wall structure, including *NlpD* (Supp. Data Table S1A). We also identified a region containing *damX*, which plays a role in cell division (*87*) and was previously shown to inhibit cell division when overexpressed, causing a filamentous phenotype (*88*). DamX has recently been specifically implicated in the conditional filamentation seen during urinary tract infections (*56*).

Amongst the expanded list of 55 identified regions that were common to both screens at the default threshold (Supp. Data Table S1B) were other genes with a known involvement in cell division, such as *zapA* (*71*), suggesting that this list also contains significant inducers of filamentation. We expect that some regions only identified with strong significance in one of the screens may also contain genuine inducers of filamentation (Supp. Data Table S1A), but they might have been missed in one of the screens due to limited sampling of cells from the library by the flow cytometry. Other regions encoding a capacity to cause

filamentation may not have been cloned in the correct position with respect to P_{BAD} in the library, and were therefore not detected in either screen.

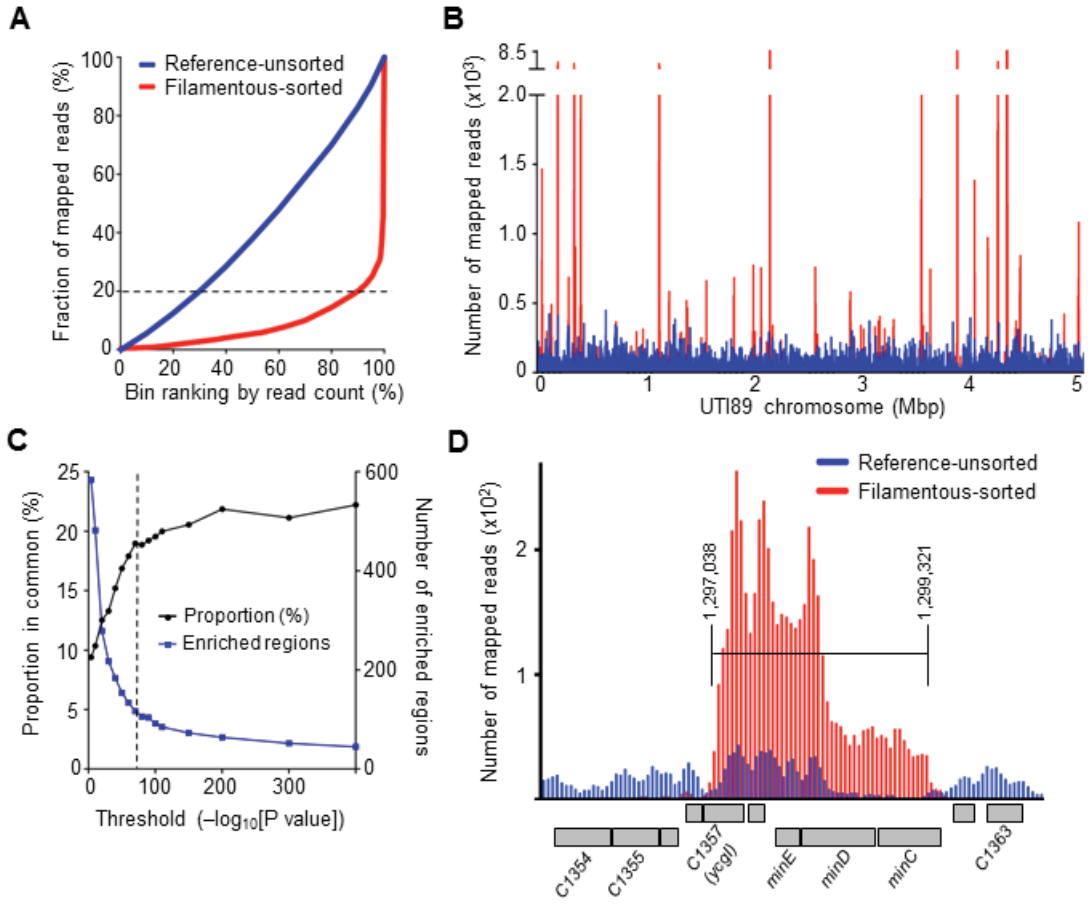


Figure 3.2 High-throughput DNA sequencing of plasmid libraries from the reference-unsorted and filamentous-sorted populations

(A) The distribution of mapped read-counts in 10 kb UTI89 chromosomal bins. Bins were ranked on their read count, and plotted against the percentage of total mapped reads. Deviation from the diagonal reflects the extent of enrichment in particular bins. (B) The distribution of read depth along the ~5.06 Mbp UTI89 chromosome in the filamentous-sorted sample (red) and the reference-unsorted sample (blue), with a bin size of 1 kb. (C) The proportion of enriched regions found in common in both biological replicates versus their corresponding threshold significance score ($-\log_{10}$[P-value]) (left-hand y-axis), and the total number of enriched regions from both replicates (right-hand y-axis). The score of 70 (dotted line) represents the selected threshold criteria for display in Table 2. (D) The number of reads mapped to the region 1,295,150 – 1,302,200 bp of UTI89 in the filamentous-sort (red) and the reference-unsort (blue) samples within a bin window of 50 bp. Large read depth numbers in the filamentous-sort sample (replicate 1, red) can be seen against the reference-unsort sample from the ycgI homolog to the minCDE genes (the region identified by MACS is shown by the bracket).

Table 3.2 UTI89 chromosomal regions causing filamentation – high stringency hits in both screens

Genomic Region (bp)[+]	Locus ID	Gene Name(s)*	COG[∞]	Length (bp)	Fold Enrichment[#]	Significance score[#]
14519 - 15641	UTI89_C0017	*dnaJ*	O	1123	6.924	197.251
	UTI89_C0018	*gef (hok)*	R			
361550 - 364162	UTI89_C0348	*yahD*	R	2613	17.098	934.951
	UTI89_C0349	*yahE*	S			
	UTI89_C0350	*yahF*	C			
640717 - 641788	UTI89_C0629	*ybeM*	R	1072	4.302	112.532
	UTI89_C0630	*tatE*	U			
	UTI89_C0631	*lipA*	H			
690222 - 691856	UTI89_C0684	*glnS*	J	1635	79.229	8503.989
1016170 - 1016793	UTI89_C1021	*matP*	MD	624	9.121	575.491
	UTI89_C1022	*ompA*	M			
1139704 - 1141232	UTI89_C1147		S	1529	9.555	321.27
	UTI89_C1148	*(cbeA/yeeU)*	S			
	UTI89_C1149	*(cbtA/yeeV)*	S			
	UTI89_C1150		S			
1245504 - 1247434	UTI89_C1276	*bet*	S	1931	415.23	43744.932
	UTI89_C1277	*gamW*	R			
	UTI89_C1278		S			
	UTI89_C1279	*kil1*	R			
	UTI89_C1280	*λ-cIII*	S			
	UTI89_C1281	*(ea10)*	S			
	UTI89_C1282	*ral*	R			
1297038 - 1299321	UTI89_C1357	*(ycgI)*	S	2284	4.727	174.72
	UTI89_C1358		S			
	UTI89_C1359	*minE*	D			
	UTI89_C1360	*minD*	D			
	UTI89_C1361	*minC*	D			
1392009 - 1395131	UTI89_C1459	*exoO*	SR	3123	70.368	10266.027
	UTI89_C1460		S			
	UTI89_C1461	*dicB*	D			
	UTI89_C1462	*dicF*	RD			
	UTI89_C1463	*ydfC*	S			
	UTI89_C1464	*ydfB*	S			
	UTI89_C1465	*ydfA*	S			
	UTI89_C1466	*dicA*	KT			
	UTI89_C1467	*dicC*	R			
1961291 - 1963326	UTI89_C2055	*zwf*	G	2036	18.951	1044.291

Genomic Region (bp)[+]	Locus ID	Gene Name(s)*	COG[∞]	Length (bp)	Fold Enrichment[#]	Significance score[#]
2011010 - 2012850	UTI89_C2102	*(yecI/ftnB)*	P	1841	6.674	159.718
	UTI89_C2103	*yecJ*	S			
	UTI89_C2104		S			
	UTI89_C2105	*yecR*	SR			
	UTI89_C2106	*ftnA*	P			
2632799 - 2635972	RS13095		S	3174	14.906	1023.552
	UTI89_C2682	*λ-cIII*	S			
	UTI89_C2683	*kil2*	R			
	UTI89_C2684		S			
	UTI89_C2685		S			
	UTI89_C2686	*erf*	S			
	UTI89_C2687		S			
	UTI89_C2688	*hkaJ*	S			
	UTI89_C2689	*hkaH*	S			
	RS13145		S			
2927393 - 2928805	UTI89_C2990		G	1413	193.228	17052.854
	UTI89_C2991		S			
	UTI89_C2992		S			
	UTI89_C2993		S			
2933266 - 2935042	UTI89_C3000	*ymfN*	K	1777	36.001	3042.465
	UTI89_C3001	*ymfM*	RD			
	UTI89_C3002	*ymfL*	S			
	UTI89_C3003	*ymfT*	S			
	UTI89_C3004	*ymfK*	KT			
3029776 - 3031529	UTI89_C3112	*nlpD*	M	1754	92.257	23091.689
	UTI89_C3114	*pcm*	O			
3162699 - 3166774	UTI89_C3238	*tas*	C	4076	8.133	372.141
	UTI89_C3239	*lplT*	U			
	UTI89_C3240	*aas*	IQ			
3426907 - 3428106	UTI89_C3489	*glnE*	OT	1200	12.824	1200.04
	UTI89_C3490	*ygiF*	S			
3553759 - 3555227	UTI89_C3621	*ispB*	H	1469	11.783	666.362
	UTI89_C3622	*nlp*	K			
	UTI89_C3623	*murA*	M			
3723755 - 3724973	UTI89_C3832	*rpoB*	K	1219	42.391	7960.966
	UTI89_C3834	*rplL*	J			
	UTI89_C3835	*rplJ*	J			
3767241 - 3770673	UTI89_C3884	*rpe*	G	3433	61.279	5157.364
	UTI89_C3885	*dam*	L			

Genomic Region (bp)[+]	Locus ID	Gene Name(s)*	COG[∞]	Length (bp)	Fold Enrichment[#]	Significance score[#]
	UTI89_C3886	*damX*	D			
	UTI89_C3887	*aroB*	E			
3921231 - 3922700	UTI89_C4038	*hmuV*	P	1470	18.962	1402.281
	UTI89_C4039	*yhiD*	S			
4049515 - 4052967	UTI89_C4158	*yibP (envC)*	D	3453	51.663	12724.513
	UTI89_C4159	*yibQ*	S			
	UTI89_C4160	*yibD*	M			
	UTI89_C4162	*tdh*	ER			
4997073 - 4999597	UTI89_C5098		S	2525	14.764	584.563
	UTI89_C5099		S			
	UTI89_C5100		S			
	UTI89_C5101	*yfdN2*	S			
	UTI89_C5102		S			

[+] *Coordinates were determined from Genbank accession NC_007946.1 as the minimal region identified in common between both replicates, with an overlap >100 bp.*

***** *Genes within or overlapping the hit region are shown, and named based on characterized homologs in other strains; alternative names may also be given in parentheses. Hypothetical or proteins with no characterized homologs are left blank.*

[∞] *COG: Clusters of Orthologous Groups functional classification code.*

[#] *The fold-enrichment and significance score (log_{10}[P-value]) are shown as the means from the two replicate screens.*

3.2.3 IDENTIFICATION AND VERIFICATION OF DNA FRAGMENTS THAT CAUSE FILAMENTATION

To investigate whether particular open reading frames (ORFs) from the identified DNA regions were responsible for the observed filamentation, we initially selected eight candidate ORFs from six of the hit regions of differing statistical strength, including some that were only observed in one screen and others that overlap the edge of a hit region (Supp. Data Table S1A). These ORFs were individually cloned into pBAD24: *fsaA, ybiY, aceE, pdhR, sdaA, lipA, pptE* and *pflC*. Two of the cloned ORFs produced a reproducibly elongated or filamentous phenotype after the controlled induction period, namely *pptE* and *pdhR*, which showed a mean cell volume of 3.6 μm^3 and 2.8 μm^3, respectively, compared to 1.4 μm^3 for the strain containing the vector-only (BW25113/pBAD24) (Fig. 3.3).

As may be seen in Fig. 3.3A-B, the cell size distributions of cultures expressing *pptE* and *pdhR* differed. Overexpression of *pdhR*, encoding the pyruvate dehydrogenase operon repressor, caused a mixed population of short cells and filaments. By comparison, overexpression of *pptE*, encoding a putative phosphoenolpyruvate-protein phosphotransferase, caused a much higher proportion of the cells to form filaments. Both genes caused their same characteristic filamentation in the absence of the SOS response regulator *recA* (Fig. 3.3B**)**. Strikingly, staining of the membrane (FM4-64) and DNA (Hoechst 33342) in the *pptE* overexpression strain revealed irregularly positioned nucleoids in the filaments, and discontinuous zones of strong FM4-64 staining that corresponded to apparent inclusions in phase-contrast (Fig. 3.3C). Similar inclusions were also present during conditions expected to result in significantly lower *pptE* expression (0.02% L-arabinose), although these conditions did not cause significant filamentation (Supp. Data Fig. S2). These observations suggest that the inclusions may be lipid-rich, and that filamentation occurs in this strain as a secondary effect, requiring higher levels of expression.

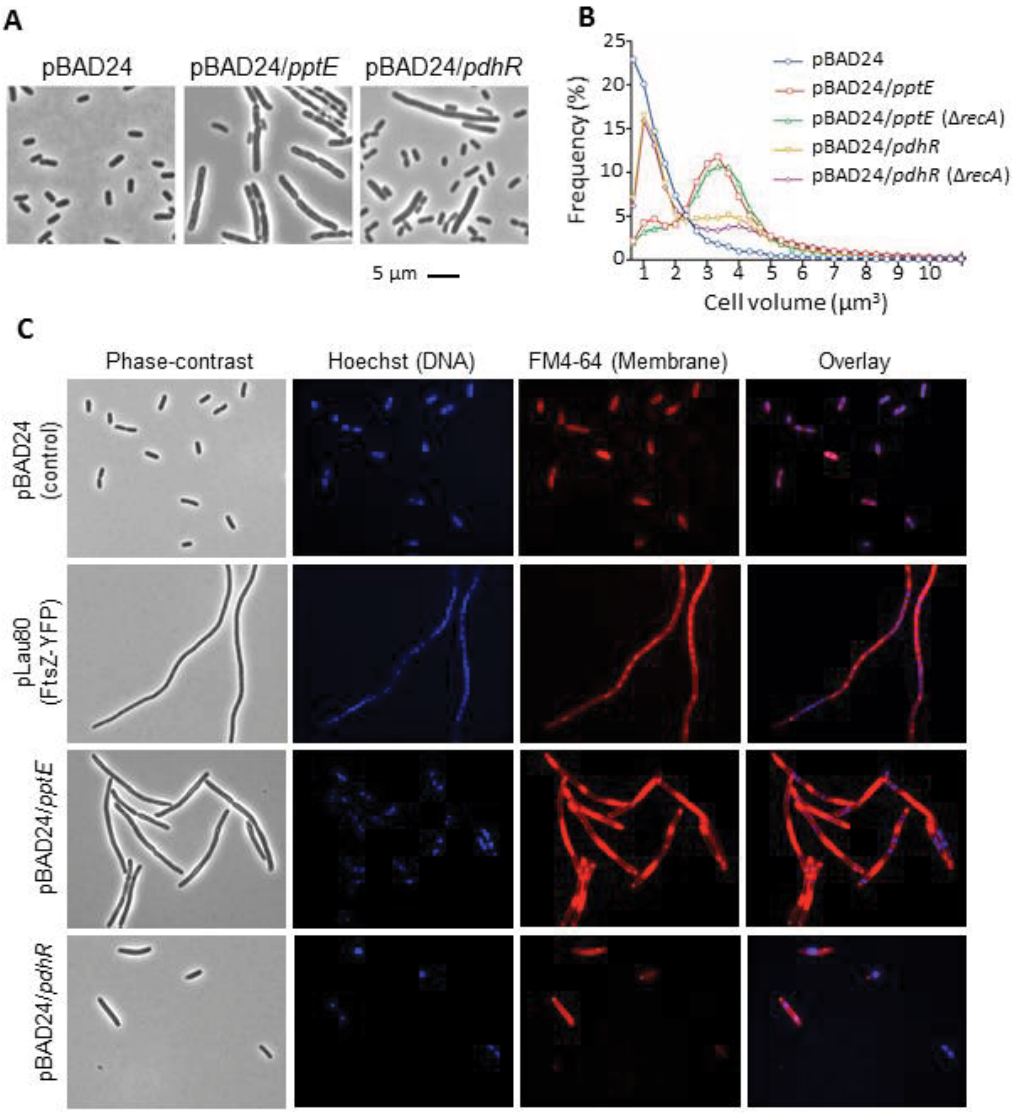


Figure 3.3 Overexpression of the pptE and pdhR ORFs cause E. coli filamentation independent of recA (SOS response)

The pptE and pdhR ORFs from UTI89 were cloned into pBAD24, transformed into BW25113, and BW25113 (ΔrecA), and cultures induced with 0.2% L-arabinose in M9 medium. Fixed cells were analysed by microscopy (A) and Coulter cytometry to obtain cell volume distributions (B). (C) Fixed cells were stained using the Hoechst 33342 (DNA) and FM4-64 (membrane) and imaged by fluorescence microscopy. An overlay of the Hoechst and FM4-64 channels is also shown, indicating the obvious anti-correlation between the two stains, suggesting a possible physical exclusion of these features.

The six tested ORFs that did not show an obvious filamentous phenotype had average cell lengths of 1.3–1.6 μm^3 (Supp. Data Fig. S3). We considered it possible that alternative ORFs, a combination of multiple ORFs, a sub-fragment of an ORF, or an expressed intragenic region might have caused the observed enrichment in in the screens. We further investigated two of the hit regions in which the preliminary whole-ORF testing above did not reveal filamentation, namely the regions containing *ybeM-tatE-lipA* (2,186 bp) and *ybiY-fsaA-moeB-moeA* (3,754 bp). Sub-libraries of the two PCR-amplified regions were created by cloning FatI partial-digest products into pBAD24, and 48 transformants from each were then placed in M9 medium with L-arabinose for 7 h before microscopy. A total of 12 clones from the two sets were classified as filamentous, and the cloned fragments from these were sequenced (Fig. 3.4 and Table 3.3).

Seven of the filamentous clones contained fragments from within the *ybeM* ORF that were out-of-frame with respect to P_{BAD} in the library. Two overlapping fragments of 112 bp and 87 bp were found in six of these clones; both orientations were observed suggesting that two different products could trigger filamentation in this fragment, or the region could conceivably contain its own Ara-inducible division inhibitor. The remaining clone (D10, Table 3.3) contained a downstream fragment of only 29 bp within ybeM (encoding an amino acid sequence MFLQRRGAH). From the *ybiY-fsaA-moeB-moeA* region, four of the five sub-clones identified as filamentous contained a 161 bp fragment in either orientation from within the *ybiY* ORF, despite the previous observation that the full-length *ybiY* ORF did not cause filamentation (Supp. Data Fig. S3B). Also, a larger downstream region of 2,197 bp was identified (clone A11, Table 3.3). This region spanned across the majority of the *ybiY-fsaA-moeB-moeA* region (Fig. 3.4K), with transcription in pBAD24 expected to start from within *ybiY* in the reverse orientation with respect to the expected transcription of *ybiY* in its genomic context. These findings show that sub-ORF in-frame fragments (e.g. clone C1 from *ybiY*) and out-of-frame fragments (e.g. clones H8, H9 and D10 from within *ybeM*) can both induce strong filamentation, even though the full-length ORFs (*ybiY* and *ybeM*) did not (Supp. Data Fig. S3B), strongly suggesting that these fragments artificially (experimentally) inhibit cell division.

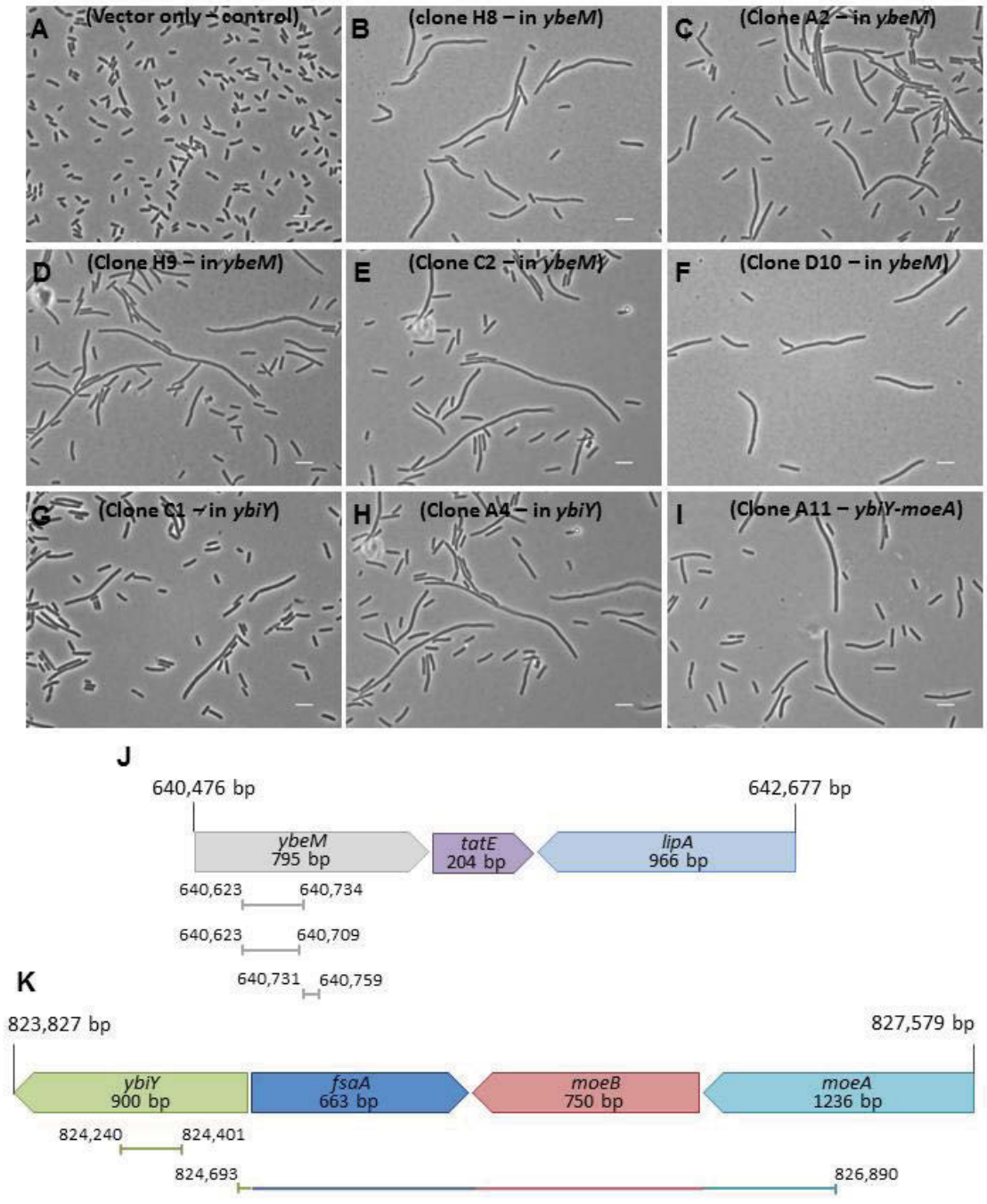


Figure 3.4 Partial fragments of ORFs cause filamentation.

FatI fragment sub-clones from the ybiY-fsaA-moeB-moeA and ybeM-tatE-lipA regions were screened (48 colonies each) for filamentation after growth and induction for 7 hours at 37°C. The indicated clones were identified as causing filamentation. (A) BW25113/pBAD24 control, (B) Clone H8 (in ybeM), (C) Clone A2 (ybeM), (D) Clone H9

(ybeM), (E) Clone C2 (ybeM), (F) Clone D10 (ybeM), (G) Clone C1 (ybiY), (H) Clone A4 (ybiY), (I) Clone A11 (ybiY-moeA). (J) The ybeM-tatE-lipA region and sub-clones identified as capable of causing filamentation; three fragments throughout the 7 clones, of 112 bp, 87 bp and 29 bp were identified within the ybeM gene. (K) Map of fragments within the ybiY-fsaA-moeB-moeA locus capable of causing filamentation.

Table 3.3 Genomic FatI fragments from the ybiY-moeA and ybeM-lipA regions of UTI89 that caused a filamentous phenotype

Genomic region (bp)	Locus ID	Gene	Length (bp)	Clone
640623-640734	UTI89_C0629	*ybeM*	112	H8, *A2
640623-640709	UTI89_C0629	*ybeM*	87	H9, *C2, *B4, *C8
640731-640759	UTI89_C0629	*ybeM*	29	D10
824240-824400	UTI89_C0827	*ybiY*	161	C1, *A4, *B2, *E5
824693-826889	UTI89_C0827,	*ybiY,*	2197	*A11
	UTI89_C0828,	*fsaA,*		
	UTI89_C0829,	*moeB,*		
	UTI89_C0830	*moeA*		

** Indicates the expressed fragments in these clones were in the opposite orientation in pBAD24 relative to the expected wild-type transcription of the indicated genes in UTI89.*

3.3 Discussion

We have developed a widely applicable high-throughput screen, combining flow cytometry cell sorting with high-throughput sequencing of a plasmid-based DNA expression library, to identify genes or genomic DNA fragments that act as inducible inhibitors of bacterial cell division and lead to bacterial filamentation. In this study, the filamentous clones isolated after expression of the genomic fragment library were recovered and amplified by overnight growth under conditions that suppress expression. Since unlimited growth of bacterial filaments is expected to result in lysis, the recovery period was expected to enrich for reversible inhibitors of division, which were considered likely to be more physiologically relevant. We applied the method to *E. coli* UTI89, and identified 55 genomic regions that were enriched in the replicate screens. The identified genomic regions included numerous genes with known roles in cell division and its regulation, confirming that the method successfully identifies genes of biological relevance. Several previously uncharacterised genes were also identified that may represent new inhibitors or mediators of division, or be part of the UTI-associated or other filamentation responses. In addition, we identified a number of short predicted transcripts or peptides that are not predicted to be expressed from their genomic location in wild-type cells but were detected as sub-ORF, out-of-frame, or intergenic fragments. One fragment of only 29 bp, encoding a potential peptide (MFLQRRGAH), acts as an inhibitor of division in our screens. These or other similarly identified sequences could represent novel peptide or small RNA inhibitors of bacterial division. It is therefore expected that expanded screens of large and diverse DNA libraries could reveal other new inhibitors with potential applications.

An expectation of using a genomic fragment library for the screens described here is that not all genes will be appropriately positioned to be expressed from the plasmid. Furthermore, the size of our library (~ 35,000 clones) may also somewhat limit the screens, and they are therefore not considered fully comprehensive. Consistent with this, we identified six out of the 12 genomic regions previously identified by Sanger sequencing of clones obtained after a similar flow-cytometry screen was performed with *E. coli* K-12 (DH5α) (*88*). The genomic regions in common with this study contained genes such as

damX (that has been implicated in the UTI-filamentation response pathway (*56*), *rplL-rpoB*, and two regions of prophage elements e14 (*ymfL/ymfM*) and Rac (*kil*). Another possible limitation relates to the yields obtained from the cell sorter flow cytometer (~20,000 cells per screen). Consistent with such a sampling limitation, a number of enriched regions were detected with statistical significance in only one of the two screens (Fig. 3.2 and Supp. Data Table S1A). Genes from some of these regions play known roles in cell division (e.g. *dedD*). We therefore recommend that other such regions are also considered as potential candidates for verification. These data may be directly assessed by referring to Supp. Data Table S1 and the genomic alignments available from the European Nucleotide Archive (accession number PRJEB23331).

Given our identification of both physiological and likely non-physiological classes of genomic fragments that lead to filamentation, we suggest that further screens aiming to focus on physiologically-relevant gene identification could utilise comprehensive complementary DNA (cDNA) expression libraries, whereas others focusing on artificial peptide inhibitors could use diverse small fragment libraries (e.g. from environmental or randomized synthetic DNA sources). The cDNA screen results should still be followed up by determining whether the identified ORFs relate to a specific pathway in controlling wild-type cell division, or whether they cause filamentation via an artefactual consequence of overexpression (e.g. via substantive changes to envelope physiology that secondarily inhibits normal cell division). Both approaches would benefit from the development of the flow cytometry or other methods to provide a significantly greater yield of filamentous bacteria from the expression library.

An intriguing aspect of our screen results was the identification of genes within the regions causing inducible filamentation that have known roles in central carbon metabolism, including gluconeogenesis (*zwf*) (Table 3.2 and Supp. Data Table S1). Glucose metabolism has previously been implicated in inhibiting cell division in *Bacillus subtilis*, via the glucosyltransferase, *ugtP*, which acts on FtsZ to delay cell division in the presence of UDP-glucose (*63*). A similar UDP-glucose-dependent mechanism has also been found in *E. coli* with the functional analog, *opgH* (*148*). Interestingly, our screens revealed a region

containing a fragment of *rssB* and the whole downstream *galU* gene (Supp. Data Table S1). GalU catalyzes the synthesis of UDP-glucose, which could conceivably stimulate the division-inhibition mediated by OpgH. However, previous overproduction of a GalU fusion protein did not cause filamentation (*148*). Preliminary hits from our screen such as the *rssB-galU* region need to be checked by defining the specific regions and conditions that cause filamentation.

We chose several genomic regions for following up and verifying the results of the screens, some of which encoded potentially novel division-metabolism links. Of particular note was the identification of a genomic region in one of our screens containing the *pdhR* and *aceE* genes, encoding the pyruvate dehydrogenase complex regulator and e1 subunit, respectively. We confirmed that over-expression of the PdhR ORF caused filamentation, independent of the SOS pathway (*recA*) (Fig. 3.3). The PdhR regulator is controlled by the concentration of pyruvate in the cell. When ample pyruvate is present, PdhR does not repress its target genes (including its own PDH operon), whereas in low-pyruvate conditions, PdhR specifically binds DNA and represses expression of numerous metabolic genes and the cell division gene cluster, *dcw*, which includes *ftsZ* (*149*). In this way, low pyruvate (nutrient poor) conditions might reduce the frequency of division, coupling it to the reduced cell growth rate. Our finding that *pdhR* overexpression inhibits division therefore lends support to this notion. In *B. subtilis*, pyruvate metabolism also plays an important role in regulating division; pyruvate dehydrogenase was found to affect the assembly of the FtsZ-ring in a pyruvate-dependent manner (*150*). It was suggested that this positively regulates FtsZ assembly and division in nutrient-rich conditions to appropriately couple division with growth rate. Therefore, this important final stage of glycolysis appears to be a nexus in the coupling of cell growth and division in these two highly divergent bacteria, but the underlying mechanisms that evolved to establish this regulation appear to be very different. A similar occurrence of convergent evolution was noted for the UDP-glucose-dependent mechanisms of division control, mentioned above (*148*).

The second whole-ORF that we discovered to encode a strong inducer of filamentation was *pptE* (Fig. 3.3). PptE is uncharacterised in UTI89, but is a putative phosphoenol-pyruvate-

protein phosphotransferase, that is present in extraintestinal pathogenic *E. coli* (ExPEC) strains (phylogenetic group B2) including UPEC, avian pathogenic *E. coli* (APEC) and neonatal meningitis *E. coli* (NMEC). The *pptE* gene varies in size substantially amongst *E. coli* strains. A smaller remnant of *pptE* (0.34 kb out of the total 2.38 kb ORF) is evident across some UPEC strains (*151*), although UTI89 has the full length ORF. Overexpression of *pptE* led to major cell structure and morphology effects in *E. coli* (Fig. 3.3, Supp. Data Fig. S2). It is possible that *pptE* plays an additional uncharacterised role in UTI89 cell division or structure, but an understanding of its normal physiological role(s) awaits further characterization. The overexpression phenotypes of *pptE* and other genes identified here offer new leads towards defining their roles.

Some of the strongest and most frequent inducers of filamentation from our screens were prophage genes, several of which have been previously studied as strong inhibitors of *E. coli* cell division. For example, the *kil1* gene in a λ-prophage was identified in our screens, and this gene encodes a known inhibitor of cell division that acts on FtsZ assembly (*152*). Other prophage regions in the list include a large region of the Qin prophage, containing known division inhibitors *dicB* and *dicF*. The gene *dicB* encodes an activator of the endogenous *minC* division regulator/inhibitor, whereas *dicF* is a small RNA inhibitor of *ftsZ* gene expression (*153, 154*). The *ymfN-ymfK* region of the e14 prophage was identified (Table 3.2), which was previously reported to contain an SOS-inducible cell division inhibitor, originally termed *sfiC* (*155*). Furthermore, known prophage toxins from toxin-antitoxin pairs affecting cell division, including *cbtA* (*yeeV*) (*156*) and *gef* (*hok*) (*157*), were identified. These common functions of the prophage-associated genes therefore appear to reflect a wide-spread bacteriophage strategy to block cell division during lysis, which likely aids bacteriophage proliferation or dissemination. At least some of these genes in remnant or defective prophages appear to have been co-opted by the host bacteria for regulatory purposes in stress responses (*158, 159*).

We applied our method to *E. coli* UTI89, as it might contain specific regulators of the UPEC filamentation response seen in UTIs that are not present in other strains of *E. coli*. Alternatively, the filamentation response in UTI may involve a more general response pathway that is activated strongly under the conditions bacteria experience during UTI,

such as the SOS response that was first postulated for UTI-associated filamentation (*58*). We identified seven genomic regions in the reproducibly filamentous expression clones that did not contain direct homologs in *E. coli* K-12. These included regions containing prophage genes that are implicated in inhibition of cell division (Supp Data Table S1), and some of these might have been co-opted by cells for regulatory purposes (*158*). At present it is not clear whether these or other more wide-spread genes are involved in the UPEC filamentation response. However, the relevance of our screen towards the goal of identifying the components that are potentially involved in the UPEC filamentation response is exemplified by our identification of *damX*, which was previously identified amongst a set of genes that were upregulated during the filamentation stage of infection (*56*). DamX has a known involvement in division, and its overexpression causes filamentation, particularly when the bacteria are surface bound (*56, 144*). DamX was essential for the UPEC filamentation response observed during infection (*56*). Other components of the regulatory pathway that leads to UPEC filamentation (via *DamX*) are still unknown. Interestingly, amongst our list of 55 UTI89 genomic regions causing inducible filamentation (Supp Data Table S1), there were 16 ORFs that showed elevated expression during the filamentation phase of bladder cell infection (*56*). These genes encode the outer membrane protein, OmpA (which is also under strong positive selection in UPEC strains (*102*)) and other proteins involved in peptidoglycan cell wall structure (*NlpD* and *PbpG*). Several relatively poorly characterized genes, such as ygaV, were also detected in both studies. YgaV is a predicted transcriptional regulator that is part of a two-component regulator (YgaVP) responsive to tributyltin exposure, but is of unknown function (*160*). These genes were therefore recognised as candidates that might act as regulators in the UTI-filamentation response.

A high-throughput sequencing-based method for genome-wide identification of genes and DNA fragments that encode a capacity to induce filamentation in E. coli was successfully developed and applied. This revealed genes from several prophages, as well as loci that have known or novel roles in bacterial cell division or filamentation, and has helped refine a list of candidate genes that may be involved in the UPEC filamentation response pathway seen extensively in UTIs. Several short DNA fragments that cause filamentation were also

identified and these may have potential uses as inhibitors of cell division. Our approach can be used to carry out comprehensive surveys of diverse DNA libraries, to identify new sequences encoding the capacity to arrest cell division and induce filamentation.

3.4 METHODS

3.4.1 BACTERIAL STRAINS AND PLASMIDS

E. coli JW0941-1 (*125*) was obtained from the *E. coli* Genetic Stock Center (Yale University, USA). *E. coli* UTI89 (*102*), plasmids pBAD24(*161*) and pLau80(*162*) were kindly provided by J. Moller-Jensen. Strains were routinely cultured at 37°C on solid or in liquid Luria Broth (LB), with 1% (w/v) NaCl.

For the cloning of individual ORFs, DNA fragments containing *fsaA, ybiY, aceE, pdhR, sdaA, lipA, pptE* and *pflC* were amplified from UTI89 gDNA with Phusion DNA polymerase (NEB). DNA oligonucleotide sequences and the integrated restriction sites are shown in Supp. Data Table S1C. The products were purified using the Isolate II PCR and gel-extraction kit (Bioline) before digestion and cloning into pBAD24, utilising the restriction enzymes indicated in Supp. Data Table S1C. The ligation mixtures were transformed into electro-competent *E. coli* BW25113 cells, and clones selected on solid LB agar supplemented with ampicillin (100 μg/ml). Ampicillin resistant colonies were selected, the expected plasmid inserts were confirmed by single-read sequencing (Australian Genome Research Facility).

3.4.2 CONSTRUCTION OF THE SHOTGUN UTI89 GENOMIC DNA EXPRESSION LIBRARY

Purified genomic DNA from *E. coli* UTI89 was partially digested with FatI (NEB), a 4-bp-recognition endonuclease that includes the ATG start codon, which would increase the frequency of in-frame ORF ligation to the expression vector. The gDNA fragments (1-5 kb range) were gel purified and ligated to pBAD24, which had been prepared by digestion with NcoI (NEB), dephosphorylated with Antarctic Phosphatase (NEB) and verified not to self-ligate by transformation of a vector-only ligation mixture. The library ligation mixture was used to transform *E. coli* JW0941-1 (Δ*sulA*) by electroporation, and clones were selected on LB agar supplemented with 100 μg/mL Ampicillin. Colonies arising after ~18

h growth at 37°C were counted and then pooled by suspension and mixing in LB with 16% (v/v) glycerol, and stored in aliquots at -80°C.

3.4.3 CELL SORTING FOR ENRICHMENT OF FILAMENTOUS CLONES FROM THE EXPRESSION LIBRARY

Mid-log cultures of *E. coli* JW0941-1/pBAD24 (vector only), JW0941-1/pLau80 (encoding FtsZ-YFP under P_{BAD} control) and the UTI89 gDNA expression library were diluted to A_{600} = 0.05 and induced with 0.2% (w/v) L-arabinose. Incubation was continued until A_{600} = 0.80. Culture samples were placed on ice and then analysed and sorted with an Aria II flow cytometer (BD Biosciences), as described previously (*88*). Specifically, 500,000 events from the "filamentous" gate were initially collected ("yield sort"), and then the sample was re-sorted to obtain ~20,000 events of higher purity ("purity sort"). Cultures were analysed at an average of 37,000 events per second for the initial yield sort, and at 14,000 events per second for the purity sort. Purity sorted cells and 1 mL of the unsorted sample were used to inoculate 5 mL liquid LB media containing ampicillin and 0.2 % (w/v) D-glucose and grown overnight at 37°C with shaking (~150 rpm). Plasmid DNA from both cultures was extracted and purified using the Isolate II Plasmid Mini Kit (Bioline).

3.4.4 HIGH-THROUGHPUT MULTIPLEX DNA SEQUENCING AND DATA ANALYSIS

Plasmid DNA (5 ng) was prepared for sequencing using the Nextera DNA Library Prep Kit (Illumina), according to the manufacturer's instructions. After the Nextera tagmentation and PCR amplification (tagging and barcoding) procedures, the DNA fragments were quantified on an Agilent Bioanalyzer and then underwent SPRI-select magnetic bead clean-up (Beckman-Coulter) at 0.8-0.5x left and right ratios (to select DNA of ~150-800 bp). The sorted and unsorted libraries were pooled, and quantified with an Agilent Bioanalyzer. The DNA was then diluted to 20 pM, including 5% PhiX genomic adapter-ligated DNA control library (Illumina). Sequencing was performed with an

Illumina V3 150-cycle paired-end (2 x 75 cycles) flow cell and MiSeq instrument, according to manufacturer's instructions.

Sequence reads from the primary FASTQ files were demultiplexed according to their sample barcodes and then underwent quality control analysis and filtering using the FastQC software tool (see *Section 2.6.5*). Sequence read mapping to the UTI89 genome (Genbank NC_007946.1)(*102*) was done using Bowtie2 (*127*). Reads were first strictly mapped (zero mismatches allowed) to the pBAD24 plasmid. Unaligned reads were then strictly mapped to the plasmid pUTI89. Finally, the unaligned reads remaining were mapped strictly to the chromosome of UTI89. Statistics of sequencing reads and the data in Table 3.1 was generated by using the SAMtools software package and Flagstat command (*130*). The frequency of mapped reads and the analysis of data for Figure 3.2 were determined with deepTools2 software (*129*). Artemis (*131*) was used to visualize the alignments of the reference genome with the read data (BAM files) for inspecting hit regions and annotating the output from the peak detection (described below).

Genomic regions causing filamentation were identified by comparing the read depth of the filamentous-sorted sample compared to the read-depth of reference-unsorted sample with MACS (version 2.1.1) peak-detection software (*145*). The alignment data of reads matched to the UTI89 genome were loaded into MACS using the Galaxy interface environment (*163*). The paired-end data for each sample were then combined, and the filamentous-sorted data were input into MACS as the test/tag file and the reference-unsorted data as the control file. Sequence reads of 75 bp and genome length of 5.06 Mbp were specified. The P-value threshold cutoff for peak detection and the bandwidth parameter were the default values of 0.00005 and 300, respectively. The ranges for calculating regional lambda were defined as 1000, 5000 and 10,000 bp in the control file. The value *d*, as determine by MACS, was 200 for both replicate screens.

3.4.5 EXPRESSION OF CLONED ORFs AND ANALYSIS OF CELL PHENOTYPES

Single colonies of BW25113-based strains, containing the cloned ORF's in pBAD24, were used to inoculate liquid LB medium supplemented with ampicillin (100 μg/mL), and incubated overnight at 37°C (~150 rpm). Cultures were then diluted in M9 minimal-media (supplemented with 0.4% (v/v) glycerol and 0.1% (w/v) casamino acids) to A_{600} = 0.02 and induced with the addition of 0.2 % L-arabinose, and repressed with the addition of 0.2% (w/v) glucose, respectively. Cultures at A_{600} = 0.8 were fixed with volume of 2% (w/v) formaldehyde (prepared from a paraformaldehyde stock) and stored at 4°C. Cell volume distributions were determined using a Multisizer 4 Coulter counter/cytometer (Beckman-Coulter), equipped with a 50-micron aperture tube, running with Isoflow electrolyte (Beckman-Coulter) and calibrated with 2-micron latex beads. For microscopy, the cells were stained with FM4-64 (12.5 μg/mL) and Hoechst 33342 (0.5 μg/mL) and were then mounted on a 1% (w/v) agarose gel pad (in phosphate-buffered saline) and then imaged in phase-contrast and epifluorescence with a Zeiss Axioplan 2 microscope, with a 1.4 NA objective lens.

3.4.6 AMPLIFICATION AND SUB-CLONING OF FRAGMENTS FROM IDENTIFIED GENOMIC REGIONS

The two loci, *ybiY-fsaA-moeB-moeA* and *ybeM-tatE-lipA* were first amplified from UTI89 gDNA using Phusion High-Fidelity DNA polymerase (NEB). The amplicons were purified and then partially digested with FatI (NEB). This DNA was ligated to the compatible NcoI site of pBAD24 using T4 DNA ligase. The ligation mixture was transformed by electroporation into *E. coli* BW25113 and clones selected on solid LB agar supplemented with 100 μg/mL ampicillin. Forty-eight ampicillin resistant colonies were picked into 96-well plates containing LB or M9 media with 100 μg/mL ampicillin and 0.2% (w/v) L-arabinose. BW25113 containing pBAD24 only was also used as a control strain. Plates were incubated for 7 h at 37°C (~150 rpm) and cells adhered to 15-well glass slides using

poly-L-lysine and examined at 100x magnification on either a Zeiss Axioplan 2 microscope or an inverted Nikon Ti microscope.

CHAPTER 4

GENOME-WIDE COMPARISON OF *E. COLI* GENES REQUIRED FOR GROWTH IN COMPLEX AND MINIMAL MEDIA

DISCLOSURE

This chapter was written in manuscript format for submission to *Journal of Bacteriology* for publication under the title "Genome-wide comparison of *E. coli* genes required for growth in complex and minimal media" (refer to page *xiv* for list of publications). I certify that I carried out significant work presented in this manuscript.

- Daniel G. Mediati and Iain G. Duggin conceived and designed the experiments.
- Daniel G. Mediati performed the experiments and acquired the data.
- Daniel G. Mediati, Leigh G. Monahan and Iain G. Duggin analysed and interpreted the data.
- Daniel G. Mediati and Iain G. Duggin wrote the paper.
- Daniel G. Mediati, Leigh G. Monahan and Iain G. Duggin revised and proof-read the manuscript.

4.1 INTRODUCTION

Microbial cells have very limited control over their environment. Minimal or variable supply of essential elements and readily assimilable compounds for growth are commonplace challenges in most microbial lifestyles. Microbes therefore must have a flexible and adaptable metabolic capacity. Bacteria such as uropathogenic *Escherichia coli* (UPEC) must rapidly adapt their metabolism according to the nutrients available in highly variable host environments such as the gut and bladder. During a urinary tract infection, UPEC must adapt to utilise cycles of exposure to urine of varying constitution and an intracellular environment in the bladder epithelium (*28*).

Identifying genes encoding the transporters, regulators, and metabolic enzymes that allow microbes to grow under certain conditions provides a way of assessing microbial metabolic and survival capacity in limited or complex environments. These core microbial capacities are essential for infection and virulence. Genome-wide approaches to understanding gene function are also a powerful way of identifying potential targets for antimicrobial therapies. Several previous studies have assessed *E. coli* gene essentiality under nutrient-rich laboratory conditions (*164-166*). However, gene essentiality information under more nutrient-restricted conditions is limited, despite its likely greater relevance to survival and infection.

The genetic requirements of commensal *E. coli* K-12 for growth in rich and minimal media have been assessed by analysis of the Keio collection of knock-out strains (*125, 166*). More recently, multiplexed analyses of pooled mutant strain libraries, such as transposon-directed insertion-site sequencing (TraDIS), have opened up new possibilities for detecting gene requirements on a genome-wide scale under many different conditions of growth (*123, 167*). TraDIS and related approaches allow the simultaneous assessment of the relative abundance of individual transposon-insertion mutants from a culture containing a large number of different mutants. The individual mutants are detected *en mass*e by PCR amplification of the specific transposon-genomic junctions from the library, followed by high-throughput sequencing of the amplified DNA and mapping the sequence reads to the reference genome. The read count at each site can be a relative measurement of the

abundance of each mutant. Essential genes may be identified by the absence of transposon insertions within those genes in a high-density transposon-insertion library. In addition, conditional requirements for particular genes may be identified by comparing the relative abundance of mutants between two different growth conditions. An attractive goal of such comparisons is to determine the genetic requirements of bacterial pathogens at various stages of infection. However, significant limitations exist in the number of mutants that can be simultaneously analyzed with the limited scale of culture that is typically feasible with infection models; with a large mutant library, the experiment may not be possible due to insufficient representation of each mutant in the library during an infection, a problem termed "bottlenecking" (*168*).

To begin addressing these challenges and provide new insight into the genetic and metabolic requirements for growth of an important human pathogen, here we constructed a transposon-insertion mutant library in the model uropathogenic *E. coli* (UPEC) strain, UTI89 (*29, 102*), at a moderate scale (~20,000 mutants) that represents insertions into the majority of genes and is well suited to infection models of UTI—such that sufficient data can be collected for assessment of the fitness of the individual mutants (*i.e.* avoiding bottlenecking). We combined this library with a modified TraDIS protocol that simplifies sample preparation and requires much less total DNA than the original method—also potentially more amenable for infection studies. TraDIS analysis of genes required for growth in a defined M9-glycerol minimal medium compared to a rich LB medium identified 60 mutants with a significant fitness defect and reduced capacity to survive in the M9, the majority of which encode gluconeogenic and amino acid catabolism proteins. Comparison of the results with those from a previous survey of the *E. coli* K-12 Keio collection of knock-out mutants (*166*) highlighted important differences and several apparent discrepancies in the metabolic requirements and adaptations between uropathogenic and commensal *E. coli*. Several uncharacterised and UPEC-specific genes were identified that likely underlie metabolic capacities of UPEC strains during infection. Two of these, *neuC* and *hisF*, were verified as important during UPEC infection of bladder epithelial cells *in vitro*.

4.2 RESULTS

4.2.1 DESIGN OF A MODIFIED TraDIS PROTOCOL

We modified the original TraDIS protocol to significantly reduce the amount of input DNA required (an expected advantage with limited scale cultures, such as with infection models) and to eliminate the need for custom sequencing primers and dark-cycles during sequencing, which are non-standard options that increase the complexity of performing sequencing runs and may reduce sequencing method reliability. This was achieved by firstly swapping the DNA shearing, end-repair and linker-ligation steps of the original TraDIS method with the "tagmentation" procedure for DNA fragment library preparation (Supp. Data Fig. S4A). Tagmentation involves a one-step transposase-mediated random cutting of DNA and ligation of adapter oligonucleotides to the generated ends, which is required for the downstream amplification and high-throughput sequencing steps. Next, PCR amplification of the transposon-genomic DNA junctions in the library was achieved by combining a transposon-specific outward-directed primer with a primer that anneals to one of the end-adapter sequences (Supp. Data Fig. S4B). Both primers include 5' tag sequences that make them suitable for downstream sequencing, as described below. The transposon-specific primer incorporates (from 5' to 3'): (i) the "P5" or "P7" flow-cell annealing sequence (Illumina), (ii) the standard "read 1" or "read 2" sequencing primer binding site (Illumina), and (iii) a ~20 bp transposon specific annealing sequence, designed to amplify ~40–50 bp of the corresponding transposon end followed by its adjacent genomic sequence. Two such transposon-specific primers were designed—one for each end of the transposon; one end was represented by the P7/read 1 tag, and the other by the P5/read 2 tag (Supp. Data Fig. S4B). The second primer in each reaction follows the more standard design for Illumina sequencing, containing (from 5' to 3'): (i) the alternate P5 or P7 adapter sequence for binding to the Illumina sequencing flow cell, (ii) a sample-specific unique barcode sequence for later sample identification (de-multiplexing), and (iii) a sequence that anneals to the tagmentation adapter on the opposing end of the fragment, which then serves as the read primer binding site during sequencing. PCR reactions carried out with only the one adapter primer did not generate significant amplified product (data

not shown). After PCR, the purified products could then be subjected to standard paired-end sequencing (Illumina) (Supp. Data Fig. S4C). The dual-end sequencing design above ensured that only ~50% of reads contained a common transposon end sequence during the early cycles of either read 1 or read 2. The remaining reads comprised a wide range of sequences from the background genome, thereby eliminating the technical problems associated with sequencing low diversity samples on Illumina instruments.

4.2.2 CONSTRUCTION OF A TRANSPOSON MUTANT LIBRARY IN UTI89

To perform transposon mutagenesis with cystitis isolate UTI89, we used a mini-Tn5 synthetic transposon, Ez-Tn5<Kan-2>, carrying a kanamycin resistance cassette between inverted repeat (IR) end regions. Transposition was carried out via electroporation of *E. coli* with transposasomes (*i.e.* transposon-transposase complexes), and then ~97,500 Km^R colonies (SEM = 882, n = 3) that arose on agar plates were pooled and homogenously resuspended in liquid medium, and then stored as frozen aliquots. One aliquot was thawed for viable cell assay by plate counting, giving 1.18×10^{10} cfu/mL.

TraDIS sequencing reads, obtained from total library DNA and the method outlined above, were quality filtered and then mapped to the chromosome and pUTI89 plasmid. The number of sequence reads successfully mapped to UTI89 and the number of unique insertion sites these reads correspondingly mapped to on the genome were determined using the SMALT read mapper software (*128*). An average of 18,756 unique insertion sites within the main chromosome, and 464 unique insertion sites within the plasmid (pUTI89) were identified in the total library samples (Supp. Data Table S2A). This equates to an average insertion site every 270 base pairs within the chromosome. Strong correlation was also evident ($R^2 = 0.878$) between the insertion site counts on the Tni5 and Tni7 ends of the transposon, demonstrating both ends are producing consistent reads per gene (Supp. Data Fig. S5). Approximately 22 % of total gene open reading frames (ORFs) within the chromosome (1,184 out of 5,363 genes) did not contain an insertion site (see Supp. Data Table S2B-C for raw read count per 150 bp and Supp. Data Table S2D for the zero-insert gene list). Many of these genes are likely to be essential for viability. However, since it has

been predicted that only 8-10 % of the *E. coli* genome contains genes individually essential for viability in LB medium (*125, 169*), a significant proportion of the genes not represented by transposon-insertion in the present library are likely non-essential, and were not present in the library by chance due to the moderate library size (transposon density). Consistent with this, the average size of the ORFs containing no insertions was 569 bp, compared to the average for the whole genome of 901 bp (*102*). Nonetheless, the insertion sites appear otherwise quite consistently distributed throughout the chromosome and plasmid pUTI89 (Fig. 4.1A).

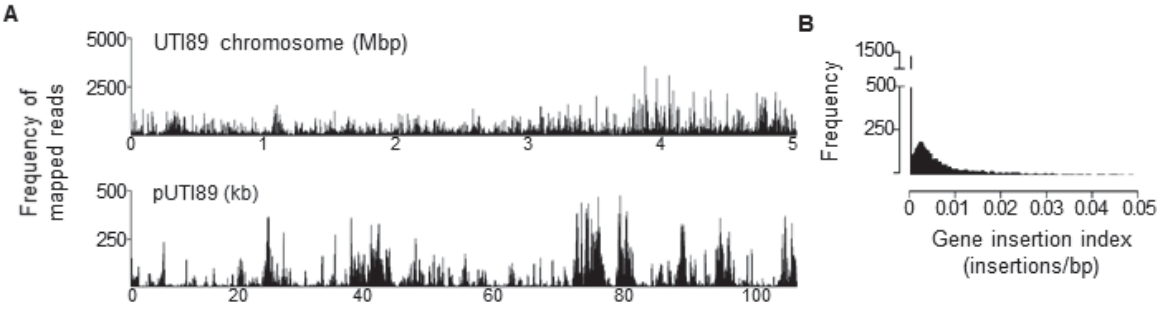


Figure 4.1 Genomic overview of the UTI89 transposon mutant library.

(A) The distribution of mapped-read counts along the ~5.06 Mbp UTI89 chromosome and ~110 kb native plasmid pUTI89, with a bin size of 150 bp for the base library samples (Library Tni5 and Tni7 in Supp. Data Table S2). (B) Frequency distribution of the insertion index (the number of insertions per gene divided by the gene length in bp) for both UTI89 chromosome and pUTI89. Genes with insertion index of 0 (leftmost peak) have no transposon insertions. These genes are either essential and/or not present by chance in our library. The broad peak of genes with positive insertion indexes represents genes that can tolerate transposon insertions and are represented in this library.

To further characterise the library, the frequency distribution of the gene insertion index, *i.e.* the number of transposon insertions within a gene divided by that gene's length in bp, for the whole UTI89 genome was determined (Fig. 4.1B). This normalisation of insertion frequency based on gene length has been used previously as a measure of gene essentiality (*123, 124, 169, 170*). Genes with an insertion index of 0 (leftmost spike, ~1,200 genes) indicate those genes that have no transposon insertions (Supp. Data Table S2D). These genes are either essential and/or not present by chance within our library. The broader right-hand peak represents those genes that can tolerate transposon insertions in at least one position within the ORF and are present within our library. This population of genes appeared as a single symmetrical peak, consistent with a near-random insertion frequency in these genes. The observation that this peak is not completely distinct from the zero-read genes suggests that the library contains a minority of non-essential genes that are not represented by transposon insertions, as noted above. For a very high density library, the right-hand peak of insertion-containing genes is clearly distinct from the leftmost spike of genes without insertions (*123*) whereas in moderate density libraries (*170*) they are not fully distinct.

4.2.3 GENES ADVANTAGEOUS AND DISADVANTAGEOUS FOR GROWTH

Gene essentiality is a dynamic and context-dependent situation. In many cases a gene inactivation may reduce growth rate but not be essential for growth in culture (*123*). Such mutants would show a reduced frequency over time in relation to others with no or lesser growth rate defects. The opposite may also be true for certain genes, where there exists the possibility that mutants may cause an increase in the growth rate in culture. Such genes would be expected to encode proteins whose expression or activity is burdensome for growth in the artificial environment of culture that differs from natural environments where such genes are expected to confer their specific fitness advantages. We sought to characterise the change in representation of mutants over time by passaging the *E. coli* UTI89 Tn-library five consecutive times (5 days) by 1/400 dilutions in Luria-Bertani (LB) media at 37°C. Bacteria were harvested after passage one, three and five and then total purified DNA was subjected to our TraDIS sequencing protocol. The frequency and

distribution of mapped-reads throughout the UTI89 chromosome was plotted in 150 bp bins along the genome. By comparing the mutant composition over the course of the five passages, it became clear that some genes contained progressively relatively fewer sequence reads, whereas others appeared to increase relative to the vast majority (Fig. 4.2A).

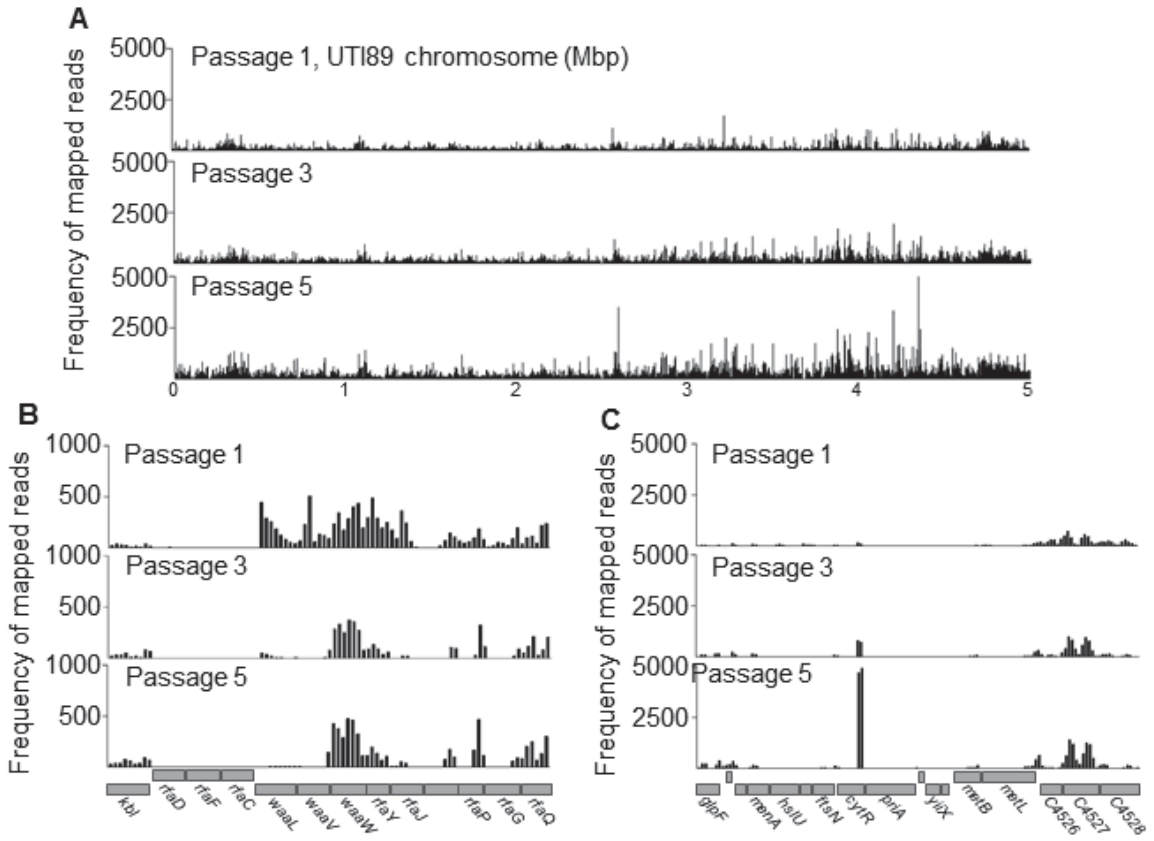


Figure 4.2. Mutant compositional changes during library growth in LB.

(A) Frequency and distribution of the number of mapped-reads in 150 bp windows along the ~5.06 Mbp UTI89 chromosome following growth through one, three, and five 24 h passages (1/400 dilutions) in LB medium. (B-C) The frequency and distribution of mapped-reads corresponding to selected regions of the chromosome within a window of 150 bp through day one, three and five. The position of the annotated genes relative to the mapped-reads is aligned with the x-axis. (B) Read-count distribution encompassing the region 4,053,040 – 4,066,540 bp, displaying the waaLV and rfaYJ genes, which demonstrate a decrease in the read depth following growth to passage 5. (C) Read-count distribution encompassing the region of 4,394,185 – 4,413,535 bp demonstrating the cytR gene, which displays an increase in the read depth following growth to passage 5.

Insertion site read counts in individual genes were then compared between the first and fifth day samples using edgeR (*128, 170*). This analysis identified 196 genes that displayed reduced abundance of their corresponding mutants in the library at day five (*i.e.* $\log_2$FC ≤ -2 and p-value ≤ 0.05) and are likely to show reduced growth rates (Supp. Data Table S2E). As an example, Fig. 4.2B shows the frequency and distribution of mapped-reads in the region 4,053,040 – 4,066,540 bp, displaying the *waaLV* and *rfaYJ* genes that were identified as disadvantageous to growth during passaging. We also identified numerous fimbriae-associated proteins, including FimB, part of the Type 1 pili; P fimbriae adhesins, *papBEF*; S fimbriae adhesins, *sfaDEHS* and the putative *yad*-type adhesin operon, *yadKLMN*.

Sixteen mutants displayed a significant increase in relative abundance by day five ($\log_2$FC ≥ 2, p-value ≤ 0.05) and these mutants may therefore confer an increased growth rate compared to the majority under these conditions (Supp. Data Table S2E). As an example, Fig. 4.2C shows the frequency and distribution of mapped-reads in the region of 4,394,185 – 4,413,535 bp containing the *cytR* (the cytidine regulator) gene, which displays an increase in the read frequency following growth to passage five. In each of the gene distribution plots (Fig. 2), passage three displays an intermediate in the read count between the first and fifth passages. Finally, those genes that could not be assigned as significantly advantageous or disadvantageous ($\log_2$FC between -2 and 2) are also listed in Supp. Data Table S2E.

4.2.4 IDENTIFICATION OF GENES REQUIRED FOR GROWTH IN M9-GLYCEROL MINIMAL MEDIUM

To better understand UPEC metabolic adaptation and responses during nutrient-limiting conditions, we next sought to identify genes required for UTI89 survival and growth in M9-glycerol minimal medium, compared with LB medium. We also wished to assess whether the size of the mutant library described here (~20,000), when combined with our modified TraDIS protocol, was suitable in experiments where the practical limits of the scale of cultures, such as in infection models, restrict the number of mutants that can be simultaneously and robustly analysed (bottlenecking). We therefore bottlenecked the

cultures by inoculating with 1.5 x 10^5 viable bacteria, which is a realistic input size in laboratory scale-restricted infection models. This represents approximately 8-fold more viable bacteria than the number of mutants in the library, *i.e.* each mutant would be represented in the starting culture by an average of 8 viable bacteria, and follow a Poisson distribution around this. Assuming equal representation of mutants in the initial library, with this probability distribution we can estimate that this degree of bottlenecking would cause ~7 mutants (0.00035 %) from a library of this size (~20,000) to be randomly excluded from any particular culture. On the other hand, a typical high density transposon library of 10^6 mutants used for bulk-culture gene requirement screens (*123, 171*) would clearly not be usable with 1.5 x 10^5 viable starting cells, since there would be an ~86 % chance of not selecting any bacteria from a particular mutant, and others would be represented by too few bacteria.

To identify mutant genes that are important in M9-glycerol conditions, we prepared duplicate cultures for both M9 and LB media, and aimed to identify the mutant genes that were robustly detected as important in M9 media but not LB in both duplicates. The four cultures were each grown for three consecutive 24 h passages at 37°C (*i.e.* the initial culture of 1.5 x 10^5 bacteria, then two further 1/10 dilution passages). After TraDIS sequencing of total DNA, an average of 1.7 million reads were obtained for each sample, with between 350, 000 – 1.4 million reads identified containing the transposon-specific end sequence (Supp. Data Table S2A). Of those reads tagged with the transposon, between 76.3 – 89.4 % were reliably mapped to the chromosome of UTI89, and 1.2 – 8.8 % reliably mapped to pUTI89 (Supp. Data Table S2A).

Read-count comparisons of the M9 and LB samples were performed with BioTraDIS; the complete list of genes and their P-values that were output can be found in Supp. Data Table S2. All those genes with a P-value of ≤ 0.05 in at least one of the replicates are plotted by functional class in Fig. 4.3. The prominent functional classes include genes with known involvements in metabolism, such as amino acid, carbohydrate, inorganic ion transport and metabolism, and energy production and conversion. Of these genes, 488 met the significance criteria (P-value ≤ 0.05) in only one of the replicates, and therefore may be of

minor consequence for growth in M9-glycerol. Seventy one genes that met the significance criteria (P-value ≤ 0.05) were also found in common between replicates.

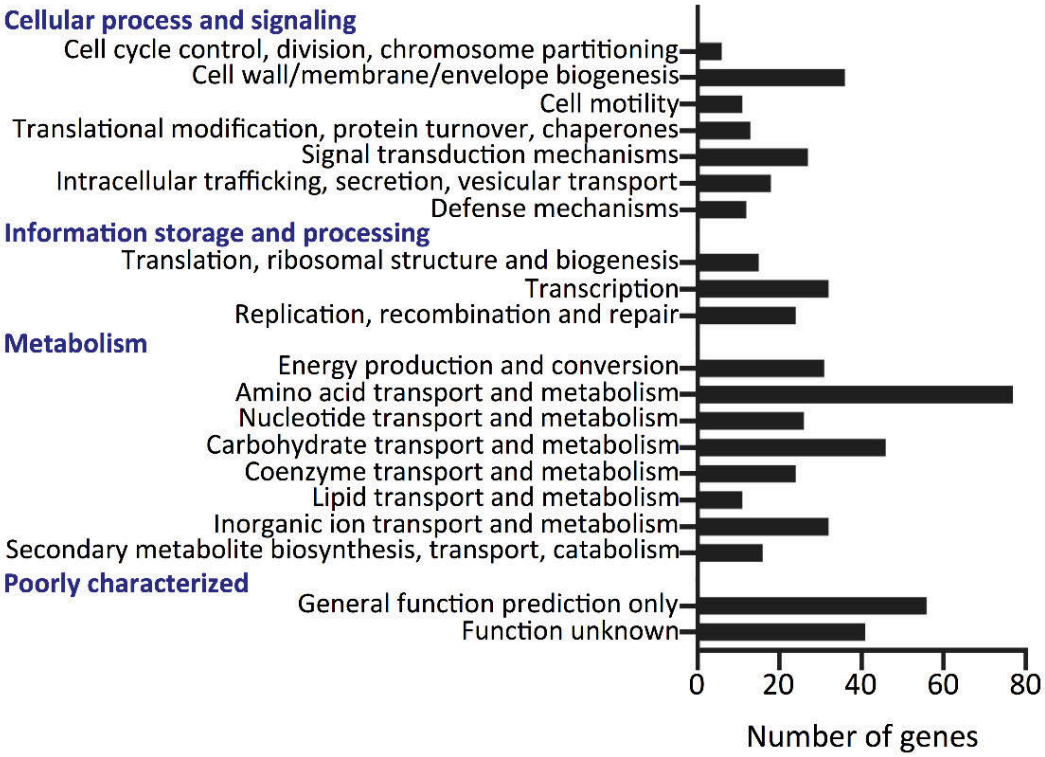


Figure 4.3. Number of essential genes according to COG functional category.

Number of genes identified by TraDIS with a P-value ≤ 0.05 in either duplicate experiment. The full dataset can be found in Supp. Data Table S2F.

By applying a more stringent selection criteria (P-value ≤ 0.05 and $\log_2$FC < -1, in both duplicates), 60 mutant genes showed a reproducibly reduced fitness to grow and survive within M9-glycerol compared to LB (Table 4.1 and Fig. 4.4A). These genes therefore are taken as important for growth of wild-type *E. coli* UTI89 in M9-glycerol. The majority of these have roles in metabolism and metabolic regulation in *E. coli*, such as *glpDK*, *purEHL*, *pyrDE* and *cysCK* (Table 4.1). Interestingly, a proportion of uncharacterized genes and those with only hypothetical functions, for example *YjiFY*, UTI89_C0737, *YgeZ* and UTI89_C29889, were amongst those identified.

To further highlight the data pertaining to this 60-gene list, three genomic regions are shown in Fig. 4.4 that showed significant differences between LB and M9. Fig. 4.4B**,** showing region 2,225,678 – 2,239,628 bp, displays genes of the *his* (histidine biosynthesis) operon, and demonstrates a reduced read depth in the M9 condition when compared to LB (particularly *hisG*, *hisH*, *hisF* and *hisI*). Another region showing reduced read depth in the M9 condition was 4,431,758 – 4,445,258 bp (Fig. 4.4C). This contains the *argCBH* gene operon and downstream *argE* gene involved in arginine biosynthesis, which is controlled by the sigma factor *rpoS* during nutrient-limiting conditions. Thirdly, the region containing the *neuC* gene was also apparently important for growth in M9-glycerol medium (region 3,288,222 – 3,302,042 bp) (Fig. 4.4D). NeuC is required for sialic acid biosynthesis and production of the K1 capsule of pathogenic strains of *E. coli*.

Table 4.1 UTI89 genes required for growth in M9-glycerol – high stringent hits

Gene ID	Gene Name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C0551	*purE*	phosphoribosylaminoimidazole carboxylase	-8.21	4.37E-04	2.12E-02
UTI89_C2298	*hisF*	imidazole glycerol phosphate synthase subunit hisF	-7.29	4.26E-04	2.06E-02
UTI89_C1010	*pyrD*	dihydroorotate dehydrogenase	-6.72	1.74E-02	1.63E-01
UTI89_C3926	*glpD*	aerobic glycerol-3-phosphate dehydrogenase	-6.57	5.53E-04	1.91E-02
UTI89_C0736	*tolB*	Involved in the tonB-independent uptake of colicins	-6.00	9.30E-03	1.01E-01
UTI89_C3111	*rpoS*	sigma factor RpoS	-5.95	3.58E-03	6.66E-02
UTI89_C3110	UTI89_C3110	hypothetical protein	-5.94	3.66E-03	6.77E-02
UTI89_C0737	UTI89_C0737	hypothetical protein	-5.90	1.93E-02	1.12E-01
UTI89_C1127	UTI89_C1127	putative transposase	-5.83	6.98E-04	2.53E-02
UTI89_C3887	*aroB*	3-dehydroquinate synthase	-5.66	4.12E-03	7.37E-02
UTI89_C2448	*yeiP*	putative elongation factor YeiP	-5.64	1.92E-03	3.81E-02
UTI89_C3726	*aroE*	shikimate 5-dehydrogenase	-5.62	1.20E-02	9.33E-02
UTI89_C3429	*yghB*	hypothetical protein	-5.55	7.78E-03	8.00E-02
UTI89_C0733	*tolQ*	TolQ protein	-5.49	1.51E-02	1.06E-01
UTI89_C4186	*pyrE*	orotate phosphoribosyltransferase	-5.44	1.31E-02	1.50E-01
UTI89_C3121	*cysC*	adenylylsulfate kinase	-5.38	7.30E-03	7.32E-02
UTI89_C2889	UTI89_C2889	hypothetical protein	-5.19	1.11E-02	1.05E-01
UTI89_C4551	*argH*	argininosuccinate lyase	-5.01	1.21E-04	6.01E-03
UTI89_C4510	*glpK*	glycerol kinase	-4.98	6.55E-09	1.02E-06
UTI89_C3219	*argA*	amino-acid acetyltransferase	-4.81	1.24E-05	7.18E-04
UTI89_C3814	*purH*	bifunctional AICAR formyltransferase	-4.80	3.29E-03	6.53E-02
UTI89_C1621	UTI89_C1621	hypothetical protein	-4.49	4.81E-03	8.12E-02
UTI89_C1620	UTI89_C1620	hypothetical protein	-4.21	9.48E-03	1.26E-01
UTI89_C0037	*carB*	Carbamoyl-phosphate synthase large chain	-4.20	3.86E-05	4.38E-03
UTI89_C2877	*purL*	phosphoribosylformyl-glycineamide synthetase	-4.19	2.72E-04	1.66E-02
UTI89_C0465	*clpP*	ATP-dependent proteolytic subunit of clpA-ClpP	-4.14	2.06E-02	1.96E-01
UTI89_C5159	*serB*	phosphoserine phosphatase	-4.07	1.87E-02	1.65E-01
UTI89_C2296	*hisH*	imidazole glycerol phosphate synthase holoenzyme	-3.94	2.23E-03	5.04E-02
UTI89_C0049	*fixC*	FixC protein	-3.54	8.81E-03	9.40E-02
UTI89_C3096	*ygbA*	conserved hypothetical protein	-3.48	2.45E-02	1.85E-01
UTI89_C2933	*tyrA*	bifunctional chorismate mutase/prephenate dehydratase	-3.29	5.97E-03	4.79E-02
UTI89_C1105	UTI89_C1105	hypothetical protein	-3.28	2.63E-02	2.01E-01
UTI89_C0231	*gloB*	probable hydroxyacylglutathione hydrolase	-3.27	2.67E-02	2.25E-01
UTI89_C2747	*cysK*	O-acetylserine sulfhydrolase A	-3.23	1.62E-02	1.58E-01
UTI89_C5098	UTI89_C5098	putative antirepressor	-2.95	3.11E-03	5.96E-02
UTI89_C4549	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase	-2.93	1.48E-02	1.58E-01
UTI89_C0888	*clpA*	ATP-dependent clp protease ATP-binding subunit clpA	-2.83	3.14E-02	2.16E-01
UTI89_C3077	*ascF*	PTS enzyme II-asc	-2.82	1.46E-02	1.58E-01
UTI89_C4548	*argE*	acetylornithine deacetylase	-2.32	2.91E-03	5.37E-02
UTI89_C3370	*neuC*	polysialic acid biosynthesis protein P7	-2.20	7.64E-04	2.61E-02
UTI89_C0118	*ampD*	AmpD protein	-2.17	4.27E-03	7.63E-02

Gene ID	Gene Name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C4223	*uhpC*	regulatory protein UhpC	-2.16	3.17E-02	2.51E-01
UTI89_C3222	*ptrA*	protease III precursor	-1.99	8.55E-04	2.84E-02
UTI89_C4307	*rbsK*	ribokinase	-1.98	1.89E-02	1.73E-01
UTI89_C4799	*yjfY*	hypothetical protein	-1.97	1.65E-02	1.70E-01
UTI89_C3258	*ygeZ*	hypothetical hydrolase YgeZ	-1.91	1.70E-02	1.61E-01
UTI89_C2292	*hisG*	ATP phosphoribosyltransferase	-1.90	3.42E-02	2.46E-01
UTI89_C3938	*glgA*	glycogen synthase	-1.81	5.88E-03	9.81E-02
UTI89_C4127	*sgbE*	probable sugar isomerase SgbE	-1.78	1.30E-03	3.71E-02
UTI89_C3736	*yhdL*	putative regulator	-1.76	2.03E-02	1.79E-01
UTI89_C3307	*yggB*	hypothetical protein YggB	-1.69	2.36E-04	1.57E-02
UTI89_C4128	UTI89_C4128	hypothetical protein	-1.67	1.42E-03	3.88E-02
UTI89_C0372	*lacI*	lac operon repressor	-1.66	1.34E-03	3.72E-02
UTI89_C0080	*leuB*	3-isopropylmalate dehydrogenase	-1.63	2.14E-02	1.91E-01
UTI89_C1497	UTI89_C1497	DNA packaging protein of prophage	-1.53	2.03E-02	1.93E-01
UTI89_C4190	*yicG*	hypothetical protein	-1.47	1.54E-02	1.55E-01
UTI89_C3267	*ygfO*	putative transport protein	-1.35	9.56E-03	1.22E-01
UTI89_C4835	*yjfF*	hypothetical ABC transporter permease protein yjfF	-1.27	2.95E-02	2.33E-01
UTI89_C0032	*yaaF*	hypothetical protein YaaF	-1.26	4.24E-02	2.92E-01
UTI89_C4242	*dgoT*	D-galactonate transporter	-1.15	5.59E-03	9.38E-02

**Hypothetical or proteins with no characterised homologs are listed with the gene ID.*

#Q-value as determined by TraDIS analysis. The Q-value is the P-value that has been adjusted for the false discovery rate (FDR) for each gene hit.

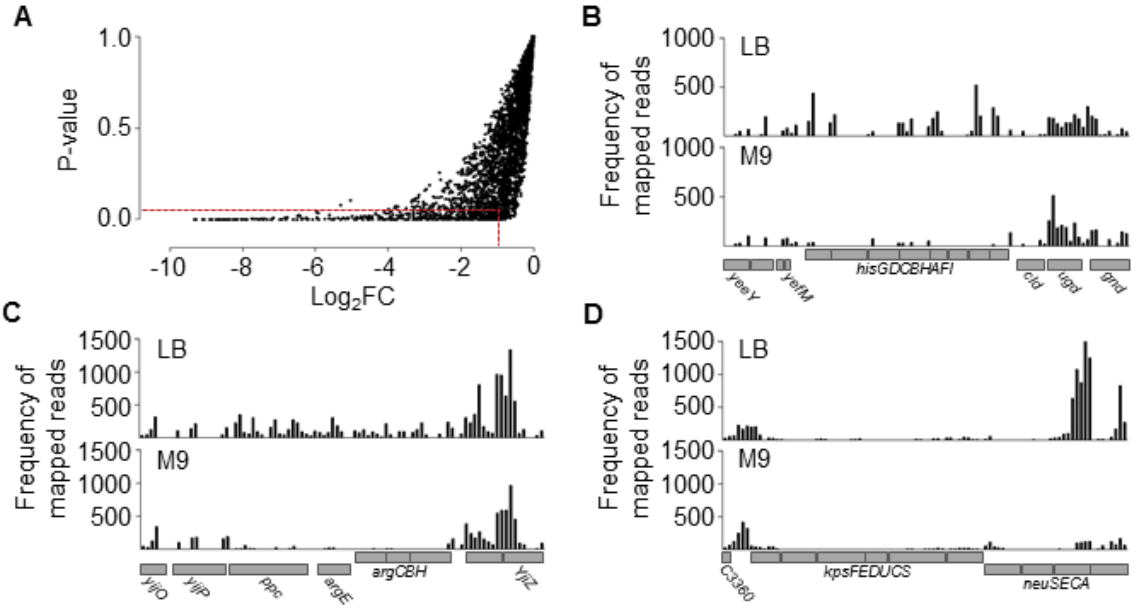


Figure 4.4 Identification of genes required for growth of UPEC in minimal media by TraDIS.

(A) The distribution of genes identified by TraDIS, with each gene's P-value plotted against its corresponding log_2FC value. The P-value score of ≤ 0.05 and Log_2FC value of < -1 indicates the selected high stringent threshold criteria (red dotted line). Those genes that were found in common between replicates within this threshold are displayed in Table 4.1. (B-D) The frequency and distribution of mapped-reads corresponding to selected regions of the UTI89 chromosome for both LB (top panel) and M9 (bottom panel) conditions, within a bin of 150 bp. The position of the annotated genes relative to the mapped-reads is indicated below the distribution plot. (B) Encompasses region 2,225,678 – 2,239,628 bp, displaying genes of the his operon, such as hisG, hisH, hisF and hisI which were identified by TraDIS and are evident with a reduced read depth in the M9 condition. (C) Similar to (B), showing the argECBH genes within the region of 4,431,758 – 4,445,258 bp identified by TraDIS. (D) Similar to (C-B), showing the neuCA genes within the region of 3,288,114 – 3,301,914 bp identified by TraDIS.

4.2.5 VERIFICATION OF GENES REQUIRED FOR GROWTH IN M9

To investigate whether particular genes identified from TraDIS analysis were responsible for the observed fitness reduction in M9-glycerol medium, we selected five candidate gene ORFs of differing statistical strength (Supp. Data Table S2F). These gene ORFs, *hisF, neuC, yggB, pdhR* and *ykgC*, were then individually deleted from UTI89 using the λ Red recombinase system. Three of the gene deletions, *hisF*, *neuC* and *pdhR*, resulted in a significant growth defect when grown in M9 media compared to LB media, when compared to WT UTI89 (Fig. 4.5). One of the gene deletions, *yggB*, resulted in a modest fitness defect in M9 media compared to WT UTI89, while the respective LB cultures were indistinguishable (Fig. 4.5). Consistent with this, *yggB* showed a moderate reduction in frequency in the TraDIS experiments (average Log_2FC of -1.69, P-value = 2.9×10^{-4}). This demonstrates the sensitivity of the TraDIS analysis to moderate differences in growth that are highly-specific to the conditions under investigation. The *ykgC* gene deletion resulted in no apparent difference in growth in M9 when compared to growth in LB and WT (Fig. 4.5). TraDIS data for this gene was output by BioTraDIS software analysis, but the statistical strength was considered insufficient between replicates (an average P-value > 0.3 and log_2FC -0.38), confirming that this gene is not important for growth in M9-glycerol.

To quantify the role of various culture growth parameters in reducing the proliferation of these mutants, the growth curves were analyzed to obtain the log-phase growth rates and maximal optical densities for these strains. The average growth rates (μ) of each mutant and WT were determined during exponential growth in M9 and LB according to growth curve modeling methods of Baranyi and Roberts(*172*), and the values are displayed as percentages relative to WT (Supp. Data Table S2G). All four mutants displaying clear growth defects, *hisF*, *neuC*, *pdhR* and *yggB* displayed significantly reduced growth rates when compared to WT UTI89 grown in M9 (Fig. 4.5G). Similarly, the average maximum optical density (OD_{600nm}) of each mutant and WT were determined and displayed as percentages relative to WT. All four mutants, *hisF*, *neuC*, *pdhR* and *yggB* also displayed significantly reduced maximum OD when compared to WT grown in M9 (Fig. 4.5H).

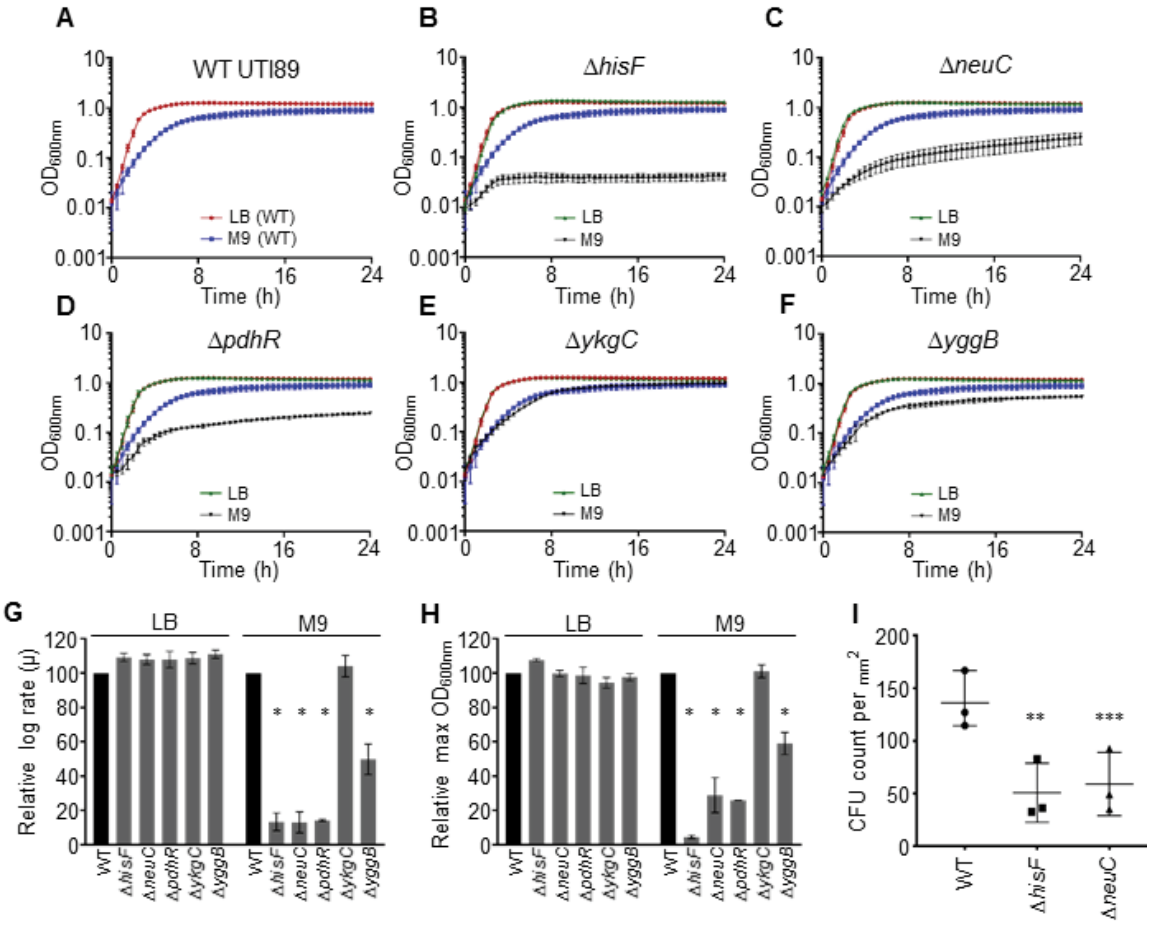


Figure 4.5. Verification and characterisation of deletion mutants in UTI89

*(A-F) Growth rate absorbance measurements (600_{nm}) of (A) UTI89 WT and (B-F) various mutants identified from TraDIS over a period of 24 hr grown in LB and M9. The WT UTI89 growth curves in LB (red) and M9 (blue) from (A) are also shown throughout B-F for comparison. In B-F, each mutant grown in LB (green) and M9 (black). (B) ΔhisF, (C) ΔneuC, (D) ΔpdhR, (E) ΔykgC and (F) ΔyggB. (G) Histogram plot of the proliferative log rate (μ) determined for the mutant growth curves relative to WT UTI89 (normalised to 100) grown in both LB (left-hand side panel) and M9 (right-hand side panel). (H) Similar to (G), displaying the maximum absorbance of each mutant relative to WT UTI89 (normalized to 100) (*P-value ≤ 0.05). (I) Human bladder epithelial cell model of UTI using a gentamicin protection assay. Human bladder cell line PD07i was grown to confluent on 6-well plate. UTI89 WT, hisF and neuC mutants were then inoculated, grown for 2 hours and gentamicin added and left to grow stationary for 20 h at 37°C. Bladder cells were then lysed and bacteria plated onto selective LB media, where appropriate. *denotes the significance level between two biological replicates (P-value ≤ 0.05). ** P-value = 0.0131. *** P-value = 0.0276.*

We considered it likely that UTI89 may rely on *de novo* biosynthesis of certain metabolic pathways during intracellular colonisation of human bladder cells, in some cases similar to minimal media conditions. Therefore, the effect of *neuC* and *hisF* mutations on the ability of UPEC to infect and colonise human bladder epithelial cells using a gentamicin protection assay was examined. To test the requirement of *hisF* and *neuC* in UTI89, the corresponding knock-out mutants were first grown statically overnight in liquid LB to induce fimbriae expression. The bacterial cultures were then used to inoculate confluent bladder epithelial cells grown to give a multiplicity of infection (MOI) of 100 and gently spun down to allow bacterial attachment and invasion. After 2 h of growth at 37°C with 5 % CO_2, the supernatant was then discarded and the bladder cells resuspended in gentamicin supplemented media. Cultures were then incubated statically at 37°C with 5% CO_2 for a further 22 h (total of 24 h post-infection). Following this, cultures were washed gently remove the gentamicin solution and then a lysis solution was added to break open the bladder cells and separate the intracellular bacteria for viability counting. The loss of *neuC* or *hisF* resulted in a significant reduction of UPEC colony forming units (CFU) 24 h post-infection when compared to wild-type, as shown in Fig. 4.5I, suggesting that these genes and the functions they encode, offer a significant fitness advantage to UPEC during infection.

4.2.6 COMPARISON OF A SUBSET OF GENES TO E. COLI K-12

Systematic testing of gene requirements for growth in M9-glycerol was performed previously in *E. coli* K-12 (*166*). We wanted to determine if our TraDIS results were congruent with the findings from *E. coli* K-12, and in doing so further characterize the metabolic profile and requirements of UTI89 during minimal media growth. To investigate whether particular genes verified from the UTI89 TraDIS analysis also gave rise to fitness defect in BW25113 grown in M9 media, we selected seven candidate gene ORFs: *argC, bioF, pdhR, sgbE, yggB, ykgC and hisF*, of which four were in common to the UTI89 verification mutants (*hisF*, *yggB*, *pdhR* and *ykgC*; *neuC* is not conserved in *E. coli* K-12 strains). Mutants for the remaining three, *argC*, *bioF*, and *sgbE*, were selected on the basis of their differing statistical strength output by TraDIS analysis. All cultures were grown

for 24 h in both M9 and LB media and their optical density (OD_{600nm}) measured at 30 min intervals in a plate reader. Four of the gene deletions, *argC*, *bioF, pdhR* and *hisF*, resulted in a significant growth defect in M9-glycerol media compared to WT UTI89 (Fig. 4.6). The *argC* and *hisF* mutants failed to show any detectable growth in M9 media (Fig. 4.6 and Supp. Data Table S2G). The *yggB* gene deletion resulted in a mild relative fitness defect when grown in M9 media (Fig. 4.6F). It was also observed that following the 24 h time period, the optical density of *pdhR*, although significantly reduced during this period, did regain to level par with WT (data not shown). Finally, both the *ykgC* and *sgbE* gene deletions resulted in no apparent growth defect in M9-glycerol medium (Fig. 4.6 and Supp. Data Table S2G), and the *sgbE* mutation appeared to offer a minor growth-rate enhancement compared to WT UTI89. This highlights several novel findings in the metabolic requirements of *E. coli* K-12, which have not been reported in previous gene essentiality studies.

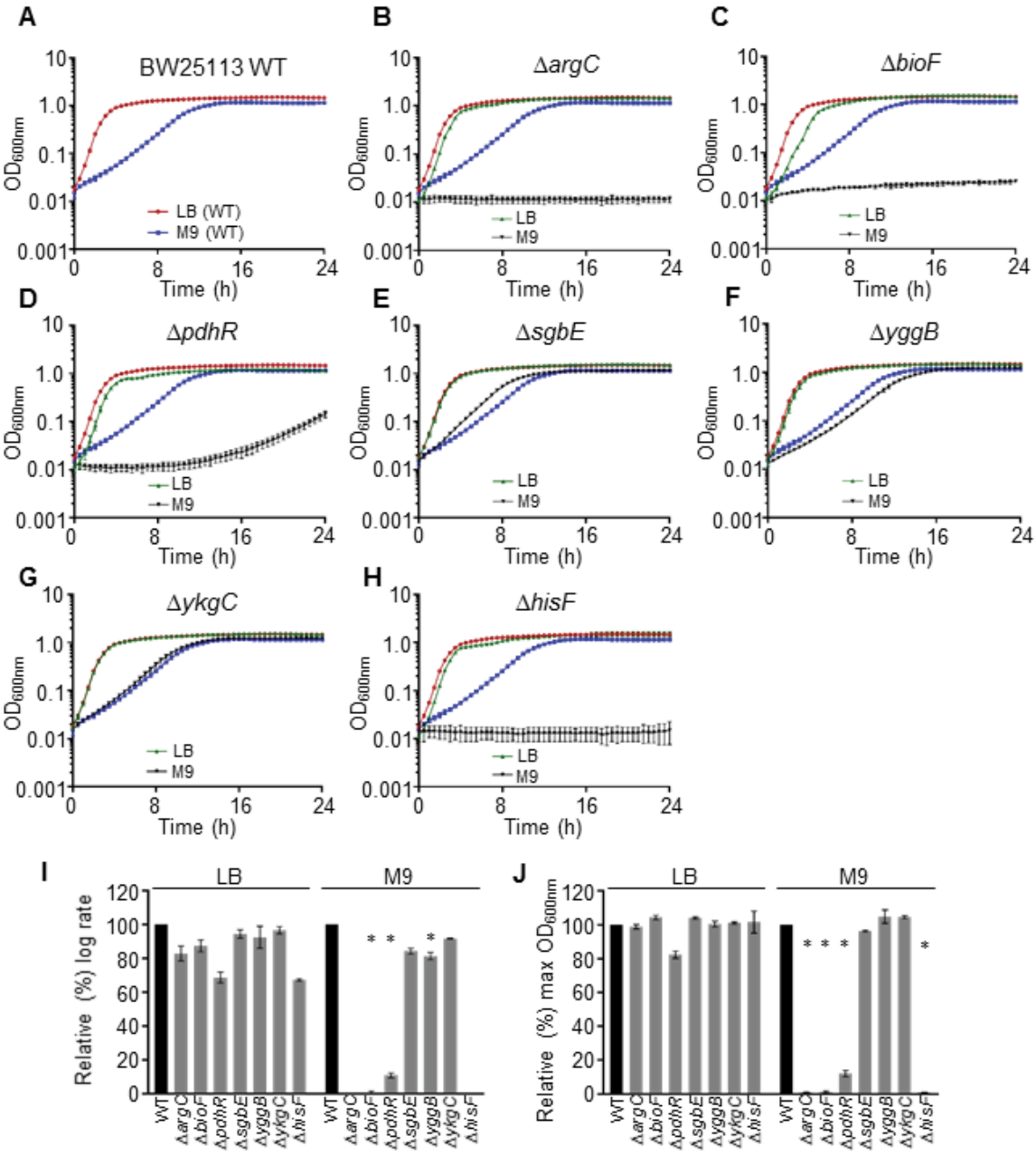


Figure 4.6. Characterization of deletion mutants in E. coli K12 identified by TraDIS.

Growth rate absorbance measurements (600_{nm}) of (A) BW25113 WT and (B-H) various mutants identified from TraDIS over a period of 24 h grown in LB and M9. The WT UTI89 growth curves in LB (red) and M9 (blue) from (A) are also shown throughout B-H for comparison. In B-H, each mutant grown in LB (green) and M9 (black). (B) ΔargC, (C) ΔbioF, (D) ΔpdhR, (E) ΔsgbE and (F) ΔyggB (G) ΔykgC and (H) ΔhisF. (I) Histogram plot of the proliferative log rate (μ) determined for the mutant growth curves relative to WT

*BW25113 (normalised to 100) grown in both LB (left-hand side panel) and M9 (right-hand side panel). (J) Similar to (I), displaying the maximum absorbance of each mutant relative to WT BW25113 (normalized to 100). (*P-value ≤ 0.05).*

4.3 Discussion

Identifying the metabolic enzymes and regulatory genetic pathways that allow *E. coli* to survive under certain restricted or complex growth conditions is vital. UPEC must adapt their metabolism according to the nutrients available in highly variable environments, such as the human urinary tract and within the intracellular cytoplasm of bladder cells. Our knowledge of gene essentiality under nutrient-limited conditions remains incomplete, despite its relevance to survival and infection. The screening of mutants on M9-glycerol minimal medium using TraDIS reported here has expanded our knowledge on UPEC gene essentiality data and metabolic capacity, which is expected to provide necessary data relevant to UTIs and towards the goal of developing new treatments for these and other infections. We identified 60 mutants with significant fitness defects and reduced ability to survive in M9-glycerol minimal medium when compared to rich LB medium. Two of these identified genes, *hisF* and *neuC* were confirmed for their fitness defects in M9-glycerol minimal medium and verified as important during UPEC *in vitro* bladder cell infection when compared to the WT strain.

TraDIS was first used to study the essential gene profile and determine the role of *Salmonella enterica* serovar Typhi genes when grown in the presence of bile salts (*123*). Since its introduction, similar techniques such as Tn-seq (transposon-sequencing) (*173*), InSeq (insertion-sequencing) (*174*) and QIseq (quantitative insertion-site sequencing) (*175*) have also been used in a range of species to determine gene essentiality and track insertion-site depth in various conditions. TraDIS was first applied to UPEC to identify the gene profile required for *in vitro* survival and growth of EC958 (an *E. coli* ST131 clonal lineage) in human serum (*124*). Here we have developed a modified TraDIS DNA sequencing library preparation protocol involving tagmentation, similar to recent reports (*124, 171*), which eliminates the need for custom sequencing primers and dark-cycles during sequencing by simultaneously sequencing transposon insertion-sites from both ends of the transposon. We show here the successful application of this protocol to a moderate size mutant library—making this protocol more amenable for infection studies. The moderate library size is well suited to infection models of UTI, in that sufficient data can

be collected for assessment of the fitness of the individual mutants (*i.e.* avoiding bottlenecking). At the same time, the library is large enough to ensure that majority of genes (~ 80%) are represented. Our modified TraDIS protocol simplifies sample preparation and requires much less total DNA than the original method—also more amenable for infection studies. We observed differences in the numbers of reads matching the transposon sequence when sequencing from each end of the transposons (Supp. Data Table S2A), likely due to differing specificities of the Tn-specific primers. In some cases, only ~30% of reads contained the correct transposon sequence. This did not prevent highly-accurate insertion site mapping of the Tn-tagged reads, but future optimization of the Tn-specific primer sequence or PCR conditions may be beneficial to maximise the numbers of useful reads obtained from the data. A similar theme has been reported previously when using a transposon-specific primer for amplification purposes (*124*).

We sought to characterise the change in representation of mutants over time by passaging the *E. coli* UTI89 Tn-library five consecutive times in LB media. Insertion site depth comparison between the first and fifth day samples was performed, which revealed 196 genes that displayed reduced abundance of their corresponding mutants in the library at day five. This revealed genes within the lipopolysaccharide (LPS) core of *E. coli* including *rfaY*, *rfaY*, *waaL* and *waaV* (Supp. Data Table S2F). Interestingly, a deletion mutant of *rfaY* is reported to increase sensitivity to SDS and the antibiotic, novobiocin (*176*). The identification of the LPS and O antigen synthesis pathway mutants, which are required for core completion in *E. coli* suggests that long-term growth of these mutants may destabilise the core membrane in *E. coli*. The core structure in *E. coli* is composed of lipid A (membrane anchor), an oligosaccharide component attached to lipid A and the outermost O antigen domain; the deletion of these biosynthetic pathways suggest slower growth and possibly lethal growth over time (*177*). We also identified numerous fimbriae-adhesion systems, including the Type 1 pili, P, S and the putative chaperone-usher *yad*-operon (Supp. Data Table S2F). These fimbriae systems have been shown to adhere to abiotic surfaces and provide UPEC with a competitive advantage when expressed (*178, 179*). During LB suspension growth their effects might relate to a stabilisation of the outer membrane (*180*). The *yad*-operon in particular has been reported to be under control of the

RpoS sigma factor responsible for controlling expression of genes during stressful, starvation culture conditions (*179*).

Interestingly, we also identified 16 mutants that displayed a significant increase in relative read abundance by day 5 of passaging and these mutations are considered to be advantageous for UPEC during laboratory long-term growth (Supp. Data Table S2E). As an example, *cytR* was identified, which belongs to the *galR/lacI* family of transcriptional regulators and controls the repression of several genes required for nucleoside uptake, transport and metabolism in *E. coli* (*181*). Among the genes controlled by *cytR* include the metabolic catabolising enzymes *deoABCD*, *udp* and *cdd* (*182*). Interestingly, TraDIS analysis also implicated the *lacI* repressor mutant as a fitness advantage in LB culture (Supp. Data Table S2E). The identification of these repressor genes (*cytR* and *lacI*) suggest that they may be metabolically costly for *E. coli* during long-term growth in laboratory culture and the deletion of these repressors may confer a fitness advantage for persistent growth. Repressors within *E. coli* (such as the *lacI* gene) are known to be constitutively expressed at low levels to facilitate a rapid change in regulation for the operon, which likely leads to an energetic cost burden for the cell. Langridge *et al*. (*123*) also identified numerous other operon repressor genes, such as *purR* (purine nucleotide operon repressor), *fruR* (fructose operon repressor), *mtlR* (mannitol operon repressor) and *nagC* (N-acetylglucosamine operon repressor) to give a fitness advantage to *Salmonella* Typhi when inactivated with a Tn5 insertion and passaged long-term in LB media (*123*). Further work to understand the basis of the apparently enhanced growth of these mutants in operon repressors is needed.

Determining the genetic requirements of UPEC during host infection remains an attractive goal for researchers. However, limitations exist in the number of mutants that can be represented for analysis with the limited scale of certain infection-culture models (bottlenecking). Even when we bottlenecked the number of viable bacteria used for inoculating the M9 and LB cultures, the current Tn-library size and TraDIS protocol still enabled us to screen a majority of the UTI89 genes and clearly identify 60 statistically significant mutants within both duplicate experiments that were required for growth in M9.

A large number of these genes were found to be involved in amino acid biosynthesis, carbohydrate metabolism, inorganic ion transport, and energy production and conversion (Fig. 4.3). Several of these genes were also found to be in common with a previous study identifying genes essential in M9-glycerol minimal media utilising *E. coli* K-12 and the Keio collection of mutants (*166*). Our UTI89 TraDIS analysis of M9-glycerol conditions identified 62 genes in either duplicate culture of the 119 genes identified as important in M9-glycerol by Joyce *et al*. (*166*). Thirty-five additional genes were identified in our TraDIS screen in common with Joyce *et al*., however these genes did not pass our stringent statistical threshold (P-value < 0.05 and Log_2FC < -1.0), and may require greater sequencing depth or a higher-density Tn-library. It must also be noted that the previous study (*166*) did not take into account any gradual or modest fitness (growth rate) defects—only an assessment of the presence or absence of growth in M9-glycerol was detected. Interestingly, we identified 33 genes that were not identified in the previous work (*166*), two of these being the *hisF* and *neuC* gene, and these may reflect the *E. coli* specific strain differences (certainly *neuC* as it is not conserved in K-12). Several previously hypothetical genes were identified that may represent novel genes that play a role in metabolic pathways. Several UPEC-specific genes were also identified, such as *neuC*, and these may play a role in pathogenesis within the nutrient-restricted urinary tract environment.

During intracellular infection, UPEC aggregate into dense biofilm-like clusters within the bladder cell cytoplasm and largely rely on the capacity for *de novo* biosynthesis of essential metabolites, amino acids and sugars (*20, 21, 96*). The biosynthesis of these resources by UPEC is supported by the importance of the amino acid histidine biosynthesis and transport gene *hisF* to grow in M9-glycerol. Additionally, *neuC* plays a direct role in *de novo* biosynthesis, activation and polymerisation of sialic acid in UPEC. Sialic acid not only is required for K1 capsule production, but is also used as a carbon and nitrogen source for the cells. In *E. coli*, sialic acid becomes converted into fructose-6-phosphate, which enters the glycolysis pathway. In UPEC, the *de novo* biosynthetic pathway of sialic acid exists (through *neuC*), however in commensal strains of *E. coli* without *neuC*, only import and catabolism of sialic acid from the environment is possible. This was confirmed with an *E. coli* MG1655 deletion mutant of the *nanAT* genes (required for sialic acid import and

degradation) displaying a significant defect compared to WT in colonisation of the mouse gut, where sialic acid-rich mucins are present. For UPEC, sialic acids represent an important and preferential bacterial carbon source in the urine (*20*), bladder (*183*) and bloodstream (*16*) and no doubt allow it to thrive in the sialic acid-rich mucins of the gut until dissemination into other host environments. When UPEC is grown in M9-glycerol, the lack of sialic acid in the media forces the cells to undergo *de novo* biosynthesis, however the deletion of *neuC* results in slower growth than WT. Both *neuC* and *hisF* were tested for their role in UPEC survival during *in vitro* bladder cell infection. This revealed that inactivation of both *neuC* and *hisF* resulted in a reduction of UPEC colony forming units 24 h post-infection when compared to WT (Fig. 4.5). This highlights the important role metabolic genes play in adapting to conditions experienced by UPEC and implicate *hisF* and *neuC* in UTIs.

In a recent study by Goh *et al*. (*171*), the *neuC* gene was not identified in a TraDIS screen for genes involved in capsule production, likely due to its high A-T nucleotide content and subsequently, low number of Tn5 insertions. We reported a total of 14 unique insertion sites within the *neuC* gene giving a total of 302 generated sequence reads after overnight growth of the library pool in LB media (Supp. Data Table S2E). This equates to an average insertion site every 84 bp in the *neuC* gene (the length of *neuC* is 1,176 bp). Differences in the frequency of insertions in *neuC* may be due to slight differences in the transposon and strain of *E. coli* (Goh *et al*. (*171*) used a ST95 hyper-virulent strain). The specific selection pressure Goh *et al*. (*171*) applied to the library pool, that being exposure to lytic K1 phages to kill capsulated mutants and harvest those unencapsulated mutants when grown on solid LB media, could possibly explain the absence of *neuC* as a significant hit in their TraDIS screen. Our *neuC::kan* deletion strain grown in liquid LB media had no growth defects as its likely sialic acid was imported from the rich-complex media – this is also likely in Goh *et al*.

Uropathogenic *E. coli* must adapt their metabolism according to the nutrients available in highly variable environments. This metabolic capacity is essential for infection and virulence. Here, we constructed a transposon-insertion mutant library in UTI89 at a

moderate scale and combined this with a modified TraDIS library preparation protocol. By utilising TraDIS analysis, we identified 60 mutants with a significant fitness defect and ability to survive growth in M9-glycerol minimal medium when compared to LB, the majority of which encoded gluconeogenic and amino acid catabolism proteins. We confirm a subset of mutants for their importance in growth in M9-glycerol and highlight novel differences in the metabolic requirements between uropathogenic and commensal *E. coli*. Both the loss of *hisF* and *neuC* resulted in a reduction of UPEC colony forming units 24 h post-infection when compared to WT. Several other uncharacterised and UPEC-specific genes were identified that likely underlie metabolic capacities of UPEC strains during infection.

4.4 METHODS

4.4.1 BACTERIAL STRAINS AND GROWTH CONDITIONS

E. coli BW25113 WT and mutants in the following genes: *hisF*, *argC*, *bioF*, *yggB*, *ykgC*, *sgbE* and *pdhR*, were obtained from the *E. coli* Genetic Stock Center (Yale University, USA). Uropathogenic *E. coli* UTI89(*102*) was kindly provided by J. Moller-Jensen. Mutant knockouts in UTI89 were constructed for this study and detailed below. Strains were routinely cultured at 37°C on solid or in liquid Luria Broth (LB), including 1% (w/v) NaCl, unless otherwise stated. For culturing the UTI89 transposon mutant library, kanamycin at 50 μg/mL was routinely used.

4.4.2 CONSTRUCTION OF A MINI-Tn5 TRANSPOSON MUTANT LIBRARY IN UTI89

E. coli UTI89 was made electro-competent and prepared for electroporation as previously described (*184*). Briefly, a 1 L culture suspension of UTI89 was grown in LB (no added NaCl) media to an OD_{600} of 0.4. The culture was harvested by centrifugation, then washed three times by centrifugation and resuspension in ice-cold 10 % (v/v) glycerol. The cells were then resuspended in a final volume of 2 mL of 10 % (v/v) glycerol and stored for subsequent use. A volume of 50 μL of electro-competent UTI89 was mixed with 1 μL of EZ-Tn5 <R6Kγori/KAN-2> Tnp transposomes (Epicenter Biotechnologies) and placed on ice for 5 min in a 2 mm electroporation cuvette. The cells were then electroporated using a BioRad GenePulser set to 2.5 kV, 25 μF and 200 Ω, giving a time constant of 4.5 – 5. Cells were then immediately resuspended in 1 mL of fresh SOC medium (*184*). This process of electroporation was repeated 11 times to achieve 11 batches of mutant cells. The resuspended cells were then pooled and incubated at 37 °C for 1 h. Following recovery, cells were spun down and concentrated to 1 mL and spread-plated onto 11 LB-agar petri plates supplemented with kanamycin (50 μg/mL). The total number of mutant colonies was estimated by plate counting, with 1/100 and 1/500 dilutions. All plates were incubated overnight at 37°C. Colonies from each agar plate were then resuspended with 1 mL of LB

supplemented with kanamycin (50 μg/mL) and 15 % (v/v) glycerol. The kanamycin resistant colonies were scraped off using a sterile biological scraper and mixed at room temperature to give a homogenous suspension, which was aliquoted and stored at -80°C for subsequent use.

4.4.3 MUTANT SELECTION IN MINIMAL MEDIA

An aliquot of the mutant library was thawed at room temperature and diluted accurately in sterile LB to precisely 300,000 individual cells. These cells were then used to inoculate 2 mL pre-warmed LB media and allowed to grow for 1 hr. A volume of 1 mL from this culture was used to inoculate 9 mL fresh LB supplemented with kanamycin (50 μg/mL) and the other 1 mL was used to inoculate 9 mL fresh M9 minimal media with 1 % glycerol supplemented with kanamycin (50 μg/mL). Fresh M9 minimal-medium with 1 % glycerol (v/v) as the carbon source was made as previously described (*166*). The cultures of LB and M9 were grown for 24 hours at 37 °C with 200 rpm shaking. A volume of 1 mL from each culture was then used to inoculate another 10 mL of fresh media, both LB or M9 appropriately, supplemented with kanamycin (50 μg/mL) and grown for another 24 hours at 37 °C with 200 rpm shaking. A final volume of 10 mL from this culture was then used to inoculate a final volume of 90 mL of fresh media, either LB or M9 appropriately, supplemented with kanamycin (50 μg/mL) and grown for a final 24 hours at 37 °C with 200 rpm shaking. After passaging, the cultures were then centrifuged down at 4,500 rpm for 30 min at room temperature and underwent genomic DNA extraction using Bioline Isolate II Genomic DNA (Bioline). A 1 mL aliquot of the UTI89 mutant library was thawed at room temperature and similarly underwent genomic DNA extraction using Bioline Isolate II Genomic DNA (Bioline). This experiment was performed in duplicate. A concentration of 5-10 μg of gDNA was purified across all samples.

4.4.4 TRANSPOSON-DIRECTED INSERTION-SITE SEQUENCING (TraDIS)

Genomic DNA (5 ng) from the transposon mutant library (grown as indicated in LB or M9 media) was first subjected to tagmentation using the Nextera DNA Library Prep Kit (Illumina). The DNA was mixed with 10 µL of 2x TD Buffer and 5 µL of Tagment DNA Enzyme TDE1 (diluted 1/50 in 0.5x TE buffer and 50% (v/v) glycerol) and made up to a final volume of 25 µL, then incubated for 5 min at 55°C. A volume of 5 µL of 0.2 % (w/v) SDS was immediately added to stop the tagmentation reaction and incubated for 5 min at room temperature. The sample was then PCR amplified with a transposon-specific primer and a barcoded Nextera primer (see Supp. Data Fig. S4 and Supp. Data Table S2H) using the KAPA HiFi Library Amplification Kit (KAPA Biosciences) as per manufacturer's instructions. As a control for non-specific amplification, identical PCR reactions containing a single primer only were also performed. An initial extension step was performed for 3 min at 72°C. This was followed by a denaturation step for 30 sec at 98°C, followed by 28 cycles of; 98°C for 15 sec, 55°C for 30 sec, and 72°C for 30 sec. A final extension for 5 min at 72°C was performed. Primer sets for both transposon ends were used (i.e. two PCR amplification reactions per sample). Ten sequencing libraries from the five samples (LB and M9 in duplicate; Stock mutant library in single) were pooled to equimolar concentrations and underwent SPRI-select magnetic bead clean-up (Beckman Coulter) at 0.9-0.5x left and right ratios, respectively (to select DNA of ~200-800 bp). Pooled sequencing libraries were quantified on an Agilent Bioanalyzer, before dilution to 10 pM and sequenced on a V3 paired end (PE) Illumina flow cell using an Illumina MiSeq sequencer for 300 cycles. PhiX genomic adapter-ligated DNA control library (Illumina)

4.4.5 ANALYSIS OF NUCLEOTIDE SEQUENCE DATA

Sequence reads from the FASTQ files were separated according to their unique index barcodes and underwent quality control analysis and filtering using the FastQC software tool as described in *Chapter 2*. Samples were then matched to their homology of the i5 and i7 ends of the transposon sequence (40 – 50 bp region) allowing for 3 bp mismatches. The transposon matching sequence was then trimmed from the reads and the reads were then

strictly mapped to the genome of UTI89 and plasmid pUTI89 using the SMALT short read mapper as described previously (*128*) (allowing 0 mismatches). Subsequent analysis steps such as raw read counts, number of insertion sites and identification of significant differences in mutant frequency between LB and M9 samples were performed as previously described(*128*). Artemis was used to visualise the alignments with the read data (BAM files) for inspecting. The sequencing depth of mapped reads and insertion sites were visualised using Artemis, version 13.0. The frequency of mapped reads and the analysis of data for Fig. 4.1, 4.2 and 4.4 were determined with deepTools2 software.

4.4.6 STATISTICAL ANALYSES FOR IDENTIFYING ESSENTIAL GENES

Using the Bioconductor package edgeR (*185*) and previously described R protocol (*128*), we identified the UTI89 genes required for survival in minimal-media (with 1 % (v/v) glycerol supplemented) by comparison of the read abundance across the genome of the LB control sample to the M9 sample. The raw read counts from each duplicate experimental sample were treated separately and used as input into the edgeR package. For essentiality between conditions, any genes that contained very low read depth (≤ 5 reads) were removed from the comparison. A criterion of P-value ≤ 0.05 was used to define a list of genes, which may have a reduced capacity to survive in M9 (Fig. 4.3 displays the COG functional classes of these genes). Within this criteria, those genes (60) that were found in common in both replicate experiments and the $\log_2$FC < -1.0 have strong implications as essential for survival in M9 are found in the stringent data of Table 4.1. All genes identified using TraDIS can be found in Supp. Data Table S2.

4.4.7 VERIFICATION OF INDIVIDUAL DELETION MUTANTS

Chromosome UTI89 gDNA was extracted and purified using Bioline Isolate II Genomic DNA (Bioline) and defined UTI89 mutants were constructed using the λ-Red recombinase protocol. The set of primers (Macrogen Inc, South Korea) used to amplify up the kanamycin cassette from pKD4 can be found in Supp. Data Table S2H. Mutants were

selected using kanamycin and confirmed using a combination of gene flanking and the internal kanamycin primer (K1, Supp. Data Table S2H). All PCRs were routinely performed at 60°C annealing temperature.

Each UTI89 mutant and WT was plated onto solid agar supplemented with kanamycin (50 μg/mL) and allowed to grow for 24 hr. Single colonies were picked and used to inoculate 5 mL of pre-warmed liquid LB and allowed to grow overnight. Next morning, 500 μL was used to inoculate 5 mL fresh M9 media (1 % glycerol supplemented) and allowed to grow for 24 hr. 200 μL of fresh M9 and LB media was placed into single wells of a 96-well plate (Falcon) and mutant and WT cultures were brought to a starting OD ~ 0.015. The plate was ran in a microplate reader (BioTek) at 37°C measuring absorbance at OD wavelength 600_{nm} in 30 min intervals for 24 hr (continuous shaking). LB and M9 media uninoculated was used as a blank to which readings were zeroed against to correct for background absorbance. The grow curves and analysis of Fig. 4.5 were constructed using technical replicates from two independent plate read data. Analysis involved the DMFit (DM: Dynamic Modelling, version 3.5) software based on the growth curve modelling of Baranyi and Roberts (*172*). Values such as the lag time, proliferation (μ) rate and maximum OD were extracted via the DMFit linear regression analysis tool and students t-test, two-sample assuming unequal variance was used to determine significance (P-value ≤ 0.05 was considered significant) when compared to WT. This approach was similarly applied to BW25113 and BW25113 mutants obtained from the *E. coli* Genetic Stock Center (Yale University, USA) displayed in Fig. 4.6.

4.4.8 TRANSPOSON MUTANT LIBRARY PASSAGE

50 μL of the mutant library aliquot was used to inoculate 20 mL of LB and grown for 24 h at 37°C with 180 rpm shaking. Subsequently, 50 μL of the culture was then used to inoculate another fresh 20 mL of LB and similarly grown for a further 24 h. This was continued for a total of five consecutive passages. Genomic DNA was extracted directly from harvested cells from 10 mL of passage one, three and five using Bioline Isolate II Genomic DNA (Bioline). The DNA was prepare for sequencing as described above and

quantified on an Agilent Bioanalyzer, before dilution to 10 pM and sequenced on a V3 paired end (PE) Illumina flow cell using an Illumina MiSeq sequencer for 300 cycles. PhiX genomic adapter-ligated DNA control library (Illumina) was included at 5%. The sequencing depth of mapped reads and insertion sites were visualised using Artemis, version 13.0 and can be found in Supp. Data Table S2.

4.4.9 HUMAN BLADDER CELL CULTURE MODEL OF UTI

Human bladder cell line PD07i was used routinely, cultured at 37°C with 5% CO_2 to confluency as previously described (*26*). 0.5% Trypsin-EDTA mixture and 1x Defined Trypsin Inhibitor (DTI) was routinely used to unstick bladder cells from base of flasks. Cells were routinely cultured in EpiLife media supplemented with HKGS and appropriate antibiotics. Bladder cells were grown to confluent and counted using a Coulter Counter (Beckman) to a minimum of $1x10^5$ cells/mL. Bladder cells were concentrated to 4 x 10^5 /mL and 1 mL was dispensed into a 6-well sterile culture dish (BD Falcon; 35 x 18 mm^2) and incubated at 37°C with 5% CO_2 for 24 h. Bacterial culture was grown overnight and then diluted down to an OD_{600nm} of 1.0. This was then used to inoculate the bladder cells by adding 0.1 μL/mm^2 of bacteria to give a multiplicity of infection (MOI) of 100. The plates were then spun down for 5 min at 500 rpm and incubated statically for 2 h at 37°C with 5% CO_2. Following this, the supernatant was discarded and the bladder cells were resuspended in gentamicin (100 μg/mL) supplemented EpiLife and HKGS. Cultures were then incubated statically at 37°C for 22 h. Following this, culture dishes were washed gently with PBS 4 times and lysis solution (0.5% trypsin-EDTA, 0.1% triton X-100) was added and incubated for 10 min. The entire supernatant was then collected and plated onto LB-agar petri plates and left to incubate at 37°C for 18 h. Colonies were then counted from a 1/100-dilution series. This experiment was performed using 3 technical replicates.

CHAPTER 5

THE ESSENTIAL GENE PROFILE SET OF UROPATHOGENIC *E. COLI* DURING BLADDER CELL INFECTION

5.1 Introduction

Uropathogenic *E. coli* (UPEC) originate from the human intestinal microbiome where they rarely cause complications or disease. However, UPEC have adapted the ability to disseminate and colonise the human urinary tract, which is facilitated by numerous virulence factors that play specific roles in a multi-stage intracellular infection cascade within the bladder cells (*17*). During this cascade, UPEC undergo various morphological changes resulting in a biofilm-like intracellular bacterial community (IBC), an elongated filamentous phenotype, subsequent rupture of the host cell and reversal to bacillary rod-shape (*17, 26*).

In the course of a UTI, UPEC first adhere to the luminal surface of the bladder epithelium. Binding is mediated by different types of bacterial surface adhesive fibers, particularly Type-1 pili that recognise mannosylated glycoprotein receptors present on the host epithelium (*27, 34*). Once bound, UPEC may become internalised via endocytosis, which results in host cell membrane rearrangement, engulfing UPEC into membrane-bound vacuoles (*43*). UPEC can then escape from these compartments to the cytoplasm and proliferate rapidly, forming IBCs (*52*). The mechanism of escape into the host cytoplasm is not completely clear (*48*). In the formation of the IBC, rod-shaped bacteria divide into small coccoid daughter cells that are tightly packed together (*52*). Here, regulatory systems sense and respond to bladder environmental cues to alter expression of toxins, flagella, iron-acquisition and other virulence factors that play specific roles to maintain intracellular infection (for more detail, see *Section 1.3.2 Intracellular Bacterial Communities*). Despite this, the complete suite of factors that are involved in the initial formation of the IBC, and the regulatory systems that sense and respond to bladder cell cues to maintain continued community growth still remain to be determined.

As the IBC reaches an advanced state, with up to 10^5 bacteria overwhelming the host cell cytoplasm, subpopulations of highly-motile and filamentous UPEC develop (*17*). The development of filamentous bacteria accompanies host cell rupture and dispersal (*17*). Filamentation is also known to assist in providing greater attachment to urinary tract

surfaces and avoid consumption by the host immune response (*31*). Additionally, filaments have the potential to revert to bacillary rod-shaped morphology and can thereafter divide as normal bacterial cells, thereby reinitiating a new infection cycle (*26*). This highlights the important role that filamentation, its reversal and regulation play during the course of a UTI.

There is one well-characterised *E. coli* response pathway that leads to filamentation: the "SOS" response to DNA damage (*141*) (see *Section 1.3.3.3 The SOS Response and Involvement of SulA*). A gene known to be responsible for filamentation during the SOS response in *E. coli*, called *sulA*, was proposed to be involved in the filamentation seen in UPEC during a UTI(*58*). However, extensive SulA-independent filamentation by UPEC in the human flow chamber-based cell culture model has been observed (*56*), which suggests that UTI-filamentation can be induced via alternative pathways. It was demonstrated that UPEC filamentation during UTIs requires the known cell division protein, DamX, which normally contributes to cell envelope synthesis (*56*) (see *Section 1.3.3.4 damX and Reversible Filamentation during UTIs*). However, the complete suite of factors that are involved in the filamentation and reversal response pathway (sensing external stimuli and transducing the signal to effector molecules to arrest division) still remain largely unknown and could involve an array of other known and novel cell division regulators.

In the work described in this chapter, we developed an up-scaled *in vitro* human bladder cell culture model of UTI, and coupled this with high-throughput TraDIS analysis to determine the essential gene profile set of UPEC harvested from the IBC, dispersal and recovery (filament reversal) phases of infection. Our up-scaled cell culture model of UTI allows large yields of bacteria to be harvested at distinct time points in the course of the infection. Used in combination with the TraDIS technique, this allowed us to identify the genetic requirements of UPEC during *in vitro* infection. This will help uncover the UPEC strategies and mechanisms underpinning survival within the bladder and facilitate development of novel treatments and therapies for UTIs in future.

5.2 RESULTS

5.2.1 DEVELOPMENT OF AN UP-SCALED INFECTION MODEL TO ASSAY UPEC GENE ESSENTIALITY DURING UTIs

The intracellular infection cascade was first visualised in infected mouse bladder explants using time-lapse microscopy (*29*). However, the *in vivo* model has limitations for dissecting bacterial molecular pathways and responses, particularly for genetic screening. Such limitations are the low bacterial yields harvested from mouse bladders. For this reason, *in vitro* cell culture models have become attractive in helping to elucidate the UPEC molecular pathways at play during infection, in particular the later phases. However, few studies have aimed to systemically analyse the infection cascade of UPEC and much remains unknown during the late phases of infection, such as the IBC, dispersal and recovery phases. In order to better elucidate these later phases of infection, we first developed and characterised an up-scaled *in vitro* bladder cell infection model for the purpose of harvesting large bacterial yields at distinct time points. The up-scaled model makes use of the gentamicin protection to select for bacteria within the bladder cells (IBCs), followed by concentrated urine exposure to induce filamentation, dispersal and rupture of the host cells. It was found that a continuous flow of fresh urine, as used in previous infection *in vitro* models of UTI, was not needed to induce IBC dispersal and filamentation, although periodic replacement of urine was performed to withdraw and collect the dispersed bacteria for analysis. The released bacteria were then collected and recovered in liquid LB media to induce bacterial recovery and filament reversal from the infection conditions (irreversible bacterial filamentation is known to be eventually lethal).

In order to characterise the up-scaled infection model, we used UTI89 transformed with the pGI5 plasmid (to provide constitutive GFP expression; Iosifidis, G., *et al.*, unpublished data) to infect bladder cells. The first 2 h of infection had conditions for UPEC attachment and invasion of the bladder cells. Following this, the supernatant liquid was removed and bound cells washed, and then incubation was continued, including gentamicin to kill the extracellular bacteria, until 24 h post-infection, to allow intracellular UTI89 to form IBCs.

At 18 h and 22 h post-infection, the bladder cell cultures were visualised by microscopy (Fig. 5.1). This revealed bladder cells that contained dense bacterial aggregates, which resembled previously identified IBCs formed in an *in vitro* cell culture model (*52*) and the murine model (*17*). As an observation throughout and evident in Fig. 5.1, not all bladder cells contained an IBC, despite the high MOI of 100, and this is also consistent with previous reports (*26*) (Iosifidis, G., *et al.*, unpublished data). The sizes of the aggregates at 22 h post-infection were larger on average, almost certainty due to the extra 4 h of intracellular growth.

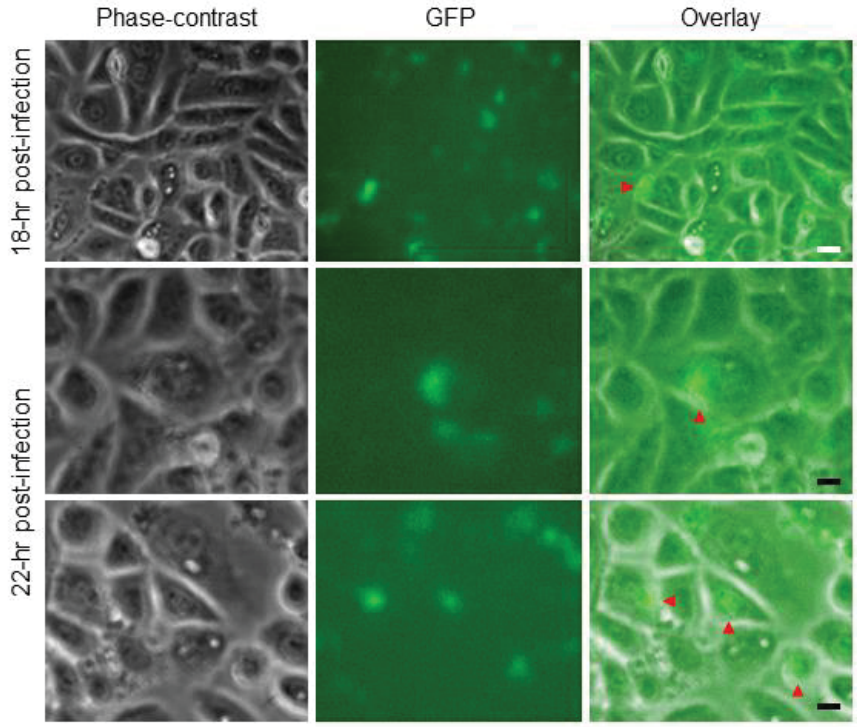


Figure 5.1 Up-scaled bladder cell infection model permits UTI89 IBCs

UTI89/pGI5 (wild-type with GFP constitutively expressed) formed IBCs within the up-scaled bladder model (red allow heads), which represent previously described IBCs, albeit with lower resolution due to the limitations imposed by the culture dishes. PD07i bladder cells grown in 100 x 20 mm culture dishes were inoculated with UTI89/pGI5 after static growth in liquid LB media. After gently spinning down culture dishes to induce UTI89-bladder attachment, the cultures were left to grow statically for 2 h. Gentamicin containing media was applied and culture dishes were incubated statically for a further 22 h to induce intracellular bladder cell growth. Bladder cells were imaged using GFP fluorescence at 18 and 22 h. Some bladder cells are colonised by UTI89, while others appeared to remain uninfected. The images are from the first of two biological replicates of the infection model characterisation experiments. Bars, 15 μm (white) and 10 μm (black).

We next sought to characterise the dispersal phase of infection by quantifying the extent of filamentation and bacterial biomass released from the bladder cells upon rupture. Concentrated urine induces a sub-population of bacteria as filaments (*26*), dispersal and rupture of the host cells. In order to achieve this, 20 min collection windows and longer 100 min windows of urine treatment were carried out during the 8-20 h period after commencing urine exposure. The first 8 h of urine exposure results in scarce bladder cell rupture and dispersal events so we did not sought to collect these (Iosifidis, G., *et al.*, unpublished data). The harvested cells were immediately fixed and analysed using microscopy and flow cytometry (Fig. 5.2). Microscopy of the 20 min collection windows revealed large bacterial filaments entangled in rod-shaped cells and remnants of bladder cells (Fig. 5.2A-G). These filaments are characteristic of previously identified filaments dispersing from bladder cells (*26*) and some were observed to be ~100 μm in length (Fig. 5.2B-D). We used flow cytometry analysis as previously described (*Chapter 3*) to quantify the extent of filamentation of cells harvested from the dispersal phase within the 20 min and 100 min windows of urine collection and compared this to the original inoculate culture (pre-infection). Fig. 5.2H demonstrates the gradual right-hand shift of increasing cell size (as represented by the flow cytometer's side-scatter – signal-area parameter) as the window collections progressed from 8-20 h. The cells harvested at 8 h (0-8 h of urine exposure; orange peak in Fig. 5.2H) depicts a bimodal population of short and filamentous cells, suggesting that any filamentous cells dispersed during this large window may have had time to reverse. This is consistent with the 20 min windows of collection displaying more filamentous cells on average than the longer window collections (Fig. 5.2H).

To determine what proportion of cells harvested within the 20 min windows were considered filamentous, we gated 3 populations of urine collection; early (10-10:20 h), middle (14-14:20 h) and late (20-20:20 h) (Supp. Data Fig. S6). Most filamentous cells upon dispersal were present in the middle population (~ 14 h) of urine exposure, with approx. 30% of filaments, whereas the early and late populations contained ~ 25% of filamentous cells (Supp. Data Fig. S6). This data is consistent with the quantified biomass of cells harvested at each 20 min and 100 min window (Fig. 5.2I). We measured the biomass at A_{600} in order to estimate the distribution of bacterial yield released from the

bladder cells overtime. This revealed a biomass peak at 14 h with a gradual plateau to 20 h of urine exposure (Fig. 5.2I).

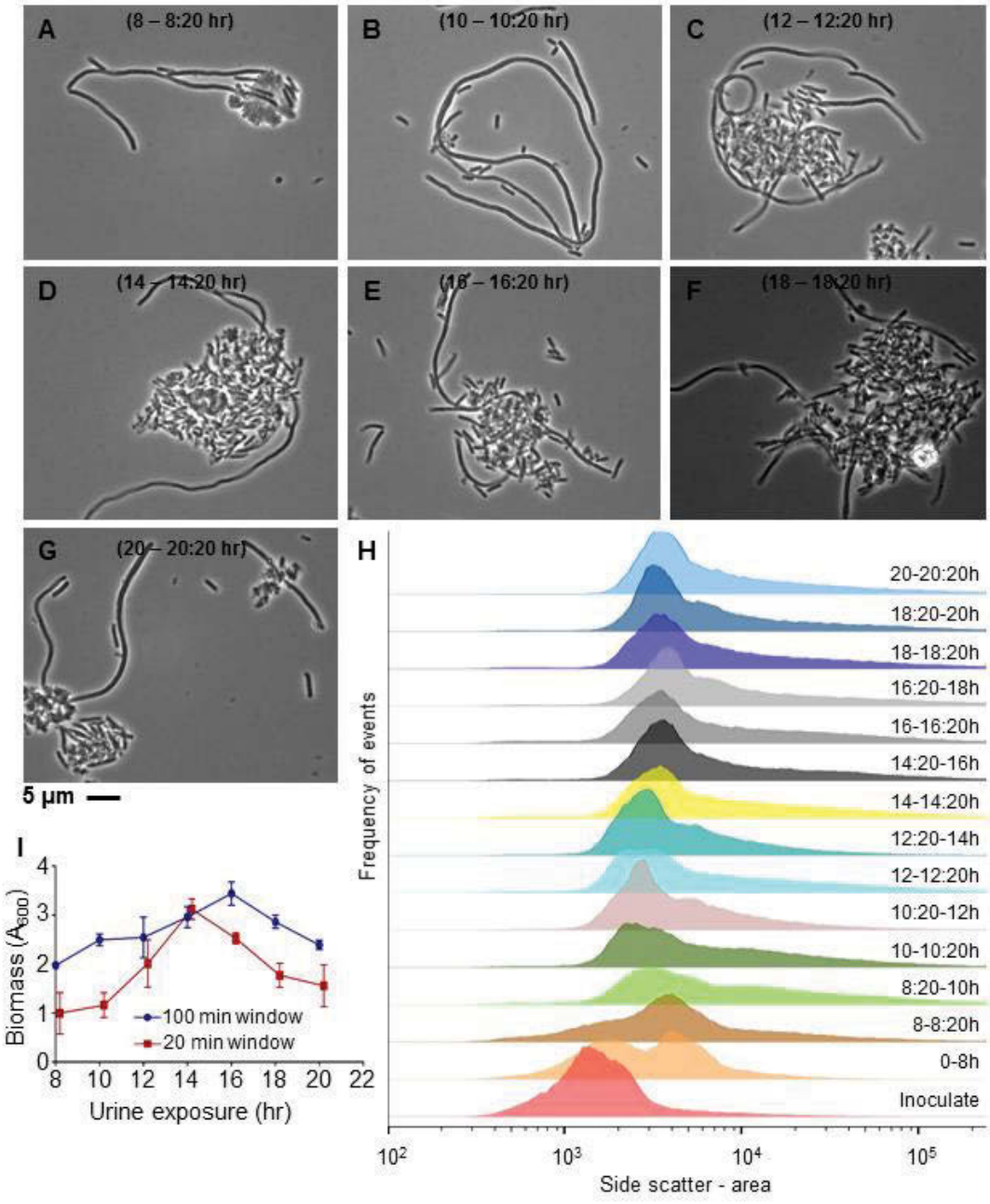


Figure 5.2 Characterisation of UPEC dispersal in vitro and filamentous cells

At 24 h post-infection of PD07i bladder cells, concentrated urine is added to induce dispersal and the culture dish left to incubate for 20 h at 37°C with gentle orbital shaking. During this 20 h period, short 20 min windows and long 100 min windows of urine was collected from the 8th h. These collected cells were immediately fixed with 2% (w/v) formaldehyde and analysed. (A-G) Phase-contrast microscopy reveals characteristic

filamentous UTI89 as previously reported (26). The sub-population of filaments appear to be tangled with remnant bladder cells and short cells. Bar, 5 μm. (H) Flow cytometry analyses was performed to determine the size distributions over the course of the dispersal phase. The Inoculate sample culture was treated as the starting control reference point. The histograms show a gradual shift towards a filamentous cell population after the first 10 h of the urine exposure phase, after a biphasic distribution of cell sizes appears early on. (I) The 20 min window and 100 min window collected cultures were measured on a spectrophotometer at A_{600nm} to estimate bacterial yield. A peak biomass is evident at ~14 h of urine exposure, which appears to plateau by 20 h. A notable different is seen between the biomass of the 100 min cultures when compared to the 20 min cultures, certainly due to the extra growth time allowed.

From this data we suggest that during the early stages of urine exposure, UTI89 are within the IBC and many have not yet developed to the dispersal phase. As time progresses, peak dispersal and bladder cell eruption occurs releasing a sub-population (~30% within our model) of filamentous cells. This peak dispersal period plateaus off, possibly due to large irreversible damage of the bladder epithelium and sustained urine exposure. From this data we decided to collect cells for our subsequent TraDIS analysis every 20 min window from 10-20 h (10 h window) of urine exposure to maximise the bacterial yield and filamentous cells collected.

5.2.2 THE ESSENTIAL GENE PROFILE SET OF UPEC DURING IN VITRO INFECTION

In order to identify bacterial genes important to bladder cell infection, we performed TraDIS on samples collected at defined stages of the infection cycle, with particular focus on the IBC, dispersal and recovery/reversal phases. We made use of the previously described UTI89 transposon library of ~20,000 mutant clones to infect bladder epithelial cells in the up-scaled cell culture model of UTI described above, (see *Chapter 4* for transposon library construction and application in TraDIS). Mutant clones were harvested just before infection (Inoculate sample), at 24 h post-infection (IBC sample), during a period of 34-44 h post-infection (Dispersal sample; 20 min collection windows) and immediately following 44 h post-infection (Non-dispersed sample; to collect those IBCs that did not disperse out of the bladder cells). Cells harvested from the Dispersal sample also underwent 18 h of growth in liquid LB to induce filament reversal and bacterial recovery (Recovery sample), and then this culture was diluted to undergo a further 24 h of growth in liquid LB in order to more clearly define the surviving mutants (Recovery-24h sample) (Fig. 5.3). Each of 2 biological sample replicates harvested from infection was made up of 3 technical culture-dish replicates pooled together (further scale-up). The Inoculate sample was done as a single biological replicate. This Inoculate sample, the UTI89 mutant library grown statically in liquid LB, was used as the first reference point representing the viable mutants just prior to inoculation of the bladder cells (Fig. 5.3).

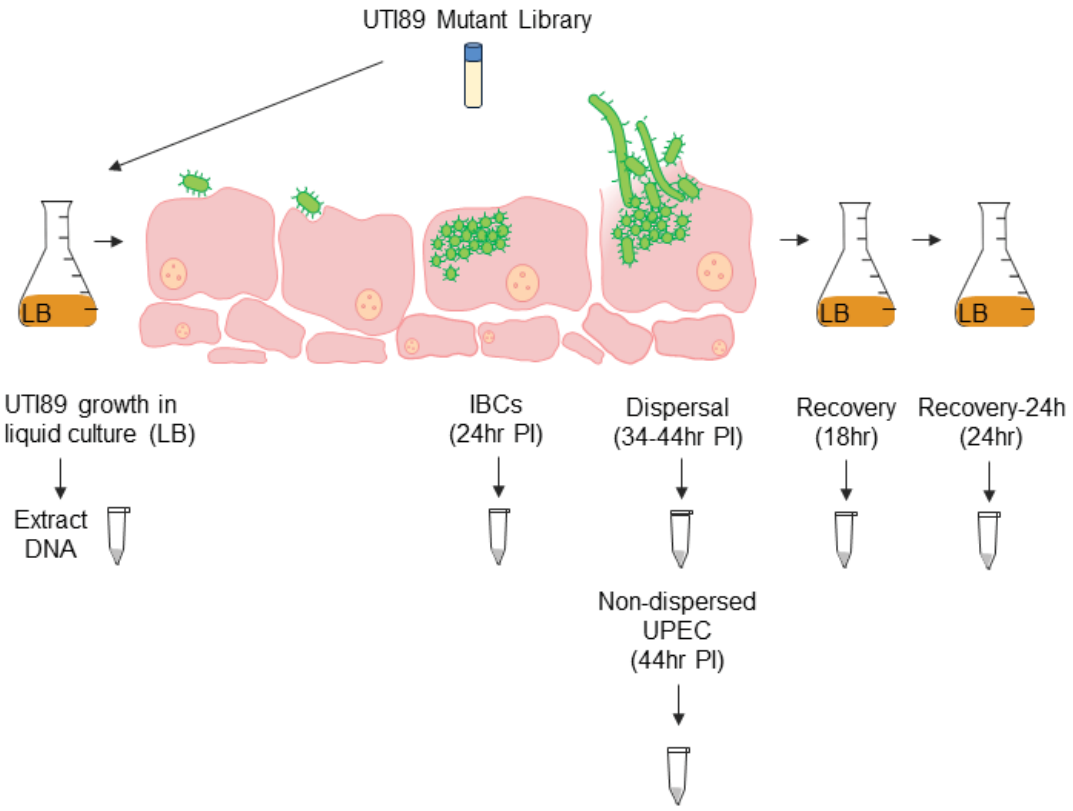


Figure 5.3 Schematic diagram of TraDIS approach

The UTI89 transposon mutant library, grown statically in liquid LB media (pre-infection; Inoculate sample), was used to inoculate PD07i bladder cells grown in culture dishes. After centrifuging the culture dishes to increase bacterial attachment, the cultures were left to grow statically for 2 h, to allow invasion to take place. Gentamicin-containing media was applied and culture dishes were incubated statically for a further 22 h to induce intracellular bacterial growth. The bladder cells were lysed and total culture was collected (IBC sample). Immediately following the IBC phase of infection, bound cells were gently washed and sterile human urine was applied. UTI89 was also harvested over a 10 h period (34-44 h post-infection) of urine exposure (Dispersal sample; collecting urine every 20 min). At 44 h, the bladder cells were lysed and total culture was collected (Non-dispersed). Cells harvested from the Dispersal sample also underwent 18 h of growth in liquid LB to induce filament reversal and recovery (Recovery sample), and a sample of this culture underwent a further 24 h of growth in liquid LB (Recovery-24h). All 6 samples underwent DNA extraction and purification for subsequent TraDIS. (PI = post-infection).

Genomic DNA was extracted from each sample and sequenced using the modified multiplex TraDIS protocol described previously (see *Chapter 4* for TraDIS methodology and sequencing preparation of DNA). Sequencing was performed on an Illumina HiSeq 2500 sequencer for 250 cycles; generating an average of 5.5 million reads per sample, with an average of 49% of reads identified containing the transposon specific sequence (Supp. Data Table S3A). Of these reads tagged with the transposon, between 73 – 91% were reliably mapped to the chromosome of UTI89, and 2 – 5% reliably mapped to pUTI89 (Supp. Data Table S3A).

In order to identify the transposon mutants that were not able to survive progression through the infection cascade, each sample was compared to each other as such: The Inoculate sample was compared to the IBC, the IBC sample was compared to the Dispersal, the IBC sample was also compared to the Non-dispersed bacterial sample, and finally the Dispersal sample was compared to both the Recovery and Recovery-24h samples. The raw sequence read counts and unique insertion sites identified within each gene throughout the samples can be found in Supp. Data Table S3B. Insertion-site depth and frequency between conditions was analysed using the Bioconductor package edgeR as previously described in *Chapter 4* and by Barquist *et al*. (*128*). Important genes were identified based on a significant reduction in read frequency between selected conditions (i.e. less mutants survived the IBC sample when compared to the Inoculate sample). The duplicate samples were used as input into edgeR analysis. The complete output of genes between the compared conditions and their statistical strength can be found in Supp. Data Table S3C-G. We applied a stringent gene selection P-value (P-value as determined by edgeR analysis) cutoff criterion of ≤ 0.05 and Log_2FC (fold-change) cutoff of < -1.0 to reduce the likelihood of false-positive results, and identify the most clearly important genes for the IBC, dispersal and recovery phases. Further, in order to determine the subset of these genes required specifically for infection, we eliminated those genes that were previously shown to be important for growth in LB (cultures passaged daily, over 5 days) in liquid LB media (data presented in *Chapter 4*). This revealed a total of 143, 333 and 322 statistically significant genes required for the IBC, dispersal and recovery phases, respectively, when compared to their preceding phases as described above (Figure 5.4). Within this stringent gene selection

criterion, 389 and 385 genes were also identified to be important in the Non-dispersed and Recovery-24h samples, respectively (Figure 5.4). When comparing the genes identified between the Dispersal and Non-dispersed samples, 209 genes were found in common, whereas 271 genes were found in common between the Recovery and Recovery-24h samples (Fig. 5.4 and Supp. Data Table S3).

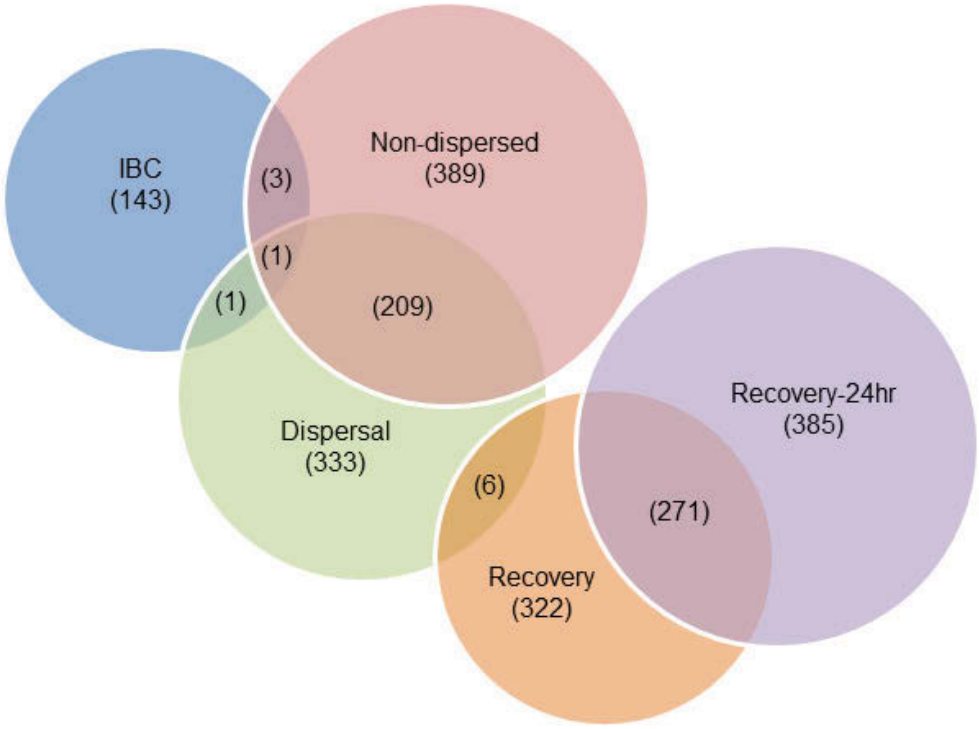


Figure 5.4. The numbers of genes identified as important to stages of infection by TraDIS, indicating the numbers of genes common to multiple stages

Venn diagram of genes identified by TraDIS in the phases of UPEC infection with P-value < 0.05 and log_2FC < -1.0. The IBC, Dispersal and Recovery samples showed 143, 333 and 322 genes, respectively, that were statistically significant. 209 genes were identified in common between the Dispersal and Non-dispersed samples when both were compared to the IBC sample (Supp. Data Table S3). Similarly, 271 genes were identified in common between the Recovery and Recovery-24h samples when both were compared to the Dispersal sample (Supp. Data Table S3).

5.2.2.1 The genes required for progression to IBCs

As described in the overview to this experiment in *Section 5.2.2*, in order to identify the UPEC genes required for progression to 24 h post-infection (the IBC phase of infection), the UTI89 transposon library was used to infect bladder epithelial cells. At 24 h post-infection, the bladder cells were lysed to reveal the intracellular bacteria and the entire culture harvested. This IBC sample of infection was compared to the inoculate culture used to infect the bladder cells. A total of 143 genes meeting the criteria of P-value ≤ 0.05 and $Log_2FC < -1.0$ were identified and are shown in Table 5.1. A large number of these genes were found to be involved in central carbon metabolism, amino acid biosynthesis and transport, flagella biosynthesis and motility, lipopolysaccharide (LPS) biosynthesis and iron homeostasis. During intracellular infection, UPEC aggregate into dense biofilm-like clusters within the cytoplasm and largely rely on the capacity for *de novo* biosynthesis of essential metabolites, amino acids and sugars (*28*). Correspondingly, the carbon metabolic genes *araA*, *phnA*, *zwf*, *pgm* and *galU* were all identified by TraDIS to be required for IBCs within the bladder (Table 5.1 and Supp. Data Table S3C). Interestingly, *pgm* (phosphoglucomutase) and *galU* (glucose-1-phosphate uridylyltransferase), both involved in UDP-glucose biosynthesis, were found to be important for coordinating *E. coli* cell size with nutrient availability (*148*).

The ability of UPEC to also exploit diverse amino acids and small peptides during bladder infection has long been considered an important adaptation (*98*). The utilisation of these resources by UPEC is supported by the identification of the amino acid biosynthesis and transport genes *hisF* (histidine biosynthesis), *purF* and *purD* (purine nucleotide biosynthesis), *aroK* (aromatic amino acid biosynthesis) and *lysP* (lysine transport) (Table 5.1 and Supp. Data Table S3C). Fig. 5.5A shows the insertion-site read frequency and distribution of *hisF* and surrounding genes; the read depth within the *hisF* gene is markedly reduced (Log_2FC -2.48) in the IBC sample (bottom panel) when compared to the Inoculate sample (top panel). Similarly, Fig. 5.5B shows the insertion-site read frequency and distribution of the *tol-pal* system; the genes *tolA*, *tolB*, *tolR*, *tolQ* and *pal* were all identified by TraDIS to be required for progression to IBCs (Table 5.1 and Supp. Data Table S3C).

Interestingly, the *tol-pal* system genes are known to control outer cell membrane integrity and may suggest that certain modifications are made to the cell membrane during IBC formation. The IBC begins as rod-shaped bacteria loosely organised within the host cytosol that multiply, producing progressively smaller coccoid daughter cells that are tightly assembled (*52*).

UPEC also express a variety of iron uptake and transport systems such as iron-chelating siderophores that play major roles during intracellular infection (*100*). The iron acquisition and homeostasis gene, *iroE* was identified in our study (Table 5.1). Additionally, within the IBC, flagella are known to facilitate weak surface contact inducing organised growth and subsequent motility (*186*). We identified the flagella motor switch complex protein FliN and the biosynthetic component FliR, as well as the sensory detection protein CheA, which transmits signals from the chemoreceptors to the flagella motors (Table 5.1).

Several genes already known to play key virulence roles within the IBC phase were also identified in our study. These included the well-characterised K1 capsule synthesizing gene *neuE* (*187*), the periplasmic prolyl chaperone SurA (*97*), the outer membrane protein OmpA (*188*) and various genes involved in LPS biosynthesis such as, *rfaF*, *rfaG* and *rfaP* (*189*) (Table 5.1 and Supp. Data Table S3).

Table 5.1 UTI89 mutants that have a reduced ability to survive to IBCs (24 h post-infection)

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C4063	yhjM	endoglucanase precursor	-8.36	2.48E-07	1.92E-04
UTI89_C0216	yaeD	hypothetical protein YaeD	-8.08	9.20E-06	2.28E-03
UTI89_C1432	galU	glucose-1-phosphate uridylyltransferase	-7.37	8.08E-09	2.11E-05
UTI89_C3001	ymfM	hypothetical protein	-7.28	3.49E-05	5.66E-03
UTI89_C0552	UTI89_C0552	hypothetical protein	-5.78	2.81E-07	1.92E-04
UTI89_C2018	yobD	hypothetical protein	-5.56	7.70E-06	2.19E-03
UTI89_C1392	ycgS	protein YcgS	-5.52	4.36E-05	6.46E-03
UTI89_C1951	ynjC	hypothetical ABC transporter permease protein ynjC	-5.41	2.92E-06	1.11E-03
UTI89_C0736	tolB	periplasmic protein (uptake of group A colicins)	-5.40	1.04E-06	5.07E-04
UTI89_C2031	prc	carboxy-terminal protease	-5.07	9.66E-08	1.10E-04
UTI89_C0507	fsr	fosmidomycin resistance protein	-5.03	3.81E-05	5.90E-03
UTI89_C0060	surA	peptidyl-prolyl cis-trans isomerase	-4.96	2.51E-05	4.76E-03
UTI89_C4165	rfaF	lipopolysaccharide core biosynthesis	-4.87	3.28E-06	1.12E-03
UTI89_C0738	pal	peptidoglycan-associated lipoprotein precursor	-4.67	2.81E-05	5.04E-03
UTI89_C0775	bioC	biotin synthesis protein BioC	-4.66	9.37E-06	2.28E-03
UTI89_C1677	UTI89_C1677	hypothetical protein	-4.56	1.23E-05	2.80E-03
UTI89_C2596	purF	amidophosphoribosyltransferase	-4.53	3.05E-05	5.20E-03
UTI89_C2768	amiA	probable N-acetylmuramoyl-L-alanine amidase	-4.40	8.81E-04	5.89E-02
UTI89_C3888	aroK	shikimate kinase I	-4.26	3.29E-04	2.91E-02
UTI89_C0773	bioB	biotin synthase	-4.25	3.87E-04	3.26E-02
UTI89_C0693	pgm	phosphoglucomutase	-4.01	2.57E-06	1.09E-03
UTI89_C2059	msbB	myristoyl acyltransferase	-3.87	4.56E-03	1.64E-01
UTI89_C0617	UTI89_C0617	hypothetical protein	-3.78	1.81E-02	3.48E-01
UTI89_C1022	ompA	outer membrane protein A precursor	-3.72	1.49E-05	3.17E-03
UTI89_C0892	UTI89_C0892	conserved hypothetical protein	-3.71	1.07E-04	1.30E-02
UTI89_C3378	gspK	hypothetical protein	-3.70	1.32E-04	1.56E-02
UTI89_C3666	degS	inner membrane serine protease DegS precursor	-3.68	1.14E-03	6.79E-02
UTI89_C0618	citF	citrate lyase alpha chain	-3.62	1.70E-02	3.35E-01
UTI89_C2954	smpB	SsrA-binding protein	-3.41	7.34E-04	5.11E-02
UTI89_C0067	araA	L-arabinose isomerase	-3.31	1.45E-02	3.16E-01
UTI89_C0181	glnD	[protein-PII] uridylyltransferase	-3.13	1.26E-03	7.28E-02
UTI89_C4231	yidG	hypothetical protein	-3.06	2.52E-03	1.13E-01
UTI89_C2430	lysP	lysine-specific permease	-3.03	7.68E-04	5.23E-02
UTI89_C0735	tolA	membrane spanning protein TolA	-2.94	2.12E-03	1.02E-01
UTI89_C1294	UTI89_C1294	DNA N-6-adenine-methyltransferase	-2.92	5.60E-03	1.87E-01
UTI89_C3369	neuE	NeuE protein	-2.88	4.25E-02	4.40E-01
UTI89_C2806	gcvR	glycine cleavage system transcriptional repressor	-2.88	3.46E-02	4.30E-01
UTI89_C0934	UTI89_C0934	putative phage tail protein	-2.78	3.84E-03	1.44E-01
UTI89_C3724	yrdA	protein YrdA	-2.77	3.79E-07	2.16E-04
UTI89_C1128	UTI89_C1128	hypothetical protein	-2.76	3.92E-04	3.26E-02
UTI89_C2202	nac	nitrogen assimilation regulatory protein nac	-2.68	3.61E-03	1.38E-01
UTI89_C2119	yecS	putative transport system permease protein	-2.66	1.38E-03	7.70E-02
UTI89_C3978	ftsE	cell division ATP-binding membraine protein ftsE	-2.65	9.29E-03	2.66E-01
UTI89_C4764	glyX	Trna-Gly	-2.61	1.91E-02	3.53E-01
UTI89_C1252	potC	spermidine/putrescine transport system permease	-2.58	1.46E-02	3.16E-01
UTI89_C0070	UTI89_C0070	hypothetical protein	-2.54	7.61E-05	9.61E-03
UTI89_C5044	yjiL	hypothetical protein YjiL	-2.51	1.08E-02	2.80E-01
UTI89_C4192	UTI89_C4192	hypothetical protein	-2.50	3.44E-02	4.30E-01
UTI89_C0787	UTI89_C0787	hypothetical protein	-2.50	2.59E-02	3.86E-01
UTI89_C2298	hisF	imidazole glycerol phosphate synthase subunit hisF	-2.48	1.85E-02	3.51E-01
UTI89_C2988	UTI89_C2988	hypothetical protein	-2.44	1.08E-02	2.80E-01
UTI89_C0547	ylbE	hypothetical protein YlbE	-2.44	4.01E-02	4.38E-01

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C4540	frwB	PTS system fructose-like IIB component 1	-2.43	1.02E-02	2.73E-01
UTI89_C0734	tolR	putative inner membrane protein TolR	-2.41	1.12E-02	2.84E-01
UTI89_C1066	yccJ	hypothetical protein	-2.40	5.91E-03	1.92E-01
UTI89_C3164	yqcD	conserved hypothetical protein	-2.39	2.63E-03	1.17E-01
UTI89_C5132	UTI89_C5132	putative tail component of prophage CP-933K	-2.36	3.72E-02	4.37E-01
UTI89_C0737	UTI89_C0737	hypothetical protein	-2.31	1.39E-02	3.10E-01
UTI89_C2055	zwf	glucose-6-phosphate dehydrogenase	-2.27	1.80E-02	3.48E-01
UTI89_C5153	deoA	thymidine phosphorylase	-2.24	2.61E-02	3.86E-01
UTI89_C1369	umuD	UmuD protein	-2.23	2.48E-03	1.13E-01
UTI89_C0298	UTI89_C0298	hypothetical protein	-2.21	1.22E-02	2.95E-01
UTI89_C4280	pstS	phosphate-binding periplasmic protein precursor	-2.21	6.44E-03	2.07E-01
UTI89_C4296	gidA	glucose-inhibited division protein A	-2.21	1.68E-02	3.33E-01
UTI89_C4152	grxC	glutaredoxin 3	-2.19	1.11E-03	6.79E-02
UTI89_C2472	ccmH	cytochrome c-type biogenesis protein CcmH precursor	-2.14	2.21E-03	1.05E-01
UTI89_C1105	UTI89_C1105	hypothetical protein	-2.13	1.33E-02	3.03E-01
UTI89_C0694	potE	putrescine transport protein	-2.10	4.89E-03	1.70E-01
UTI89_C2264	UTI89_C2264	ABC transporter, FecCD transport family	-2.10	2.67E-02	3.86E-01
UTI89_C2935	yfiL	hypothetical protein	-2.09	7.84E-03	2.33E-01
UTI89_C2891	lepA	GTP-binding elongation factor LepA	-2.06	4.94E-02	4.70E-01
UTI89_C3863	pabA	para-aminobenzoate synthase	-2.06	9.47E-03	2.68E-01
UTI89_C4433	yigZ	hypothetical protein	-2.05	5.64E-03	1.87E-01
UTI89_C4737	UTI89_C4737	hypothetical protein	-2.05	1.06E-02	2.79E-01
UTI89_C4689	phnN	ATP-binding component of phosphonate transport	-2.02	3.36E-03	1.35E-01
UTI89_C4544	frwD	PTS system fructose-like IIB component 2	-2.01	4.45E-02	4.52E-01
UTI89_P019	UTI89_P019	hypothetical protein	-1.98	2.86E-03	8.22E-02
UTI89_C3815	purD	Phosphoribosylamine--glycine ligase	-1.97	2.33E-02	3.72E-01
UTI89_C0257	UTI89_C0257	hypothetical protein	-1.96	1.98E-03	9.72E-02
UTI89_C2441	UTI89_C2441	hypothetical protein	-1.96	1.87E-02	3.51E-01
UTI89_C4609	ubiC	chorismate pyruvate lyase	-1.95	1.90E-02	3.53E-01
UTI89_C1119	iroE	IroE protein	-1.89	1.36E-03	7.70E-02
UTI89_C3121	cysC	adenylylsulfate kinase	-1.87	1.57E-02	3.25E-01
UTI89_C4941	UTI89_C4941	putative membrane protein	-1.86	3.87E-02	4.38E-01
UTI89_C3036	nrdF	ribonucleoside-diphosphate reductase 2 beta chain	-1.86	7.35E-05	9.61E-03
UTI89_C3160	yqcA	hypothetical protein YqcA	-1.86	3.92E-02	4.38E-01
UTI89_C2009	yeaB	hypothetical protein	-1.85	4.99E-02	4.70E-01
UTI89_C0235	aspV	tRNA-Asp	-1.84	4.81E-03	1.70E-01
UTI89_C2249	UTI89_C2249	hypothetical protein	-1.83	2.39E-02	3.74E-01
UTI89_C4151	secB	protein-export protein SecB	-1.81	1.00E-02	2.73E-01
UTI89_P035	UTI89_P035	hypothetical protein	-1.81	9.88E-03	2.27E-01
UTI89_C2146	fliN	flagellar motor switch protein FliN	-1.77	3.36E-02	4.24E-01
UTI89_C2150	fliR	flagellar biosynthetic protein FliR	-1.76	4.00E-02	4.38E-01
UTI89_C1621	UTI89_C1621	hypothetical protein	-1.74	3.26E-02	4.19E-01
UTI89_C1620	UTI89_C1620	hypothetical protein	-1.74	3.25E-02	4.19E-01
UTI89_C4150	UTI89_C4150	hypothetical protein	-1.70	1.28E-02	2.99E-01
UTI89_C0413	ykiA	hypothetical protein YkiA	-1.70	2.15E-02	3.61E-01
UTI89_C3172	UTI89_C3172	hypothetical protein	-1.67	3.06E-02	4.06E-01
UTI89_C4093	yiaE	putative dehydrogenase	-1.66	2.93E-02	4.06E-01
UTI89_C0529	tesA	acyl-CoA thioesterase I precursor	-1.65	2.89E-03	1.20E-01
UTI89_P089	UTI89_P089	hypothetical protein	-1.63	3.41E-02	3.56E-01
UTI89_C2275	yafX2	hypothetical protein YafX	-1.62	4.35E-02	4.46E-01
UTI89_C3981	yhhL	conserved hypothetical protein	-1.62	1.53E-03	8.26E-02
UTI89_C4735	dcuA	DcuA dicarboxylate transporter	-1.61	1.52E-02	3.20E-01
UTI89_C2486	napF	ferredoxin-type protein: electron transfer	-1.59	3.99E-02	4.38E-01
UTI89_C4621	UTI89_C4621	putative transmembrane transport protein	-1.58	2.32E-02	3.72E-01
UTI89_C4525	metL	Aspartate kinase II subunit	-1.56	1.84E-04	2.09E-02
UTI89_C0219	ileV	tRNA-lle	-1.56	4.10E-02	4.38E-01

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C4153	UTI89_C4153	hypothetical protein	-1.54	9.25E-04	6.06E-02
UTI89_C1317	UTI89_C1317	putative capsid protein gp7 of prophage	-1.50	3.96E-02	4.38E-01
UTI89_C0852	grxA	glutaredoxin 1	-1.49	6.80E-03	2.17E-01
UTI89_C2448	yeiP	putative elongation factor YeiP	-1.48	1.44E-02	3.16E-01
UTI89_C0406	yaiI	hypothetical protein	-1.45	1.25E-02	2.99E-01
UTI89_C4174	rfaG	lipopolysaccharide core biosynthesis	-1.42	4.89E-02	4.70E-01
UTI89_C1016	pqiB	paraquat-inducible protein B	-1.41	1.98E-02	3.56E-01
UTI89_C1159	csgF	curli production assembly/transport	-1.40	7.85E-03	2.33E-01
UTI89_C4173	rfaP	lipopolysaccharide core biosynthesis protein rfaP	-1.38	4.56E-05	6.48E-03
UTI89_C3918	malP	maltodextrin phosphorylase	-1.37	2.40E-04	2.48E-02
UTI89_P142	finO	fertility inhibition protein	-1.37	4.68E-06	5.38E-04
UTI89_C3170	fucO	lactaldehyde reductase	-1.34	3.08E-02	4.07E-01
UTI89_C3455	mdaB	modulator of drug activity B	-1.33	4.27E-02	4.40E-01
UTI89_C2128	amyA	cytoplasmic alpha-amylase	-1.30	2.88E-03	1.20E-01
UTI89_C0258	UTI89_C0258	hypothetical protein	-1.28	2.17E-04	2.31E-02
UTI89_C1438	adhE	CoA-linked acetaldehyde dehydrogenase	-1.27	7.85E-03	2.33E-01
UTI89_C4771	miaA	tRNA-adenosine transferase	-1.25	4.94E-02	4.70E-01
UTI89_C3682	UTI89_C3682	hypothetical protein	-1.25	3.17E-02	4.12E-01
UTI89_C2384	yehC	putative periplasmic chaperone YehC precursor	-1.23	1.91E-03	9.58E-02
UTI89_P058	rsvB	resolvase	-1.23	7.30E-04	2.80E-02
UTI89_C2989	UTI89_C2989	enterobacteria phage Sf6 gene 63 protein	-1.20	3.33E-04	2.91E-02
UTI89_C3471	ygiB	hypothetical protein	-1.19	4.27E-02	4.40E-01
UTI89_C3258	ygeZ	hypothetical hydrolase YgeZ	-1.19	1.68E-03	8.66E-02
UTI89_C5120	UTI89_C5120	putative protease/scaffold protein	-1.18	5.74E-05	7.82E-03
UTI89_C1847	sodB	superoxide dismutase, iron	-1.17	3.55E-02	4.33E-01
UTI89_C2386	yehE	hypothetical protein YehE precursor	-1.14	3.65E-03	1.38E-01
UTI89_C0778	ybhK	hypothetical protein YbhK	-1.14	3.05E-02	4.06E-01
UTI89_C1699	bdm	hypothetical protein	-1.13	2.39E-02	3.74E-01
UTI89_C0691	ybfF	putative esterase/lipase YbfF	-1.13	3.69E-02	4.37E-01
UTI89_C0733	tolQ	TolQ protein	-1.13	2.05E-02	3.59E-01
UTI89_C2091	cheA	chemotaxis protein CheA	-1.09	3.43E-02	4.29E-01
UTI89_C5158	smp	Smp protein precursor	-1.07	1.64E-02	3.32E-01
UTI89_C4213	UTI89_C4213	hypothetical protein	-1.06	1.98E-02	3.56E-01
UTI89_C0136	cueO	blue copper oxidase CueO precursor	-1.05	3.59E-02	4.33E-01
UTI89_C1117	focX	putative regulatory protein	-1.02	3.69E-02	4.37E-01

**Hypothetical or proteins with no characterised homologs are listed with the gene ID.*

[#]Q-value as determined by TraDIS analysis. The Q-value is the P-value that has been adjusted for the false discovery rate (FDR) for each gene hit.

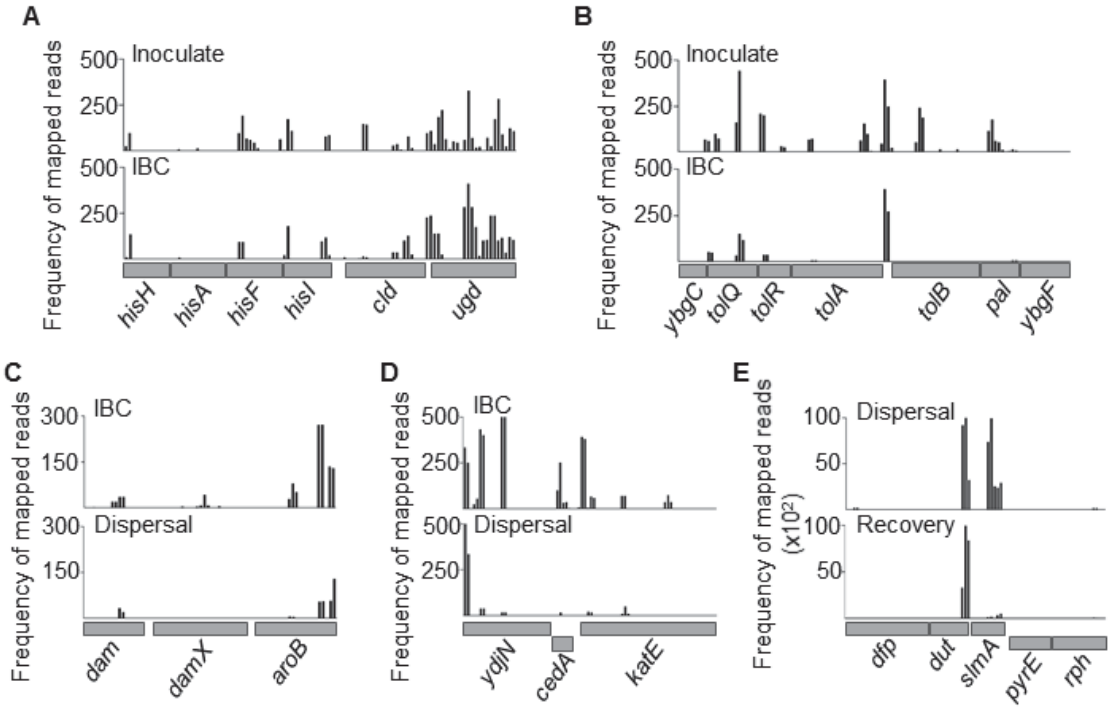


Figure 5.5 Identification by TraDIS of UPEC genes required for bladder infection

Frequency and distribution of mapped-reads from the chromosome of UTI89 harvested from bladder cell infections. Each distribution plot has a bin of 50 bp and corresponds to selected regions in the genome. The position of the annotated genes relative to the mapped reads is indicated below the distribution plot. (A-B) Depicts the mapped reads from the Inoculate (top panel) and IBC (bottom panel) samples encompassing genes from the hisF operon (A) and tol-pal system (B) that were identified, which both demonstrate a decrease in the read depth in the IBC sample. (C-D) Depicts the mapped reads from the IBC (top panel) and Dispersal (bottom panel) samples encompassing genes from the damX operon (A) and cedA region (B). Both genes (damX and cedA) show a reduction in read depth in the Dispersal sample when compared to the IBC sample (Table 5.2). (E) Depicts the mapped reads from the Dispersal (top panel) and Recovery (bottom panel) samples encompassing the slmA gene that was identified (Table 5.3). SlmA can be seen with a decrease in the read depth in the Recovery when compared to the Dispersal sample. The frequency and distribution of the mapped-reads are from the first of two biological replicates (sequencing data A; Supp. Data Table S3).

5.2.2.2 The genes required for UPEC dispersal and filamentation

The IBC continues to grow over time, however at some advanced point the coccoid bacteria differentiate into highly-motile and filamentous cells, which dissociate from the IBC and disperse out from the host cells into the bladder lumen (*17*). The filamentation process in UPEC involves ongoing bacterial cell growth, but a block to the normal cell division process (*17, 26*). The exact mechanism of how this is regulated is yet to be fully understood. In order to identify genes involved at this phase, we harvested UTI89 bacteria released from the bladder cells by collecting the supernatant culture in short 20 min windows between 34-44 h post-infection. During this period (and 10 h prior), the cell culture infection model was exposed to human urine (ρ = 1.026 g/mL) with gentle oscillatory shaking. TraDIS sequencing data obtained from analysis of this dispersal sample of infection was compared to the IBC culture sample, to determine those genes that were specifically important for the IBC to dispersal transition.

A total of 333 stringent genes (P-value ≤ 0.05 and $Log_2FC < -1.0$) were identified and are shown in Table 5.2. As anticipated, a large number of these genes were found to be involved in the regulation of cell division. Among the strongest cell division genes identified in UTI89 were *cedA*, *ytfB*, *dedD* and the previously identified gene *damX (56)* (Table 5.2). Along with *damX* (Figure 5.5C), *dedD* is another gene that encodes a protein with a SPOR domain, which has been reported to localise to the division septal ring to facilitate cell membrane integrity (*143, 190*). Interestingly, one of the filamentous, temperature-sensitive (*fts*) mutants, *ftsX* was identified in our study (Log_2FC -4.95, Table 5.2). FtsX has previously been shown to be a non-essential division protein, which localises to the septal ring (*191*). As mentioned, the newly discovered cell division gene, *ytfB* was identified in the dispersal phase with a strong log_2 fold-change decrease of 8.13 (Table 5.2). Additionally, the *cedA* gene has a role in cell division regulation and our data revealed a log_2 fold-change decrease of 4.38 (Figure 5.5D). We also identified genes encoding membrane-associated proteins, such as *yedZ* and UTI89_C2133, and others involved in peptidoglycan cell wall structure, including *dacA* and *mrcB* (Table 5.2 and Supp. Data Table S3D).

Some of the strongest genes identified were located in several bacteriophages, prophages or prophage remnants (Table 5.2 and Supp. Data Table S3). These include UTI89_C1149 (prophage CP-933L), UTI89_C0943 (prophage CP-933R), *ybcQ1* and *ybcS1* (prophage DLP12). There were also numerous genes found to be involved in DNA replication. We identified the *iciA* gene (log_2FC -2.65) that encodes the protein inhibitor of *oriC*-initiated replication, which has also been reported to be nucleoid-associated and acts as a transcriptional regulator of genes involved in DNA replication, such as *dnaA* (*192*). We also identified the negative regulator of DNA replication initiation, SeqA, which ensures that the initiation of replication occurs only once per chromosome every cell division cycle.

Interestingly, a large number of genes were also identified in cell stress and signaling response pathways (Table 5.2 and Supp. Data Table S3D). In order to colonise the intracellular bladder cell environment and disperse out into the bladder lumen, UPEC must be able to sense and respond to certain stresses that may compromise the cell envelope. In our study we identified genes from four of the cell envelope stress response systems. The genes *envR* and *ompR* were identified (3.56 and 4.0 log_2 fold-change decrease, respectively) and are involved in the well-characterised EnvZ/OmpR two-component system mediating signal transduction and response to changes in osmotic stress. We also identified the *cpxR* gene (log_{2FC} -3.02) involved in the Cpx cell envelope stress response system, which regulates expression of outer membrane porins OmpF and OmpC to changes in osmolarity and nutrient availability. We identified the Rcs phosphorelay gene, *rscD* and the *baeS* gene involved in the BaeSR two-component system (Table 5.2). Both these systems are induced by osmotic shock, which regulates the expression of genes involved in membrane integrity, motility and production of the K1 capsule.

Due to the dynamic nature of the infection cascade where many individual events are predicted to overlap, genes within the functional classes of carbon metabolism, amino acid catabolism and iron acquisition were also identified in the Dispersal and Non-dispersed samples (Table 5.2 and Supp. Data Table S3D-E). Those mostly strongly implicated include genes such as *aceE*, a component of the pyruvate dehydrogenase complex (log_2FC

-4.02), *pyrE* involved in pyrimidine biosynthesis (log_2FC -4.62) and the heme binding and transport protein, ChuU (log_2FC -5.70).

Figure 5.2 UTI89 mutants that have a reduced ability to survive to the dispersal phase

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_P002	repB	replication protein RepB-like protein	-9.80	5.72E-07	2.59E-05
UTI89_C4825	ytfK	hypothetical protein YtfK	-9.65	3.20E-07	4.02E-05
UTI89_C2550	elaB	hypothetical protein	-9.34	4.43E-08	1.26E-05
UTI89_C1318	UTI89_C1318	hypothetical protein	-9.32	6.88E-08	1.50E-05
UTI89_C1655	yncJ	hypothetical protein YncJ precursor	-9.14	1.06E-06	7.93E-05
UTI89_C1425	UTI89_C1425	Tpn1330-like protein	-8.48	2.59E-09	1.69E-06
UTI89_C1261	ycfC	hypothetical protein YcfC	-8.24	3.31E-06	1.60E-04
UTI89_C4815	ytfB	hypothetical protein YtfB	-8.13	1.14E-07	1.97E-05
UTI89_C1171	UTI89_C1171	hypothetical protein	-8.13	5.31E-08	1.26E-05
UTI89_C2898	UTI89_C2898	ATP-dependent RNA helicase	-7.73	4.65E-07	5.42E-05
UTI89_C4595	UTI89_C4595	hypothetical protein	-7.66	1.14E-07	1.97E-05
UTI89_C0608	ahpC	alkyl hydroperoxide reductase, C22 subunit	-7.47	7.60E-07	6.44E-05
UTI89_C0519	UTI89_C0519	hypothetical protein	-7.39	2.89E-08	1.08E-05
UTI89_C2599	dedD	putative lipoprotein	-7.38	3.24E-07	4.02E-05
UTI89_C3152	UTI89_C3152	hypothetical protein	-7.35	4.91E-08	1.26E-05
UTI89_C2105	yecR	hypothetical protein YecR	-7.08	1.38E-05	4.19E-04
UTI89_C0471	ybaW	hypothetical protein YbaW	-6.76	3.43E-05	8.22E-04
UTI89_C3491	ygiM	hypothetical protein	-6.75	8.75E-08	1.76E-05
UTI89_C2458	yejG	hypothetical protein	-6.75	2.01E-06	1.15E-04
UTI89_C1251	potD	spermidine/putrescine-binding periplasmic protein	-6.66	3.61E-06	1.68E-04
UTI89_C0705	ybgA	hypothetical protein YbgA	-6.61	6.47E-06	2.58E-04
UTI89_C4746	ecnA	antitoxin of osmotically regulated toxin-antitoxin	-6.61	6.12E-07	5.91E-05
UTI89_C4989	ykfF	hypothetical protein YkfF	-6.61	7.74E-06	2.88E-04
UTI89_C4511	glpF	glycerol MIP channel	-6.53	9.43E-07	7.45E-05
UTI89_C3635	yrbK	hypothetical protein	-6.49	2.06E-06	1.15E-04
UTI89_C1149	UTI89_C1149	unknown protein encoded within prophage CP-933L	-6.47	1.30E-06	8.93E-05
UTI89_C1527	yciF	putative structural proteins	-6.40	2.10E-06	1.15E-04
UTI89_C3017	gabT	4-aminobutyrate aminotransferase	-6.36	3.46E-06	1.64E-04
UTI89_C0943	UTI89_C0943	prophage CP-933R unknown protein	-6.35	5.20E-07	5.42E-05
UTI89_C1035	UTI89_C1035	putative acylphosphatase	-6.28	2.52E-06	1.30E-04
UTI89_C4771	miaA	tRNA-adenosine transferase	-6.27	2.19E-05	5.76E-04
UTI89_C0050	fixX	putative ferredoxin	-6.24	2.21E-07	3.24E-05
UTI89_C2307	UTI89_C2307	bacteriophage HK620 O-antigen modification protein	-6.22	1.61E-05	4.62E-04
UTI89_C4734	cutA	divalent cation tolerance protein	-6.20	1.80E-04	2.97E-03
UTI89_C3576	yraO	hypothetical protein YraO	-6.14	1.01E-06	7.78E-05
UTI89_C2777	eutG	ethanolamine utilization	-6.08	2.11E-05	5.66E-04
UTI89_C4481	UTI89_C4481	hypothetical protein	-6.07	4.18E-06	1.82E-04
UTI89_C1604	ydaA	hypothetical protein	-6.02	1.76E-05	4.90E-04
UTI89_C0692	seqA	negative modulator of initiation of replication	-6.02	6.04E-07	5.91E-05
UTI89_C3285	UTI89_C3285	putative oxidoreductase	-6.01	7.36E-06	2.81E-04
UTI89_C2089	tar	methyl-accepting chemotaxis protein II	-5.94	1.41E-05	4.21E-04
UTI89_C3247	yqeF	probable acetyl-CoA acetyltransferase	-5.91	8.15E-06	2.99E-04
UTI89_C0114	hofB	transport protein HofB	-5.90	5.52E-05	1.16E-03
UTI89_C2435	yeiI	hypothetical sugar kinase YeiI	-5.83	1.12E-05	3.61E-04
UTI89_C3273	UTI89_C3273	hypothetical protein	-5.82	1.06E-04	1.93E-03
UTI89_C0214	metI	D-methionine transport system permease MetI	-5.81	2.99E-06	1.50E-04
UTI89_C3841	tufA	translation elongation factor EF-Tu	-5.81	8.90E-06	3.01E-04
UTI89_C2690	UTI89_C2690	hypothetical protein	-5.78	5.87E-06	2.39E-04
UTI89_C4037	chuU	permease of iron compound ABC transport system	-5.70	1.67E-06	1.01E-04
UTI89_C5014	fimD	outer membrane usher protein FimD precursor	-5.67	6.04E-09	3.15E-06
UTI89_C0920	pin	putative DNA-invertase	-5.65	3.28E-06	1.60E-04
UTI89_C4188	yicC	putative alpha helix protein	-5.63	1.67E-05	4.72E-04
UTI89_C2371	UTI89_C2371	hypothetical protein	-5.55	1.06E-05	3.45E-04
UTI89_C4237	ibpB	16 kDa heat shock protein B	-5.54	5.91E-05	1.21E-03

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C2816	purN	phosphoribosylglycinamide formyltransferase	-5.50	5.83E-05	1.21E-03
UTI89_C3594	pnp	polyribonucleotide nucleotidyltransferase	-5.49	3.97E-06	1.79E-04
UTI89_C2101	araF	L-arabinose-binding periplasmic protein precursor	-5.48	9.81E-06	3.24E-04
UTI89_C5017	fimH	type 1 fimbiral adhesin FimH	-5.46	9.77E-06	3.24E-04
UTI89_C4256	UTI89_C4256	hypothetical protein	-5.46	3.91E-04	5.20E-03
UTI89_C1502	UTI89_C1502	putative capsid protein of prophage	-5.43	2.56E-05	6.55E-04
UTI89_C2657	UTI89_C2657	hypothetical protein	-5.42	6.57E-07	6.12E-05
UTI89_C2598	UTI89_C2598	conserved hypothetical protein	-5.41	3.54E-05	8.39E-04
UTI89_C5013	fimC	periplasmic chaperone protein FimC precursor	-5.36	4.99E-07	5.42E-05
UTI89_C1530	trpB	tryptophan synthase beta chain	-5.35	1.20E-05	3.77E-04
UTI89_C1300	ybcQ1	lambdoid prophage DLP12 antitermination protein	-5.32	3.16E-04	4.46E-03
UTI89_C3130	ygcN	e- transfer flavoprotein-quinone oxidoreductase	-5.30	1.61E-04	2.77E-03
UTI89_C2580	UTI89_C2580	hypothetical protein	-5.23	9.24E-05	1.73E-03
UTI89_C3593	UTI89_C3593	hypothetical protein	-5.20	4.09E-06	1.81E-04
UTI89_C2549	menF	menaquinone-specific isochorismate synthase	-5.18	1.76E-04	2.97E-03
UTI89_C0634	dacA	penicillin-binding protein 5	-5.11	2.86E-05	7.02E-04
UTI89_C5011	fimA	type 1 fimbriae major subunit FimA	-5.08	3.63E-05	8.53E-04
UTI89_C1513	UTI89_C1513	minor tail protein	-5.05	1.23E-05	3.79E-04
UTI89_C1512	UTI89_C1512	hypothetical protein	-5.05	1.23E-05	3.79E-04
UTI89_P037	UTI89_P037	conserved hypothetical protein	-5.01	4.98E-05	1.10E-03
UTI89_C0748	pnuC	PnuC NMN transporter	-4.96	3.78E-05	8.64E-04
UTI89_C0615	citG	2-(5''-triphosphoribosyl)-3'-dephosphocoenzyme-A	-4.95	1.99E-04	3.22E-03
UTI89_C3977	ftsX	cell division membrane protein	-4.95	1.84E-04	3.00E-03
UTI89_C1315	UTI89_C1315	head-tail preconnector protein gp5 of bacteriophage	-4.94	5.56E-05	1.16E-03
UTI89_C2975	UTI89_C2975	hypothetical protein	-4.92	6.61E-05	1.35E-03
UTI89_C5055	yjiA	hypothetical protein YjiA	-4.90	3.63E-04	4.91E-03
UTI89_C4826	ytfL	putative transport protein	-4.87	1.97E-05	5.35E-04
UTI89_C3534	yqjB	hypothetical protein YqjB	-4.86	1.27E-04	2.23E-03
UTI89_C2845	pepB	aminopeptidase B	-4.83	4.21E-05	9.29E-04
UTI89_C4698	phnE	PhnE, subunit of alkylphosphonate ABC transporter	-4.80	1.48E-04	2.57E-03
UTI89_C3672	yhcQ	hypothetical protein YhcQ	-4.77	9.43E-05	1.76E-03
UTI89_C1643	ydcI	hypothetical transcriptional regulator YdcI	-4.76	3.77E-05	8.64E-04
UTI89_C0581	ybdK	hypothetical protein YbdK	-4.75	6.86E-05	1.37E-03
UTI89_C3271	ygfT	hypothetical protein YgfT	-4.70	2.71E-05	6.86E-04
UTI89_C0356	UTI89_C0356	putative ATP-binding component of transport system	-4.69	5.04E-05	1.09E-03
UTI89_C3886	damX	putative membrane protein	-4.69	3.07E-02	1.87E-01
UTI89_C0803	ybiB	hypothetical protein YbiB	-4.65	5.45E-05	1.16E-03
UTI89_C4379	corA	magnesium, cobalt and nickle transporter	-4.64	4.91E-04	6.21E-03
UTI89_C4186	pyrE	orotate phosphoribosyltransferase	-4.62	8.35E-05	1.58E-03
UTI89_C4589	UTI89_C4589	transcriptional regulator of sorbose uptake	-4.62	4.20E-05	9.29E-04
UTI89_C2505	atoE	short-chain fatty acids transporter	-4.53	2.53E-04	3.86E-03
UTI89_C0864	artM	arginine 3rd transport system permease protein	-4.46	6.34E-04	7.51E-03
UTI89_C0789	ybhN	hypothetical protein	-4.42	5.11E-05	1.10E-03
UTI89_C3466	icc	icc protein	-4.40	3.46E-04	4.83E-03
UTI89_C1290	UTI89_C1290	putative replication protein O of bacteriophage	-4.40	1.27E-03	1.30E-02
UTI89_C3323	tktA	transketolase 1	-4.39	1.80E-04	2.97E-03
UTI89_C4257	yidD	hypothetical protein	-4.38	8.83E-04	9.67E-03
UTI89_C1924	cedA	cell division regulatory protein CedA	-4.38	8.31E-04	9.29E-03
UTI89_C1384	modD	putative pyrophosphorylase ModD	-4.36	2.12E-04	3.41E-03
UTI89_C3567	agaS	tagatose-6-phosphate ketose/aldose isomerase	-4.34	2.22E-04	3.52E-03
UTI89_C3586	yhbV	hypothetical protein YhbV	-4.31	5.43E-04	6.65E-03
UTI89_P114	traR	conjugal transfer protein TraR	-4.29	1.18E-03	1.16E-02
UTI89_C1249	ycfZ	hypothetical protein YcfZ	-4.28	8.14E-04	9.19E-03
UTI89_C2466	yejM	putative sulfatase	-4.26	1.64E-03	1.56E-02
UTI89_C1396	UTI89_C1396	hypothetical protein	-4.25	2.54E-03	2.13E-02
UTI89_C1395	UTI89_C1395	hypothetical protein	-4.25	2.57E-03	2.15E-02

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C2837	yfgA	hypothetical protein YfgA	-4.23	2.98E-04	4.29E-03
UTI89_C4243	dgoA	bifunctional aldolase/galactonate dehydratase	-4.22	2.79E-04	4.11E-03
UTI89_C3281	ygfX	hypothetical protein	-4.18	4.93E-04	6.21E-03
UTI89_C1500	UTI89_C1500	putative capsid assembly protein of prophage	-4.15	5.22E-04	6.42E-03
UTI89_C1547	acnA	aconitate hydratase 1	-4.11	3.73E-04	5.01E-03
UTI89_C0127	aceE	pyruvate dehydrogenase E1 component	-4.02	3.52E-04	4.86E-03
UTI89_C4013	uspB	stationary phase ethanol tolerance protein	-4.00	3.99E-04	5.28E-03
UTI89_C3905	ompR	transcriptional regulatory protein OmpR	-4.00	2.47E-03	2.10E-02
UTI89_C1566	UTI89_C1566	putative membrane transport protein	-4.00	3.88E-03	2.87E-02
UTI89_C0238	UTI89_C0238	hypothetical protein	-3.95	3.04E-03	2.40E-02
UTI89_C1980	yeaH	hypothetical protein YeaH	-3.94	3.89E-04	5.20E-03
UTI89_C4409	UTI89_C4409	hypothetical protein	-3.91	2.55E-03	2.14E-02
UTI89_C3990	yhhT	hypothetical protein YhhT	-3.90	1.21E-03	1.25E-02
UTI89_C3698	UTI89_C3698	hypothetical protein	-3.89	2.58E-03	2.15E-02
UTI89_C4837	yjfG	putative ligase	-3.87	5.99E-04	7.23E-03
UTI89_P092	ssb	single strand DNA-binding protein	-3.78	9.30E-04	1.02E-02
UTI89_C2543	menE	O-succinylbenzoic acid--CoA ligase	-3.78	2.05E-03	1.81E-02
UTI89_C1207	flgK	hook-filament junction protein 1	-3.77	1.46E-03	1.44E-02
UTI89_C2443	fruK	fructose-1-phosphate kinase	-3.73	6.54E-04	7.65E-03
UTI89_C0074	thiQ	thiamine transport ATP-binding protein ThiQ	-3.71	6.45E-04	7.58E-03
UTI89_C1082	ycdB	hypothetical protein YcdB precursor	-3.70	2.76E-03	2.27E-02
UTI89_C0284	proA	gamma-glutamyl phosphate reductase	-3.70	1.44E-02	7.65E-02
UTI89_C3731	def	peptide deformylase	-3.69	3.75E-03	2.80E-02
UTI89_C0850	ybjL	putative transport protein	-3.68	4.93E-07	5.42E-05
UTI89_C4616	yjbK	putative zinc uptake regulation protein	-3.68	5.28E-03	3.64E-02
UTI89_C0541	ybbY	putative purine permease YbbY	-3.66	3.05E-04	4.35E-03
UTI89_C4447	dsbA	disulfide oxidoreductase	-3.65	6.18E-03	4.08E-02
UTI89_C3178	ygdE	hypothetical protein YgdE	-3.57	1.26E-03	1.29E-02
UTI89_C3904	envZ	osmolarity sensor protein EnvZ	-3.56	3.05E-04	4.35E-03
UTI89_C2738	xapB	xanthosine permease	-3.55	9.76E-03	5.77E-02
UTI89_C1959	selD	selenophosphate synthase	-3.55	3.27E-03	2.53E-02
UTI89_C2339	UTI89_C2339	hypothetical protein	-3.53	9.06E-03	5.48E-02
UTI89_C0075	thiP	thiamine transport system permease protein ThiP	-3.52	2.95E-04	4.28E-03
UTI89_C3631	yrbF	ATP-binding component of a transport system	-3.49	2.38E-04	3.74E-03
UTI89_C2188	fyuA	putative pesticin receptor precursor	-3.49	1.70E-03	1.58E-02
UTI89_C1437	UTI89_C1437	transposase InsG for insertion sequence element IS4	-3.48	3.63E-03	2.74E-02
UTI89_C3412	hybG	hydrogenase-2 operon protein HybG	-3.45	1.93E-02	9.08E-02
UTI89_C2171	yedZ	putative inner membrane protein YedZ	-3.45	1.88E-02	8.88E-02
UTI89_C1281	UTI89_C1281	bacteriophage lambda ssDNA binding protein	-3.45	3.66E-03	2.75E-02
UTI89_C4114	avtA	valine--pyruvate aminotransferase	-3.44	3.53E-03	2.69E-02
UTI89_C3926	glpD	aerobic glycerol-3-phosphate dehydrogenase	-3.37	8.38E-03	5.16E-02
UTI89_C1546	UTI89_C1546	conserved hypothetical protein	-3.36	3.47E-03	2.65E-02
UTI89_C5037	yjiE	putative transcriptional regulator	-3.33	4.62E-04	5.93E-03
UTI89_C0859	potH	putrescine transport protein	-3.33	5.92E-03	3.95E-02
UTI89_C0473	ybaE	hypothetical protein YbaE	-3.30	2.71E-03	2.23E-02
UTI89_C2353	baeS	sensor protein BaeS	-3.28	7.44E-03	4.66E-02
UTI89_C2641	UTI89_C2641	hypothetical protein	-3.26	1.02E-04	1.86E-03
UTI89_C4138	mtlD	mannitol-1-phosphate 5-dehydrogenase	-3.25	5.07E-03	3.52E-02
UTI89_C2415	yohK	putative transporter	-3.24	3.24E-02	1.28E-01
UTI89_C2716	ypdE	hypothetical protein YpdE	-3.24	8.50E-03	5.22E-02
UTI89_C2060	yebA	conserved hypothetical protein	-3.23	9.97E-03	5.84E-02
UTI89_C2788	talA	transaldolase A	-3.23	6.57E-03	4.25E-02
UTI89_C1731	UTI89_C1731	putative phage-related membrane protein	-3.22	1.11E-02	6.32E-02
UTI89_C5090	UTI89_C5090	hypothetical protein	-3.22	5.93E-03	3.95E-02
UTI89_C0808	ybiL	probable tonB-dependent receptor YbiL precursor	-3.20	3.88E-03	2.87E-02
UTI89_C0648	gltL	binding protein of glutamate/aspartate transport	-3.20	8.61E-03	5.26E-02

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C3433	yqhG	hypothetical protein	-3.19	1.46E-02	7.69E-02
UTI89_C1120	iroD	IroD protein	-3.13	1.60E-02	8.17E-02
UTI89_C1875	sufB	component of SufB-SufC-SufD cysteine desulfurase	-3.11	1.83E-02	8.75E-02
UTI89_C1535	yciO	hypothetical protein	-3.10	1.89E-02	8.91E-02
UTI89_C0347	UTI89_C0347	hypothetical protein	-3.09	9.33E-03	5.60E-02
UTI89_C2453	rtn	hypothetical protein	-3.07	3.33E-02	1.31E-01
UTI89_C2133	UTI89_C2133	outer membrane porin protein NmpC precursor	-3.07	2.02E-03	1.79E-02
UTI89_C0292	UTI89_C0292	putative transcription regulator	-3.06	3.65E-02	1.39E-01
UTI89_C1270	UTI89_C1270	hypothetical protein	-3.06	3.81E-03	2.83E-02
UTI89_C1271	UTI89_C1271	hypothetical protein	-3.05	1.39E-02	7.47E-02
UTI89_C2529	yfaW	hypothetical protein YfaW	-3.04	2.68E-03	2.22E-02
UTI89_C3880	yhfY	hypothetical protein YhfY	-3.03	4.93E-03	3.47E-02
UTI89_C4496	cpxR	transcriptional regulatory protein CpxR	-3.02	1.81E-03	1.64E-02
UTI89_C4856	UTI89_C4856	carbamate kinase	-3.01	9.16E-04	9.91E-03
UTI89_C1422	narH	respiratory nitrate reductase 1 beta chain	-3.01	1.00E-02	5.86E-02
UTI89_P109	traK	F pilus assembly protein TraK	-2.99	5.93E-03	2.90E-02
UTI89_C4030	UTI89_C4030	hypothetical protein	-2.99	2.36E-02	1.05E-01
UTI89_C3859	yhfA	hypothetical protein	-2.97	2.21E-02	1.00E-01
UTI89_C0501	htpG	chaperone Hsp90, heat shock protein C 62.5	-2.94	3.91E-03	2.88E-02
UTI89_P051	UTI89_P051	hypothetical protein	-2.94	3.90E-02	1.15E-01
UTI89_C3129	ygcM	putative 6-pyruvoyl tetrahydrobiopterin synthase	-2.93	1.79E-02	8.70E-02
UTI89_P112	trbD	hypothetical protein	-2.90	1.86E-02	6.83E-02
UTI89_C4461	yihW	hypothetical transcriptional regulator YihW	-2.89	2.04E-02	9.38E-02
UTI89_C2604	usg	USG-1 protein	-2.88	1.62E-02	8.22E-02
UTI89_C2394	yehQ	hypothetical protein YehQ	-2.87	4.12E-03	3.01E-02
UTI89_C0237	UTI89_C0237	hypothetical protein	-2.87	1.08E-03	1.14E-02
UTI89_C0483	ybaA	hypothetical protein YbaA	-2.85	3.03E-02	1.23E-01
UTI89_C4298	asnC	regulator for asnA, asnC and GidA	-2.84	1.54E-02	7.97E-02
UTI89_C4310	yieP	hypothetical transcriptional regulator YieP	-2.84	2.83E-02	1.18E-01
UTI89_C2219	UTI89_C2219	hypothetical protein	-2.84	1.98E-02	9.20E-02
UTI89_C3913	bioH	carboxylesterase	-2.83	1.79E-02	8.70E-02
UTI89_C0069	araC	transcriptional regulator AraC	-2.82	6.46E-04	7.58E-03
UTI89_C0913	UTI89_C0913	putative bacteriophage late gene regulator	-2.81	3.62E-02	1.38E-01
UTI89_C1601	mppA	periplasmic murein peptide-binding protein precursor	-2.80	2.46E-02	1.08E-01
UTI89_C2756	cysW	ABC-type sulfate transport system	-2.80	1.27E-02	7.00E-02
UTI89_C1835	ydhJ	hypothetical protein YdhJ	-2.79	2.90E-02	1.20E-01
UTI89_C0695	speF	ornithine decarboxylase, inducible	-2.77	6.45E-03	4.20E-02
UTI89_C5100	UTI89_C5100	putative replication protein	-2.76	2.07E-02	9.46E-02
UTI89_C3157	gudD	glucarate dehydratase	-2.75	1.31E-02	7.21E-02
UTI89_C1303	ybcS1	probable lysozyme from lambdoid prophage DLP12	-2.74	3.10E-02	1.25E-01
UTI89_C5087	UTI89_C5087	hypothetical protein	-2.73	2.87E-02	1.19E-01
UTI89_C4715	fumB	fumarate hydratase class I, anaerobic	-2.73	1.82E-02	8.72E-02
UTI89_C2612	yfcA	putative structural protein	-2.72	4.26E-02	1.54E-01
UTI89_C3400	UTI89_C3400	hypothetical protein	-2.71	4.89E-03	3.47E-02
UTI89_C0881	ybjX	hypothetical protein YbjX	-2.70	2.30E-02	1.03E-01
UTI89_C0964	UTI89_C0964	hypothetical protein	-2.70	1.96E-02	9.14E-02
UTI89_C3139	ygcE	hypothetical sugar kinase YgcE	-2.69	1.57E-03	1.50E-02
UTI89_C1742	yneG	hypothetical protein YneG	-2.68	2.75E-02	1.16E-01
UTI89_C3133	ygcQ	putative flavoprotein	-2.68	3.62E-02	1.38E-01
UTI89_C1786	UTI89_C1786	conserved hypothetical protein	-2.68	1.85E-02	8.78E-02
UTI89_C4845	treR	trehalose operon repressor	-2.68	2.06E-02	9.41E-02
UTI89_C3048	UTI89_C3048	hypothetical protein	-2.67	2.33E-02	1.04E-01
UTI89_C3711	yhdX	amino-acid ABC transporter permease protein yhdX	-2.67	1.45E-02	7.65E-02
UTI89_C1670	yncD	probable tonB-dependent receptor YncD precursor	-2.66	1.35E-02	7.33E-02
UTI89_C3302	iciA	replication initiation inhibitor	-2.65	1.14E-02	6.44E-02
UTI89_C3729	smg	hypothetical protein	-2.65	1.95E-02	9.12E-02

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C3961	livF	ATP-binding component of leucine transport	-2.63	1.44E-03	1.43E-02
UTI89_C0314	ykgM	50S ribosomal protein L31 type B-1	-2.62	8.55E-03	5.23E-02
UTI89_C3106	ygbN	hypothetical permease YgbN	-2.61	2.75E-02	1.16E-01
UTI89_C0304	ykgJ	putative ferredoxin	-2.60	3.81E-02	1.43E-01
UTI89_C4702	phnA	hypothetical protein PhnA	-2.60	4.67E-02	1.64E-01
UTI89_C0133	speD	S-adenosylmethionine decarboxylase proenzyme	-2.59	3.02E-02	1.23E-01
UTI89_C2696	UTI89_C2696	hypothetical protein	-2.57	8.37E-03	5.16E-02
UTI89_C3424	yghA	hypothetical oxidoreductase YghA	-2.57	1.15E-02	6.46E-02
UTI89_C3050	gshA	gamma-glutamate-cysteine ligase	-2.55	4.82E-02	1.68E-01
UTI89_C5094	UTI89_C5094	hypothetical protein	-2.54	1.63E-02	8.25E-02
UTI89_C4172	rafI	lipopolysaccharide 1,3-galactosyltransferase	-2.54	1.86E-03	1.67E-02
UTI89_C3530	uxaC	uronate isomerase	-2.54	1.95E-02	9.12E-02
UTI89_C3590	deaD	Cold-shock DEAD-box protein A	-2.52	1.08E-02	6.18E-02
UTI89_C0353	yahJ	hypothetical protein YahJ	-2.48	1.54E-02	7.97E-02
UTI89_C0372	lacI	lac operon repressor	-2.48	3.86E-02	1.44E-01
UTI89_C2049	yebE	hypothetical protein YebE	-2.47	1.70E-02	8.47E-02
UTI89_C3469	tolC	outer membrane protein TolC precursor	-2.46	1.32E-02	7.27E-02
UTI89_C5144	prfC	peptide chain release factor RF-3	-2.44	4.98E-04	6.24E-03
UTI89_C0164	mrcB	penicillin-binding protein 1B	-2.43	2.92E-03	2.34E-02
UTI89_C0177	degP	periplasmic serine protease DegP	-2.42	1.41E-02	7.54E-02
UTI89_C3920	rtcA	RNA 3'-terminal phosphate cyclase, chain B	-2.41	2.89E-02	1.19E-01
UTI89_C0802	dinG	probable ATP-dependent helicase DinG	-2.40	3.99E-02	1.48E-01
UTI89_C1098	UTI89_C1098	hypothetical protein	-2.40	4.84E-02	1.68E-01
UTI89_C2093	motA	chemotaxis MotA protein	-2.39	4.44E-02	1.59E-01
UTI89_C0033	UTI89_C0033	hypothetical protein	-2.39	2.40E-02	1.06E-01
UTI89_C4597	yjbH	conserved hypothetical protein	-2.38	9.53E-03	5.65E-02
UTI89_C3123	cysD	sulfate adenylyltransferase subunit 2	-2.38	1.73E-02	8.57E-02
UTI89_C2221	UTI89_C2221	putative polyketide synthase	-2.38	2.99E-02	1.22E-01
UTI89_C4425	tatC	sec-independent twin-arginine translocase subunit	-2.34	8.75E-04	9.63E-03
UTI89_C3510	aer	aerotaxis sensor receptor, flavoprotein	-2.33	3.48E-02	1.35E-01
UTI89_C3493	bacA	bacitracin resistance protein BacA	-2.32	2.11E-02	9.62E-02
UTI89_C2165	UTI89_C2165	outer membrane protein N precursor	-2.29	3.36E-02	1.32E-01
UTI89_C3848	yheO	conserved hypothetical protein	-2.26	4.05E-02	1.50E-01
UTI89_C0508	ushA	UDP-sugar hydrolase (5'-nucleotidase)	-2.26	4.82E-02	1.68E-01
UTI89_C4963	UTI89_C4963	hypothetical protein	-2.24	3.53E-02	1.36E-01
UTI89_C1090	UTI89_C1090	hypothetical protein	-2.24	3.56E-02	1.37E-01
UTI89_P067	UTI89_P067	transposase	-2.23	1.50E-03	1.16E-02
UTI89_C0112	guaC	GMP reductase	-2.23	3.38E-02	1.33E-01
UTI89_C0962	UTI89_C0962	unknown protein encoded by prophage CP-933T	-2.22	3.72E-02	1.40E-01
UTI89_C1895	ppsA	phosphoenolpyruvate synthase	-2.21	2.59E-02	1.12E-01
UTI89_P044	UTI89_P044	IS2 OrfA	-2.20	4.43E-02	1.26E-01
UTI89_C2553	yfbK	hypothetical protein YfbK	-2.19	1.84E-02	8.78E-02
UTI89_C3250	ygeR	hypothetical lipoprotein YgeR precursor	-2.17	6.43E-03	4.20E-02
UTI89_C1816	ydgK	hypothetical protein YdgK	-2.16	6.89E-03	4.39E-02
UTI89_C1705	yddW	hypothetical lipoprotein YddW precursor	-2.15	5.35E-03	3.66E-02
UTI89_C1447	cls	cardiolipin synthase	-2.14	4.25E-02	1.54E-01
UTI89_C4619	qor	Qor subunit of quinone oxidoreductase	-2.13	7.01E-03	4.46E-02
UTI89_C0363	prpB	PrpB protein	-2.13	4.95E-02	1.70E-01
UTI89_C0544	ylbB	allantoate amidohydrolase	-2.10	3.55E-02	1.36E-01
UTI89_C2642	hkcB	hypothetical protein HkcB	-2.10	9.93E-03	5.83E-02
UTI89_C1362	UTI89_C1362	hypothetical protein	-2.09	4.20E-02	1.53E-01
UTI89_C4376	uvrD	DNA helicase II	-2.09	3.98E-03	2.92E-02
UTI89_C1235	ycfO	conserved hypothetical protein	-2.08	3.04E-02	1.23E-01
UTI89_C0584	fepA	ferrienterobactin receptor precursor	-2.07	1.96E-02	9.14E-02
UTI89_C0572	cusC	probable outer membrane lipoprotein CusC precursor	-2.07	1.74E-03	1.60E-02
UTI89_C1291	UTI89_C1291	putative replication protein P of bacteriophage	-2.03	3.73E-02	1.40E-01

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value#
UTI89_C0596	entE	enterobactin synthetase component E	-2.03	3.59E-02	1.37E-01
UTI89_C2783	eutT	ethanolamine utilization adenosyltransferase	-2.03	1.94E-02	9.12E-02
UTI89_C3873	yhfL	hypothetical protein	-2.02	1.57E-02	8.10E-02
UTI89_C0794	ybhS	hypothetical protein	-1.98	4.47E-02	1.59E-01
UTI89_C3709	yhdV	hypothetical protein YhdV	-1.95	1.61E-02	8.19E-02
UTI89_C5072	bglJ	BglJ transcriptional regulator	-1.95	3.92E-03	2.88E-02
UTI89_C0364	prpC	putative citrate synthase	-1.93	4.08E-02	1.50E-01
UTI89_C3430	yqhC	hypothetical transcriptional regulator YqhC	-1.93	4.62E-02	1.62E-01
UTI89_C2985	UTI89_C2985	bacteriophage V large terminase subunit	-1.93	7.96E-03	4.97E-02
UTI89_C4460	typA	putative GTP-binding factor	-1.92	5.35E-03	3.66E-02
UTI89_C0291	UTI89_C0291	hypothetical protein	-1.92	3.03E-02	1.23E-01
UTI89_C3366	kpsC	capsule polysaccharide export protein KpsC	-1.91	4.88E-02	1.68E-01
UTI89_C2302	gnd	6-phosphogluconate dehydrogenase, decarboxylating	-1.91	2.23E-02	1.00E-01
UTI89_C0680	UTI89_C0680	putative alcohol dehydrogenase	-1.90	9.92E-03	5.83E-02
UTI89_C5021	yjiD	hypothetical protein YjiD	-1.88	3.45E-02	1.34E-01
UTI89_C1081	ycdO	protein YcdO	-1.87	3.16E-02	1.26E-01
UTI89_C2209	UTI89_C2209	putative polyketide synthase	-1.87	2.14E-02	9.72E-02
UTI89_C5012	fimI	FimI fimbrial protein	-1.87	4.06E-02	1.50E-01
UTI89_C2498	yojN	putative sensor-like histidine kinase YojN	-1.87	2.73E-02	1.15E-01
UTI89_C0518	cueR	putative HTH-type transcriptional regulator	-1.85	4.70E-02	1.64E-01
UTI89_C0081	leuA	2-isopropylmalate synthase	-1.83	6.40E-03	4.19E-02
UTI89_C0302	UTI89_C0302	hypothetical protein	-1.82	3.06E-02	1.24E-01
UTI89_C2538	yfbH	hypothetical protein YfbH	-1.82	1.82E-02	8.72E-02
UTI89_P026	UTI89_P026	immunity protein	-1.81	2.57E-02	8.70E-02
UTI89_C1177	yceE	YceE/MdtG drug MFS transporter	-1.80	3.23E-02	1.28E-01
UTI89_C4652	yjcE	CPA1 transporter YjcE	-1.79	2.37E-02	1.05E-01
UTI89_C3407	yghS	hypothetical ATP-binding protein YghS	-1.79	2.73E-02	1.15E-01
UTI89_C3540	yqjG	putative transferase	-1.77	3.17E-02	1.26E-01
UTI89_C0510	UTI89_C0510	hypothetical protein	-1.73	3.42E-02	1.34E-01
UTI89_C4783	yjfK	hypothetical protein YjfK	-1.73	1.37E-02	7.39E-02
UTI89_C4044	UTI89_C4044	hypothetical protein	-1.72	4.50E-02	1.59E-01
UTI89_C1170	mdoC	glucans biosynthesis protein MdoC	-1.71	6.78E-03	4.34E-02
UTI89_C2395	yehR	hypothetical lipoprotein YehR precursor	-1.70	4.46E-02	1.59E-01
UTI89_C4326	ilvE	branched-chain amino acid aminotransferase	-1.68	1.99E-02	9.24E-02
UTI89_C1966	ydjF	putative DEOR-type transcriptional regulator	-1.68	4.14E-02	1.52E-01
UTI89_C3315	UTI89_C3315	putative ATP-binding protein of ABC transport system	-1.68	4.22E-02	1.54E-01
UTI89_P046	UTI89_P046	hypothetical serine-threonine protein kinase	-1.68	4.69E-02	1.29E-01
UTI89_C3557	yhaF	hypothetical protein	-1.67	3.73E-02	1.40E-01
UTI89_C2192	UTI89_C2192	hypothetical protein	-1.60	2.48E-02	1.08E-01
UTI89_C3070	gutQ	GutQ protein	-1.59	4.10E-02	1.51E-01
UTI89_C4303	rbsD	high affinity ribose transport protein RbsD	-1.58	4.20E-02	1.53E-01
UTI89_C0349	yahE	hypothetical protein YahE	-1.58	1.90E-02	8.96E-02
UTI89_C0085	ilvI	acetolactate synthase isozyme III large subunit	-1.56	4.41E-02	1.58E-01
UTI89_C2622	yfcU	putative outer membrane usher YfcU precursor	-1.53	2.83E-02	1.18E-01
UTI89_C2403	yehY	hypothetical ABC transporter permease protein	-1.52	4.61E-02	1.62E-01
UTI89_C3367	kpsS	polysialic acid capsule synthesis protein KpsS	-1.51	2.92E-02	1.20E-01
UTI89_C2930	UTI89_C2930	hypothetical protein	-1.50	4.45E-02	1.59E-01
UTI89_C1814	UTI89_C1814	hypothetical protein	-1.50	4.30E-02	1.55E-01
UTI89_C3439	UTI89_C3439	ureidoglycolate dehydrogenase	-1.49	1.68E-02	8.42E-02
UTI89_C3872	cysG	multifunctional protein; siroheme synthase	-1.46	2.14E-02	9.72E-02
UTI89_C3858	prkB	putative phosphoribulokinase	-1.45	4.57E-02	1.61E-01
UTI89_P144	UTI89_P144	hypothetical protein	-1.45	1.26E-02	5.27E-02
UTI89_P113	traV	F pilus assembly protein TraV	-1.40	2.07E-02	7.27E-02
UTI89_C3851	slyD	FKBP-type peptidyl-prolyl cis-trans isomerase	-1.39	4.47E-02	1.59E-01
UTI89_C2189	UTI89_C2189	hypothetical protein	-1.28	4.84E-02	1.68E-01

**Hypothetical or proteins with no characterised homologs are listed with the gene ID.*
[#]Q-value as determined by TraDIS analysis. The Q-value is the P-value that has been adjusted for the false discovery rate (FDR) for each gene hit.

5.2.2.3 The genes required for filament reversal and bacterial recovery from dispersal

Upon dispersal, filamentous UPEC have the potential to revert to bacillary rod-shaped morphology and can thereafter divide as normal bacterial cells (*17*). This filament reversal is necessary for long-term bacterial survival and for UPEC to regain their invasive abilities in the bladder to initiate a new cycle of infection (*26*). Similarly to filamentation, the mechanisms surrounding reversal and regulation remain unclear. In order to determine the profile of genes important for filament reversal and bacterial recovery from bladder cell dispersal, we cultured the UTI89 bacteria released from the bladder cells (in 2 h collection windows) in liquid LB. All cells harvested from the dispersal phase were centrifuged and then resuspended in liquid LB and allowed to grow overnight for 18 h (Recovery sample). A subpopulation of this culture was then grown for a further 24 h in liquid LB media (Recovery-24h sample). The cells from both liquid cultures then underwent gDNA extraction and sequencing. The data obtained from these Recovery samples were then both individually compared to the Dispersal sample to identify any mutants that had a significant reduced ability to survive. A total of 322 genes meeting the stringent criteria of P-value $\leq$ 0.05 and $Log_2FC < -1.0$ were identified and are shown in Table 5.3. When comparing the genes identified between the Recovery and Recovery-24h samples, 271 genes were found in common and these are most strongly implicated as essential for bacterial reversal and recovery (Supp. Data Table S3F-G). As anticipated, a number of these genes were found to be involved in the regulation of cell division and peptidoglycan remodeling. Among the strongest genes identified were *slmA*, *yfiR, yibP* (*envC*) and *ampH* (Table 5.3). The nucleoid-binding protein, SlmA is an inhibitor of FtsZ polymerization and prevents the Z-ring from forming when in close association to the chromosomes. Our data revealed a log_2 fold-change decrease of 4.11 for *slmA* and Figure 5.5E shows the insertion-site read frequency and distribution of *slmA* and surrounding genes. A newly discovered regulatory repressor gene, *yfiR* was also identified in our screen with a log_2 fold-change decrease of 3.13 (Table 5.3). YfiR is known to prevent the location of (YfiN) DgcN to the mid-cell, which halts cell division by blocking septal peptidoglycan synthesis (*193*) The peptidoglycan regulation, synthesis and remodeling genes *yibP*, which localises to the

septal ring and *ampH* were also identified in our TraDIS screen (log_2FC -3.75 and -5.76, respectively (Table 5.3).

A number of genes were also found to be involved in DNA replication and repair (Table 5.3 and Supp. Data Table S3F-G). We identified the DNA-binding transcriptional regulator StpA (log_2FC -3.13), which is involved in chromosome packaging. We also identified *recG*, which encodes the DNA helicase that has a role in repairing double-stranded breaks within the DNA. Some of the other strong genes identified by TraDIS that have a role in DNA replication or repair include *hepA*, *radC* and *dinF* with log_2 fold-change decreases of 7.16, 5.88 and 2.67, respectively (Table 5.3). We were also interested in other notably strong genes that we identified in the recovery phase that may not be implicated in cell division or DNA replication; one such example is *ompC* (Table 5.3). The outer membrane porin, OmpC (log_2FC -6.55), is positively selected for in UPEC strains and has a role in cell membrane biogenesis and pore formation (*102*).

Table 5.3 UTI89 mutants that have a reduced ability to recover from dispersal

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C4097	cspA	cold shock protein 7.4, transcriptional activator of hns	-10.74	1.11E-07	1.63E-05
UTI89_C3970	UTI89_C3970	putative phosphotransferase system protein	-10.07	6.70E-08	1.61E-05
UTI89_C2682	UTI89_C2682	FtsH protease inhibitor of enterobacteria phage Sf6	-9.70	1.32E-07	1.63E-05
UTI89_C2810	yfgC	hypothetical protein YfgC precursor	-9.70	1.83E-06	6.83E-05
UTI89_C5129	UTI89_C5129	putative minor tail protein	-9.69	1.71E-07	1.63E-05
UTI89_C4599	UTI89_C4599	hypothetical protein	-9.66	1.53E-07	1.63E-05
UTI89_C4598	UTI89_C4598	hypothetical protein	-9.66	1.51E-07	1.63E-05
UTI89_C4265	yieE	hypothetical protein	-9.46	7.50E-07	3.96E-05
UTI89_C3891	yrfB	hypothetical protein YrfB	-9.39	6.78E-07	3.95E-05
UTI89_C3105	ygbM	hypothetical protein	-9.26	1.72E-07	1.63E-05
UTI89_C0949	UTI89_C0949	prophage CP-933T putative replication protein	-9.16	1.43E-07	1.63E-05
UTI89_C0281	crl	crl transcriptional regulator	-8.83	4.79E-06	1.35E-04
UTI89_C1162	UTI89_C1162	hypothetical protein	-8.80	2.52E-07	1.89E-05
UTI89_C3962	livG	ATP-binding component of aa transport	-8.69	1.06E-06	5.02E-05
UTI89_C3206	UTI89_C3206	putative exported protein	-8.61	2.70E-07	1.89E-05
UTI89_C3107	UTI89_C3107	hypothetical protein	-8.56	1.01E-07	1.63E-05
UTI89_C1414	chaC	cation transport protein ChaC	-8.54	1.40E-07	1.63E-05
UTI89_C4701	phnB	hypothetical protein PhnB	-8.50	1.90E-06	6.83E-05
UTI89_C1622	hslJ	heat shock protein HslJ	-8.48	7.62E-07	3.96E-05
UTI89_C0248	UTI89_C0248	hypothetical protein	-8.40	1.20E-05	2.65E-04
UTI89_C1713	UTI89_C1713	putative ARAC-type regulatory protein	-8.40	2.00E-07	1.67E-05
UTI89_C1347	ycgZ	hypothetical protein YcgZ	-8.39	1.94E-06	6.83E-05
UTI89_C0071	UTI89_C0071	hypothetical protein	-8.39	1.79E-07	1.63E-05
UTI89_C1135	UTI89_C1135	hypothetical protein	-8.36	2.19E-06	7.02E-05
UTI89_C1363	ycgJ	hypothetical protein YcgJ precursor	-8.27	4.23E-07	2.55E-05
UTI89_C4958	UTI89_C4958	hypothetical protein	-8.18	1.15E-08	4.44E-06
UTI89_C4391	yigM	hypothetical membrane protein YigM	-8.03	1.67E-08	5.35E-06
UTI89_C4909	UTI89_C4909	putative FMN-dependent dehydrogenase	-8.03	7.58E-07	3.96E-05
UTI89_C2538	yfbH	hypothetical protein YfbH	-7.98	2.73E-06	8.34E-05
UTI89_C4550	argB	acetylglutamate kinase	-7.91	1.12E-06	5.12E-05
UTI89_C1595	tpx	thiol peroxidase	-7.89	2.06E-05	3.88E-04
UTI89_C2639	mnt	regulatory protein	-7.87	1.79E-06	6.83E-05
UTI89_C1157	ycdZ	hypothetical protein YcdZ	-7.85	6.40E-06	1.64E-04
UTI89_C1072	ycdJ	hypothetical protein YcdJ	-7.79	2.53E-05	4.64E-04
UTI89_C4424	tatB	sec-independent twin-arginine translocase subunit	-7.69	9.57E-08	1.63E-05
UTI89_C2312	rfbB	RfbB, subunit of dTDP-glucose 4,6-dehydratase	-7.61	2.85E-07	1.89E-05
UTI89_C0513	UTI89_C0513	hypothetical protein	-7.61	2.11E-06	6.99E-05
UTI89_C0079	leuC	3-isopropylmalate isomerase (dehydratase) subunit	-7.60	7.80E-06	1.95E-04
UTI89_C3413	hybF	hydrogenase nickel incorporation protein	-7.56	1.07E-06	5.02E-05
UTI89_C2712	ypdA	hypothetical protein YpdA	-7.52	3.65E-07	2.34E-05
UTI89_C4147	yibK	hypothetical protein	-7.48	2.89E-05	5.04E-04
UTI89_C0713	abrB	putative transport protein AbrB	-7.35	1.77E-06	6.83E-05
UTI89_C4173	rfaP	lipopolysaccharide core biosynthesis protein rfaP	-7.29	7.04E-07	3.96E-05
UTI89_C1160	csgE	curli production assembly/transport component	-7.23	2.01E-06	6.83E-05
UTI89_C2781	eutM	detox protein	-7.19	2.31E-05	4.31E-04
UTI89_C0064	hepA	probable ATP-dependent RNA helicase	-7.16	1.42E-06	5.83E-05
UTI89_C1099	UTI89_C1099	putative aminotransferase	-7.09	2.02E-06	6.83E-05
UTI89_C0065	polB	DNA polymerase II	-7.04	3.30E-06	9.78E-05
UTI89_C0258	UTI89_C0258	hypothetical protein	-6.98	1.94E-05	3.85E-04
UTI89_C4908	UTI89_C4908	hypothetical protein	-6.91	2.03E-05	3.87E-04
UTI89_C1711	ydeM	hypothetical protein YdeM	-6.89	1.85E-05	3.80E-04
UTI89_C0288	UTI89_C0288	hypothetical protein	-6.85	6.14E-06	1.62E-04
UTI89_C0275	UTI89_C0275	hypothetical protein	-6.85	9.73E-05	1.21E-03
UTI89_C0044	caiB	crotonobetainyl-CoA:carnitine CoA-transferase	-6.84	1.21E-06	5.40E-05

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C1341	UTI89_C1341	hypothetical protein	-6.66	2.17E-06	7.02E-05
UTI89_C3895	nudE	ADP compounds hydrolase NudE	-6.59	1.10E-05	2.52E-04
UTI89_C3827	thiF	adenylyltransferase ThiF	-6.57	1.71E-04	1.92E-03
UTI89_C2497	ompC	outer membrane protein 1b (Ib;c)	-6.55	1.38E-05	2.93E-04
UTI89_C0257	UTI89_C0257	hypothetical protein	-6.45	8.18E-05	1.06E-03
UTI89_C4209	yicI	putative family 31 glucosidase YicI	-6.45	5.44E-05	7.76E-04
UTI89_C1457	intO	putative integrase for prophage CP-933O	-6.40	1.24E-05	2.68E-04
UTI89_C3596	truB	tRNA pseudouridine 5S synthase	-6.38	2.22E-06	7.02E-05
UTI89_C3380	gspI	hypothetical type II secretion protein GspI	-6.37	7.93E-05	1.04E-03
UTI89_C0823	ybiS	protein YbiS precursor	-6.30	1.26E-05	2.70E-04
UTI89_C2196	yeeN	hypothetical protein	-6.29	9.75E-06	2.35E-04
UTI89_C4488	rhaS	L-rhamnose operon regulatory protein RhaS	-6.27	1.77E-04	1.96E-03
UTI89_C4850	pyrI	subunit aspartate-carbamoyltransferase	-6.23	9.00E-05	1.15E-03
UTI89_C4425	tatC	sec-independent twin-arginine translocase subunit	-6.20	6.33E-06	1.64E-04
UTI89_C1736	UTI89_C1736	hypothetical protein	-6.18	1.14E-05	2.55E-04
UTI89_C4879	UTI89_C4879	hypothetical protein	-6.16	1.96E-05	3.85E-04
UTI89_C4607	malM	maltose operon periplasmic protein precursor	-6.15	4.00E-04	3.61E-03
UTI89_C2439	yeiN	hypothetical protein	-6.14	3.87E-04	3.52E-03
UTI89_C3874	yhfS	hypothetical protein YhfS	-6.09	2.01E-05	3.86E-04
UTI89_C3321	cmtB	PTS system, mannitol-specific enzyme II	-6.07	3.85E-05	6.23E-04
UTI89_C3854	yheR	putative NAD(P)H oxidoreductase	-6.05	4.64E-05	7.08E-04
UTI89_C0068	araB	L-ribulokinase	-6.05	1.56E-04	1.81E-03
UTI89_C4260	tnaL	tryptophanase leader peptide	-6.00	2.81E-04	2.81E-03
UTI89_C4759	yjeP	hypothetical protein YjeP precursor	-5.97	5.77E-06	1.54E-04
UTI89_C5091	yfdR2	hypothetical protein	-5.93	4.20E-04	3.76E-03
UTI89_C0158	yadB	conserved hypothetical protein YadB	-5.93	4.64E-05	7.08E-04
UTI89_C5132	UTI89_C5132	putative tail component of prophage CP-933K	-5.93	9.16E-06	2.23E-04
UTI89_C1716	fmlD	putative Fml fimbrial adhesin FmlD precursor	-5.93	8.04E-06	1.98E-04
UTI89_C2605	pdxB	erythronate-4-phosphate dehydrogenase	-5.91	5.85E-05	8.28E-04
UTI89_C3524	ygjQ	hypothetical protein YgjQ	-5.91	7.05E-05	9.49E-04
UTI89_C0781	moaB	molybdopterin biosynthesis, protein B	-5.90	3.80E-05	6.23E-04
UTI89_C4628	UTI89_C4628	NadR transcriptional repressor	-5.90	2.59E-05	4.68E-04
UTI89_C4182	radC	DNA repair protein	-5.88	3.36E-04	3.17E-03
UTI89_C3533	yqjA	hypothetical protein	-5.87	1.30E-04	1.55E-03
UTI89_C1459	exoO	exonuclease VIII, encoded by prophage CP-933O	-5.85	2.00E-05	3.86E-04
UTI89_C0361	yahN	putative cytochrome subunit of dehydrogenase	-5.83	9.10E-05	1.15E-03
UTI89_C1732	UTI89_C1732	putative phage protein	-5.79	4.64E-05	7.08E-04
UTI89_C0391	yaiH	penicillin-binding protein AmpH	-5.76	1.10E-05	2.52E-04
UTI89_C3555	yhaD	glycerate kinase 2	-5.74	1.10E-04	1.34E-03
UTI89_C1810	malY	maltose regulon modulator	-5.73	4.21E-05	6.76E-04
UTI89_C4067	yhjR	hypothetical protein YhjR	-5.73	7.59E-04	5.84E-03
UTI89_C2303	wbgM	putative galactosyltransferase WbgM	-5.70	1.12E-05	2.53E-04
UTI89_C1913	UTI89_C1913	hypothetical protein	-5.68	5.30E-05	7.67E-04
UTI89_C3093	hypE	hydrogenase isoenzymes formation protein HypE	-5.66	6.76E-05	9.28E-04
UTI89_C1161	csgD	2-component transcriptional regulator for curli	-5.64	6.19E-04	5.00E-03
UTI89_C2125	fliD	flagellar hook-associated protein 2	-5.63	3.35E-04	3.17E-03
UTI89_C3045	emrA	multidrug resistance protein A	-5.62	4.29E-04	3.78E-03
UTI89_C2382	yehA	putative Yeh fimbiral adhesin YehA precursor	-5.57	2.85E-05	5.03E-04
UTI89_C0929	UTI89_C0929	predicted tail completion phage protein	-5.55	3.85E-05	6.23E-04
UTI89_C4392	metR	regulator for metE and MetH	-5.54	1.23E-05	2.68E-04
UTI89_C0702	kdpA	potassium-transporting ATPase A chain	-5.51	2.91E-05	5.04E-04
UTI89_C1931	celA	N,N'-diacetylchitobiose PTS permease	-5.44	1.24E-03	8.49E-03
UTI89_C3557	yhaF	hypothetical protein	-5.44	6.24E-05	8.76E-04
UTI89_C5116	UTI89_C5116	unknown protein encoded by prophage CP-933K	-5.43	2.39E-04	2.45E-03
UTI89_C2114	uvrC	excinuclease ABC subunit C	-5.43	4.10E-03	2.34E-02
UTI89_C1458	UTI89_C1458	putative excisionase	-5.39	2.19E-04	2.32E-03

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C0244	UTI89_C0244	hypothetical protein	-5.36	8.71E-04	6.42E-03
UTI89_C2647	UTI89_C2647	DNA stabilization protein	-5.34	1.28E-03	8.69E-03
UTI89_C2672	UTI89_C2672	bacteriophage Nil2 gene P DnaB analogue	-5.34	4.84E-05	7.22E-04
UTI89_C4691	phnL	phosphonates transport ATP-binding protein PhnL	-5.32	7.61E-05	1.01E-03
UTI89_C0393	sbmA	inner-membrane transport protein, Microcin 25	-5.32	4.92E-05	7.28E-04
UTI89_C5050	hsdS	HsdS, type I site-specific deoxyribonuclease	-5.30	7.05E-05	9.49E-04
UTI89_C0990	ycbJ	hypothetical protein YcbJ	-5.27	1.59E-04	1.82E-03
UTI89_C2231	cobU	cobalamin biosynthesis protein CobU	-5.23	4.81E-05	7.22E-04
UTI89_C2785	eutP	ethanolamine utilization protein EutP	-5.21	7.07E-04	5.53E-03
UTI89_C0752	gpmA	phosphoglycerate mutase 1	-5.21	8.51E-04	6.37E-03
UTI89_C3787	yheJ	secretion pathway protein K precursor	-5.20	1.35E-04	1.58E-03
UTI89_C1325	UTI89_C1325	putative tail component of prophage	-5.19	1.63E-04	1.84E-03
UTI89_C0043	caiC	probable crotonobetaine/carnitine-CoA ligase	-5.17	3.88E-04	3.52E-03
UTI89_C3543	yhaK	conserved hypothetical protein	-5.05	1.67E-04	1.88E-03
UTI89_C3490	ygiF	hypothetical protein YgiF	-5.04	2.42E-05	4.48E-04
UTI89_C2643	UTI89_C2643	DNA transfer protein	-5.03	3.34E-04	3.17E-03
UTI89_C2484	napA	periplasmic nitrate reductase precursor	-4.97	1.24E-04	1.49E-03
UTI89_C2384	yehC	putative periplasmic chaperone YehC precursor	-4.96	9.84E-04	7.06E-03
UTI89_C4940	UTI89_C4940	hypothetical protein	-4.89	2.00E-03	1.29E-02
UTI89_C2779	eutE	ethanolamine utilization acetaldehyde dehydrogenase	-4.88	2.78E-04	2.78E-03
UTI89_C4155	yibN	hypothetical protein YibN	-4.84	5.03E-04	4.25E-03
UTI89_P102	UTI89_P102	hypothetical protein	-4.84	6.28E-05	2.66E-03
UTI89_P129	trbB	F pilus assembly periplasmic protein TrbB	-4.83	1.06E-04	2.66E-03
UTI89_C3272	ygfU	putative purine permease YgfU	-4.83	1.78E-06	6.83E-05
UTI89_C4706	basS	sensor kinase-phosphotransferase basS/PmrB	-4.83	2.05E-03	1.30E-02
UTI89_C3109	UTI89_C3109	hypothetical protein	-4.82	1.13E-03	7.86E-03
UTI89_C4011	yhiN	hypothetical protein YhiN	-4.76	3.24E-04	3.15E-03
UTI89_C2688	hkaJ	hypothetical protein	-4.76	4.76E-03	2.60E-02
UTI89_C3186	mltA	membrane-bound lytic murein transglycosylase	-4.76	2.23E-04	2.32E-03
UTI89_C4667	UTI89_C4667	hypothetical protein	-4.75	1.20E-03	8.26E-03
UTI89_C4353	yifK	probable transport protein YifK	-4.70	1.53E-03	1.02E-02
UTI89_C0420	phoB	positive response regulator for pho regulon	-4.68	3.59E-03	2.11E-02
UTI89_C5025	cgrD	glycerol dehydrogenase CgrD	-4.65	1.90E-03	1.24E-02
UTI89_C1700	UTI89_C1700	hypothetical protein	-4.64	5.15E-04	4.32E-03
UTI89_C5002	UTI89_C5002	hypothetical protein	-4.63	1.99E-04	2.16E-03
UTI89_C0698	kdpE	transcriptional regulatory protein KdpE	-4.62	4.62E-03	2.55E-02
UTI89_C4153	UTI89_C4153	hypothetical protein	-4.58	7.21E-03	3.49E-02
UTI89_C3917	malQ	4-alpha-glucanotransferase	-4.57	3.80E-04	3.51E-03
UTI89_C2758	cysP	thiosulfate binding protein	-4.53	9.38E-04	6.84E-03
UTI89_C3546	yhaO	putative transport system permease protein	-4.52	2.97E-03	1.81E-02
UTI89_C1718	fmlB	periplasmic chaperone FmlB protein	-4.52	5.07E-03	2.69E-02
UTI89_C3132	ygcP	putative anti-terminator regulatory protein	-4.52	1.30E-02	5.35E-02
UTI89_C4400	UTI89_C4400	transketolase 1	-4.49	4.94E-06	1.36E-04
UTI89_C4025	UTI89_C4025	hypothetical protein	-4.46	1.50E-03	1.00E-02
UTI89_C4853	UTI89_C4853	hypothetical protein	-4.45	6.65E-03	3.27E-02
UTI89_C4214	yicL	hypothetical transport protein YicL	-4.42	3.23E-04	3.15E-03
UTI89_C0360	yahM	hypothetical protein	-4.41	3.82E-04	3.52E-03
UTI89_C0998	ycbK	conserved hypothetical protein	-4.40	5.88E-04	4.82E-03
UTI89_C1368	UTI89_C1368	putative conserved protein	-4.39	4.14E-03	2.34E-02
UTI89_C4721	yjdJ	hypothetical protein	-4.29	7.01E-04	5.50E-03
UTI89_C4328	ilvA	threonine dehydratase biosynthetic	-4.28	5.04E-04	4.25E-03
UTI89_C2609	yfcK	hypothetical protein YfcK	-4.28	1.58E-04	1.82E-03
UTI89_C4348	UTI89_C4348	hypothetical protein	-4.28	1.13E-03	7.86E-03
UTI89_C0227	UTI89_C0227	putative transcriptional regulator LYSR-type	-4.23	3.53E-03	2.09E-02
UTI89_C2856	UTI89_C2856	hypothetical protein	-4.22	3.87E-04	3.52E-03
UTI89_C1305	ybcU	lambdoid prophage DLP12 precursor Bor-like protein	-4.21	1.01E-02	4.41E-02

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C4197	recG	ATP-dependent DNA helicase RecG	-4.17	5.88E-04	4.82E-03
UTI89_C3824	yjaE	regulator of sigma D	-4.15	6.19E-03	3.12E-02
UTI89_P041	UTI89_P041	putative transposase	-4.12	3.73E-04	5.60E-03
UTI89_C4974	UTI89_C4974	hypothetical protein	-4.11	2.25E-03	1.41E-02
UTI89_C4185	slmA	putative transcriptional division regulator	-4.11	4.73E-05	7.16E-04
UTI89_C0567	ybcH	hypothetical protein YbcH precursor	-4.09	4.45E-03	2.48E-02
UTI89_C1165	csgA	curlin major subunit CsgA	-4.07	1.16E-03	8.06E-03
UTI89_C4603	malF	maltose transport system permease protein MalF	-4.05	9.19E-05	1.15E-03
UTI89_C3736	yhdL	putative regulator	-4.00	2.83E-03	1.73E-02
UTI89_C1733	UTI89_C1733	putative phage protein	-3.99	7.13E-03	3.47E-02
UTI89_C0328	ykgD	hypothetical transcriptional regulator YkgD	-3.99	1.47E-02	5.72E-02
UTI89_C1322	UTI89_C1322	putative tail component of prophage CP-933X	-3.93	1.08E-02	4.69E-02
UTI89_C2204	UTI89_C2204	prophage P4 integrase	-3.93	9.03E-04	6.60E-03
UTI89_C3166	UTI89_C3166	hypothetical protein	-3.92	3.70E-03	2.15E-02
UTI89_C2687	UTI89_C2687	bacteriophage HK97 gp40	-3.91	2.26E-04	2.32E-03
UTI89_C0201	ldcC	lysine decarboxylase, constitutive	-3.90	5.56E-03	2.88E-02
UTI89_C0798	rhlE	putative ATP-dependent RNA helicase RhlE	-3.88	6.19E-03	3.12E-02
UTI89_C2390	yehI	hypothetical protein YehI	-3.87	1.80E-03	1.18E-02
UTI89_C0924	UTI89_C0924	putative phage protein gpH	-3.87	7.65E-04	5.84E-03
UTI89_C1277	gamW	host-nuclease inhibitor protein Gam of bacteriphage	-3.86	3.00E-04	2.97E-03
UTI89_C0349	yahE	hypothetical protein YahE	-3.86	1.47E-03	9.87E-03
UTI89_C2498	yojN	putative sensor-like histidine kinase YojN	-3.84	3.05E-03	1.85E-02
UTI89_C5086	UTI89_C5086	hypothetical protein	-3.82	6.11E-03	3.10E-02
UTI89_P062	stbA	stable plasmid inheritance protein	-3.80	1.68E-02	8.40E-02
UTI89_C0812	glnQ	GlnQ protein, subunit of glutamine ABC transporter	-3.76	1.08E-02	4.69E-02
UTI89_C4350	UTI89_C4350	hypothetical protein	-3.75	1.45E-02	5.70E-02
UTI89_C4158	yibP	hypothetical protein YibP	-3.75	4.85E-03	2.64E-02
UTI89_C4080	dppC	dipeptide transport system permease protein 2	-3.74	6.10E-03	3.10E-02
UTI89_C3201	UTI89_C3201	hypothetical protein	-3.73	1.36E-02	5.49E-02
UTI89_C3175	fucK	L-fuculokinase	-3.72	5.33E-03	2.78E-02
UTI89_C3406	yghR	hypothetical protein	-3.71	3.56E-03	2.10E-02
UTI89_C3516	ebgA	evolved beta-galactosidase alpha-subunit	-3.71	2.25E-04	2.32E-03
UTI89_C3373	neuD	sialic acid synthase NeuD	-3.71	6.81E-03	3.33E-02
UTI89_C3409	pitB	low-affinity inorganic phosphate transporter 2	-3.70	4.28E-03	2.41E-02
UTI89_C4862	UTI89_C4862	hypothetical protein	-3.70	2.65E-02	9.04E-02
UTI89_C5003	UTI89_C5003	hypothetical protein	-3.70	9.09E-03	4.07E-02
UTI89_C4901	UTI89_C4901	hypothetical protein	-3.68	8.98E-03	4.05E-02
UTI89_C4576	iclR	aceBAK operon repressor	-3.68	6.42E-03	3.17E-02
UTI89_C3826	thiE	thiamine-phosphate pyrophosphorylase	-3.68	7.21E-03	3.49E-02
UTI89_C3525	ygjR	hypothetical oxidoreductase YgjR	-3.67	1.21E-02	5.09E-02
UTI89_C4222	uhpT	hexose phosphate transport protein	-3.63	1.43E-02	5.69E-02
UTI89_C0891	UTI89_C0891	hypothetical protein	-3.63	1.31E-02	5.37E-02
UTI89_C0570	cusS	sensor kinase CusS	-3.62	2.81E-02	9.38E-02
UTI89_C4096	UTI89_C4096	hypothetical protein	-3.61	2.61E-02	8.98E-02
UTI89_P103	traM	tra operon transcriptional repressor	-3.60	1.96E-03	2.10E-02
UTI89_C1411	chaA	calcium/proton antiporter	-3.57	1.34E-02	5.44E-02
UTI89_C3346	ansB	L-asparaginase II precursor	-3.56	1.14E-02	4.84E-02
UTI89_C2621	yfcS	putative periplasmic chaperone YfcS precursor	-3.56	5.80E-03	2.97E-02
UTI89_C2227	UTI89_C2227	hypothetical protein	-3.54	1.86E-02	6.94E-02
UTI89_C2128	amyA	cytoplasmic alpha-amylase	-3.54	1.81E-02	6.82E-02
UTI89_C4270	yieL	putative xylanase	-3.54	7.49E-03	3.58E-02
UTI89_C4578	UTI89_C4578	hypothetical protein	-3.53	1.44E-02	5.70E-02
UTI89_C2257	UTI89_C2257	insertion sequence ATP-binding protein	-3.51	4.30E-03	2.42E-02
UTI89_C3517	ebgC	evolved beta-D-galactosidase beta subunit	-3.51	2.01E-02	7.28E-02
UTI89_C3671	UTI89_C3671	conserved hypothetical protein	-3.50	1.99E-02	7.28E-02
UTI89_C0573	ylcC	conserved hypothetical protein	-3.48	1.41E-02	5.61E-02

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C2157	UTI89_C2157	hypothetical protein	-3.48	6.30E-03	3.15E-02
UTI89_C4411	udp	uridine phosphorylase	-3.43	3.93E-03	2.25E-02
UTI89_C4223	uhpC	regulatory protein UhpC	-3.42	2.26E-03	1.41E-02
UTI89_C4626	alr	alanine racemase 1	-3.40	8.67E-03	3.96E-02
UTI89_C3268	ygfP	guanine deaminase	-3.38	1.37E-02	5.53E-02
UTI89_C0289	UTI89_C0289	hypothetical protein	-3.37	1.28E-02	5.30E-02
UTI89_C0343	betT	high-affinity choline transport protein	-3.35	2.67E-02	9.07E-02
UTI89_C4335	trxA	thioredoxin 1	-3.34	1.67E-02	6.35E-02
UTI89_C2173	UTI89_C2173	hypothetical protein	-3.34	1.31E-02	5.37E-02
UTI89_C0062	djlA	DnaJ-like protein DjlA	-3.34	2.00E-02	7.28E-02
UTI89_C1623	ldhA	fermentative D-lactate dehydrogenase	-3.33	7.46E-03	3.57E-02
UTI89_C4070	yhjU	hypothetical protein YhjU	-3.33	5.06E-03	2.69E-02
UTI89_P131	trbF	hypothetical protein	-3.32	1.25E-02	6.71E-02
UTI89_C0857	potF	putrescine-binding periplasmic protein precursor	-3.31	9.46E-03	4.20E-02
UTI89_C3511	ygjG	probable ornithine aminotransferase	-3.31	9.26E-03	4.12E-02
UTI89_C5049	yjiW	hypothetical protein	-3.30	4.37E-03	2.45E-02
UTI89_C3035	nrdE	ribonucleoside-diphosphate reductase 2 alpha chain	-3.29	5.92E-03	3.02E-02
UTI89_C3388	pppA	putative prepilin peptidase A	-3.26	4.99E-03	2.68E-02
UTI89_C2681	UTI89_C2681	enterobacteria phage Sf6 gene 34 protein	-3.26	1.80E-02	6.79E-02
UTI89_C3790	hofD	type 4 prepilin-like protein	-3.26	5.53E-03	2.88E-02
UTI89_C0931	UTI89_C0931	hypothetical protein	-3.22	1.85E-02	6.94E-02
UTI89_C0923	UTI89_C0923	tail fiber assembly protein	-3.21	8.08E-03	3.78E-02
UTI89_C2678	UTI89_C2678	hypothetical protein	-3.19	1.38E-02	5.57E-02
UTI89_C4788	yjfN	hypothetical protein	-3.15	3.56E-02	1.10E-01
UTI89_C1440	UTI89_C1440	hypothetical protein	-3.14	4.49E-02	1.31E-01
UTI89_C2936	yfiR	hypothetical protein YfiR precursor	-3.13	9.87E-03	4.33E-02
UTI89_C3025	stpA	DNA-binding protein StpA	-3.13	5.33E-03	2.78E-02
UTI89_C3316	yggC	hypothetical protein YggC	-3.10	8.69E-03	3.96E-02
UTI89_C1754	UTI89_C1754	PTS system, cellobiose-specific IIA component	-3.10	1.99E-02	7.28E-02
UTI89_C0786	ybhL	hypothetical protein	-3.10	3.08E-02	9.96E-02
UTI89_C2992	UTI89_C2992	hypothetical protein	-3.10	4.67E-03	2.57E-02
UTI89_C2869	yphH	hypothetical protein YphH	-3.08	4.27E-02	1.26E-01
UTI89_C4806	UTI89_C4806	putative hexuronate transporter	-3.06	4.63E-03	2.55E-02
UTI89_C1814	UTI89_C1814	hypothetical protein	-3.06	1.92E-02	7.11E-02
UTI89_C2370	UTI89_C2370	hypothetical protein	-3.04	2.78E-02	9.36E-02
UTI89_C4152	grxC	glutaredoxin 3	-3.02	4.95E-02	1.40E-01
UTI89_C4609	ubiC	chorismate pyruvate lyase	-3.01	1.95E-02	7.17E-02
UTI89_C0002	thrA	aspartokinase I	-2.99	3.23E-02	1.03E-01
UTI89_C0939	UTI89_C0939	putative phage protein gpP	-2.99	3.19E-02	1.02E-01
UTI89_C1660	ydcT	ABC transporter ATP-binding protein ydcT	-2.98	2.98E-02	9.73E-02
UTI89_C1851	ydhC	hypothetical transport protein YdhC	-2.96	8.36E-03	3.85E-02
UTI89_C2419	yeiT	hypothetical oxidoreductase YeiT	-2.95	3.04E-02	9.87E-02
UTI89_C1191	rimJ	ribosomal-protein-alanine acetyltransferase	-2.93	3.93E-02	1.17E-01
UTI89_C0338	fimX	type 1 fimbriae regulatory protein FimX	-2.92	4.19E-02	1.24E-01
UTI89_C3317	yggD	putative transcriptional regulator	-2.91	2.81E-02	9.38E-02
UTI89_C2689	hkaH	hypothetical protein	-2.89	3.08E-02	9.96E-02
UTI89_C1138	UTI89_C1138	hypothetical protein	-2.88	3.10E-02	9.99E-02
UTI89_C2501	atoS	sensor protein AtoS	-2.84	7.47E-03	3.57E-02
UTI89_C2706	yfdW	formyl-coenzyme A transferase monomer	-2.83	4.60E-03	2.55E-02
UTI89_C1190	yceL	hypothetical transport protein YceL	-2.82	3.27E-02	1.04E-01
UTI89_C0515	ybaR	copper-transporting P-type ATPase	-2.82	2.98E-02	9.73E-02
UTI89_C2310	rfbA	glucose-1-phosphate thymidylyltransferase	-2.81	1.25E-02	5.22E-02
UTI89_P110	traB	conjugal transfer protein TraB	-2.81	1.11E-02	6.38E-02
UTI89_C4476	UTI89_C4476	hypothetical protein	-2.81	3.39E-02	1.06E-01
UTI89_C4023	UTI89_C4023	putative conserved protein	-2.79	2.93E-02	9.63E-02
UTI89_C1173	mdoH	periplasmic glucans biosynthesis protein MdoH	-2.77	3.37E-02	1.05E-01

Gene ID	Gene name*	Annotation	Log_2FC	P-value	Q-value[#]
UTI89_C3397	UTI89_C3397	putative saframycin Mx1 synthetase B	-2.77	6.68E-03	3.28E-02
UTI89_C3235	ygdQ	putative transport protein	-2.75	3.65E-02	1.12E-01
UTI89_C3991	acpT	4'-phosphopantetheinyl transferase AcpT	-2.75	2.59E-02	8.95E-02
UTI89_C4559	yijD	hypothetical protein	-2.75	4.65E-02	1.33E-01
UTI89_C3927	UTI89_C3927	conserved hypothetical protein	-2.74	1.51E-02	5.85E-02
UTI89_C2440	yeiC	hypothetical sugar kinase YeiC	-2.74	3.73E-02	1.13E-01
UTI89_C1328	UTI89_C1328	putative tail component of prophage	-2.73	3.33E-02	1.05E-01
UTI89_C1327	UTI89_C1327	hypothetical protein	-2.73	3.32E-02	1.05E-01
UTI89_C4880	UTI89_C4880	putative DNA helicase, superfamily I	-2.72	1.25E-02	5.22E-02
UTI89_C4212	rhuM	putative cytoplasmic protein	-2.72	2.11E-02	7.57E-02
UTI89_C4034	chuW	oxygen independent coproporphyrinogen III oxidase	-2.71	1.45E-02	5.71E-02
UTI89_C4886	papX	putative regulator PapX protein	-2.70	3.73E-03	2.16E-02
UTI89_C4047	yhiW	hypothetical transcriptional regulator YhiW	-2.69	3.91E-03	2.24E-02
UTI89_C2857	csiE	stationary phase inducible protein CsiE	-2.68	4.53E-02	1.31E-01
UTI89_C1967	ydjG	hypothetical oxidoreductase YdjG	-2.67	1.13E-02	4.83E-02
UTI89_C4651	yjcD	hypothetical protein YjcD	-2.67	2.80E-02	9.38E-02
UTI89_C4210	yicJ	hypothetical symporter YicJ	-2.67	2.00E-02	7.28E-02
UTI89_C3683	yhdA	hypothetical protein YhdA	-2.67	2.39E-02	8.35E-02
UTI89_C4614	dinF	DNA-damage-inducible protein F	-2.67	3.64E-02	1.12E-01
UTI89_C4506	yiiS	hypothetical protein	-2.66	1.43E-02	5.69E-02
UTI89_C4467	UTI89_C4467	shikimate transporter	-2.64	1.55E-02	5.97E-02
UTI89_C4450	yihG	hypothetical protein YihG	-2.64	7.99E-03	3.76E-02
UTI89_C1583	ycjP	hypothetical ABC transporter permease protein ycjP	-2.57	2.95E-02	9.66E-02
UTI89_C4299	asnA	asparagine synthetase A	-2.57	2.84E-02	9.44E-02
UTI89_C1117	focX	putative regulatory protein	-2.56	1.46E-02	5.71E-02
UTI89_C0038	UTI89_C0038	transposase	-2.52	1.40E-02	5.58E-02
UTI89_C3908	yhgF	hypothetical protein YhgF	-2.51	3.02E-02	9.83E-02
UTI89_C2385	yehD	putative Yeh fimbriae subunit YehD precursor	-2.42	4.52E-02	1.31E-01
UTI89_C2210	UTI89_C2210	putative peptide synthetase	-2.42	2.00E-02	7.28E-02
UTI89_C3136	UTI89_C3136	hypothetical protein	-2.40	4.39E-02	1.28E-01
UTI89_C4121	yiaN	hypothetical protein YiaN	-2.38	7.98E-03	3.76E-02
UTI89_P052	ydiA	hypothetical protein	-2.37	4.61E-02	1.76E-01
UTI89_C4680	yjcT	D-allose kinase	-2.35	3.73E-02	1.13E-01
UTI89_C4156	UTI89_C4156	hypothetical protein	-2.32	3.37E-02	1.05E-01
UTI89_C4045	yhiU	hypothetical lipoprotein YhiU precursor	-2.28	2.68E-02	9.09E-02
UTI89_P013	UTI89_P013	putative membrane protein	-2.24	3.39E-02	1.46E-01
UTI89_C5061	UTI89_C5061	putative C4-dicarboxylate-binding periplasmic protein	-2.24	2.21E-02	7.82E-02
UTI89_C4977	UTI89_C4977	hypothetical protein	-2.21	3.55E-02	1.10E-01
UTI89_C3438	UTI89_C3438	zinc-type alcohol dehydrogenase-like protein yjjN	-2.20	3.63E-02	1.12E-01
UTI89_C4220	yicO	hypothetical protein YicO	-2.20	4.31E-02	1.27E-01
UTI89_C4554	oxyR	hydrogen peroxide-inducible genes activator	-2.12	3.83E-02	1.15E-01
UTI89_C1252	potC	spermidine/putrescine transport system permease	-2.10	3.30E-02	1.04E-01
UTI89_C0236	yafT	putative aminopeptidase	-2.10	4.55E-02	1.32E-01
UTI89_C4273	bglF	PTS system, beta-glucoside-specific IIABC component	-2.08	3.33E-02	1.05E-01

**Hypothetical or proteins with no characterised homologs are listed with the gene ID.*

[#]Q-value as determined by TraDIS analysis. The Q-value is the P-value that has been adjusted for the false discovery rate (FDR) for each gene hit.

5.3 DISCUSSION

This chapter describes, for the first time, a genome-wide identification of genes important for UPEC transitioning through the distinct phases of bladder cell infection. Our findings demonstrate the utility of an up-scaled bladder cell culture model as a tool for genetic or 'omic-style analyses that require large bacterial yields at distinct time points in the course of an infection.

We employed a stringent fold-change and statistical significance (P-value) threshold for cells harvested from the IBC, dispersal and recovery phases of infection to define the set of genes important for survival or growth of UTI89. In order to determine the most stringent set of genes require solely for survival in infection, we eliminated any genes that were previously shown to be important for growth in liquid LB media (mutant library continually passaged over 5 days; data presented in *Chapter 4*). We suggest that these genes were detected in our infection TraDIS screen due to their importance for growth, and may not be relevant to infection. We sought to focus on those genes that are important specifically for infection conditions.

We identified 143 statistically significant genes that were required for UTI89 progression to the IBC phase of bladder infection (Table 5.1). These included numerous genes with known roles in amino acid biosynthesis, motility regulation, LPS biosynthesis and iron homeostasis, confirming that our screen successfully identified genes of biological relevance. Several previously hypothetical genes were also identified that may represent novel genes that play a role in intracellular infection.

An interesting aspect of our TraDIS screen during the IBC phase was the identification of genes involved in central carbon metabolism (Supp. Data Table S3C). This was not unexpected, as UPEC pack into dense biofilm-like aggregates that rely on the capacity for *de novo* synthesis of essential metabolites and sugars from the bladder environment (*96*) However, the identification of genes *pgm* and *galU*, both involved in UDP-glucose biosynthesis, was intriguing. Glucose metabolism has previously been implicated in

regulating cell division in *E. coli* via the glucosyltransferase, *opgH*, which acts on FtsZ via a UDP-glucose-dependent mechanism to coordinate cell size with nutrient availability; reducing cell size without substantially affecting the *E. coli* doubling time (*148*). The IBC begins as rod-shaped bacteria that are loosely organised, but progresses to smaller coccoid daughter cells that are tightly assembled, resembling a biofilm-like organisation (*52*). These findings suggest that *pgm* and *galU* may possibly act, in a UDP-glucose-dependent manner to coordinate the cell size of UTI89 within the IBC after sensing and responding to cytoplasmic sugar levels. Interestingly, the *tol-pal* system genes were also identified in our screen and are known to control outer cell membrane integrity (Table 5.1 and Figure 5.5B). Expression analysis revealed that genes involved in supporting the integrity of the *E. coli* outer membrane such as, *tolB*, and *pal* were upregulated in the human flow chamber-based cell culture model of UTI at the IBC phase, highlighting the potential regulation UPEC undergo to change shape (*56*). Further, the TolQ protein has been reported to localise at mid-cell and interact *in vivo* with the essential cell division protein, FtsN (*194*) This may further suggest that certain modifications are made to the cell membrane and division machinery for IBC formation.

The relevance of our screen towards the goal of identifying the components involved in IBC formation and maturation is demonstrated by our finding of several genes known to play key virulence roles within the IBC phase of infection. These included the well-characterised K1 capsule synthesizing gene *neuE* (*187*), the periplasmic prolyl chaperone SurA, which facilitates outer membrane protein biogenesis and pilus assembly (*97*), the outer membrane protein OmpA which has a role in maintaining structural membrane integrity (*188*) and various genes involved in LPS biosynthesis such as, *rfaF*, *rfaG* and *rfaP*, which may play a role in remodeling individual cells to maintain the tightly assembled IBC structure (*189*).

Another aspect of our TraDIS screen was to identify the genes that may play a role in the dispersal of UPEC from bladder cells and also to identify any regulators of the UTI-filamentation response pathway. We identified 333 statistically significant genes that were important for UTI89 in the dispersal phase of infection, of which, some were found to be

cell division regulators (Table 5.2). As an example, the newly discovered division gene, *ytfB* was identified and has recently been reported to accumulate at the septal ring via its OapA domain (*195*). The overexpression of the *ytfB* gene from an arabinose-inducible promoter caused extensive *E. coli* filamentation (*88*). We also identified the *cedA* gene, which has a role in cell division regulation and activation (Figure 5.5D). Overexpression of CedA has been reported to prevent the filamentation phenotype that is caused by the *dnaAcos* mutant strain of *E. coli* (mutant causes over-initiation of DNA replication) (*120*). Additionally, the *cedA* gene has been reported to be under strong positive selection within UPECs (*102*) and has up-regulated expression in the dispersal phase of UPEC infection (*56*). This may suggest a more regulatory role for *cedA* in the UTI-filamentation response pathway.

The relevance of our screen towards identifying components of the UPEC filamentation response pathway is demonstrated by our finding of *damX* with strong significance (Table 5.2 and Fig. 5.5C). DamX is recruited to the septum during cell division and is involved in binding septal peptidoglycan (*87*). Recently, Khandige *et al*. (*56*) demonstrated a lack of filamentation in the *damX* deletion mutant of UTI89 (UTI89Δ*damX*) during dispersal in the human flow chamber-based cell culture model and in the murine mouse model of UTI, implicating *damX* as a requirement for UPEC filamentation during infection (*56*).

In order to colonise the intracellular bladder cell environment and disperse out into the bladder lumen, UPEC must be able to sense and respond to certain stresses that may compromise the cell envelope. In our study we identified genes from four of the cell envelope stress response systems, *envZ*/*ompR*, *cpxR*, *rcsD* and *baeS*. The urinary tract environment and urine contain large amounts of inorganic ions and urea, which impose osmotic and denaturing stresses on *E. coli* that need to be overcome for infection (*196*). It is possible that these stresses exhibited by the cell envelope during the dispersal phase also result in altered division regulatory mechanisms to arrest division, as has been reported for the Rcs system (*197*). Several previously uncharacterised genes were also identified that may represent new inhibitors or mediators of division in UPEC.

As ~70% of cells harvested from the dispersal phase were non-filamentous (Supp. Data Fig. S6), we also anticipated identifying genes important to other processes. These included genes within the functional classes of carbon metabolism (*aceE*), nucleotide biosynthesis (*pyrE*) and iron acquisition (*ChuU*), which represent considerable functional overlap with the IBC phase (Supp. Data Table S3). This was not unexpected due to the dynamic nature of the infection cascade where many individual events are predicted to overlap (i.e. single IBCs could last well into the dispersal phase before erupting from the bladder cell). Further, highly-motile cells are also dispersed out from the IBC into the bladder lumen. We identified the FlaJ (MotA) protein that facilities the formation of the flagella motor complex (Table 5.2). Again this could represent overlap between the IBC and dispersal phases as flagella do play an essential role in establishing the IBC structure (*186*). Within the IBC phase we identified the flagella motor switch complex protein FliN, the biosynthetic component FliR and the sensory detector CheA, which transmits signals from the chemoreceptors to the flagella motors (Table 5.1). It is likely that motility proteins are needed for both the establishment and maturation of the IBC, and subsequent dispersal producing a highly-motile population of bacteria into the bladder lumen (*186*).

Filament reversal is necessary for long-term bacterial survival and for UPEC to regain their invasive abilities in the bladder (*26*). We identified 322 statistically significant genes that were required for UTI89 filament reversal and bacterial recovery from dispersal of the bladder cells (Table 5.3). As anticipated, a number of genes were found to be involved in the regulation of cell division. The nucleoid-binding protein, SlmA was identified and is an inhibitor of FtsZ polymerization (Figure 5.5E). SlmA prevents the Z-ring from forming when in close association to the chromosomes (*136*). The identification of *slmA* suggests it could have a role in controlling division of filaments, to produce short rod cells during recovery from infection. Cell division in the absence of *slmA* would lead to Z-ring positioning over the nucleoid, which could result in lethal DNA-compromised daughter cells.

We also identified the repressor gene, *yfiR*, which is known to negatively regulate the localisation of YfiN (also known as DgcN) at the Z-ring to prevent cell division arrest in

response to envelope stress (*193*). When *E. coli* senses envelope stress, *yfiN* produces c-di-GMP and immediately locates to the Z-ring to inhibit cell division by preventing septal peptidoglycan synthesis (*193*). This may suggest an envelope stress-filament-mediated response in UTI89 during infection, which requires the negative regulator YfiR to reverse the YfiN-mediated division arrest and recover the bacterial cells for continued growth. Also identified in our screen in the recovery phase was the outer membrane porin, OmpC, which is known to be regulated by the EnvZ/OmpR two-component system. The expression of *ompC* is up-regulated during high osmolality conditions (*198*) and in UPEC CFT073 bacteria isolated from the urine of infected mice (*99*). Interestingly, *ompC* shows strong positive selection in both UPEC and Shigella strains (*102*). Further, Bernardini *et al*. (*198*) reported that the *ompC* mutant of *Shigella flexneri* is unable to spread from infected human caco-2 epithelial cells (mammalian colorectal cell line) and re-invade other neighbouring cells. This has yet to be determined in *E. coli*, however our finding of *ompC* in the recovery/reversal phase may suggest that it plays a vital role in the rescue of UPEC for recurrent cycles of bladder infection.

We have demonstrated the successful utilisation of an up-scaled bladder cell culture model for UTI, coupled with multiplexed TraDIS to determine the essential gene profile set of UPEC harvested from the IBC, dispersal and reversal phases of infection. This approach enabled the first systematic analysis of essential genes required for UPEC to transition through the different phases of bladder infection. This essential gene catalog provides the first comprehensive framework for further characterisation of the UPEC factors needed for UTIs and represents an attractive framework for the development of novel treatments and therapies in the future.

5.4 METHODS

For a description of the general materials and methods used throughout this thesis, refer to *Chapter 2*. The following section will describe the specific methods used in this Chapter.

5.4.1 BACTERIAL STRAINS AND GROWTH CONDITIONS

Bacterial strains were routinely cultured at 37°C on solid or in liquid LB, with 1% (w/v) NaCl, unless otherwise stated. When culturing the UTI89 transposon mutant library (see *Section 4.2* for the construction of the transposon mutant library), an aliquot of the library was thawed at room temperature and used to inoculate fresh LB media and grown overnight at 37°C. In culturing the transposon mutant library, kanamycin (50 μg/mL) was routinely used. A single mutant deletion knockout in UTI89 was constructed for this study and detailed below.

5.4.2 DEVELOPMENT OF AN UP-SCALED IN VITRO UTI MODEL

Human bladder cell line PD07i was cultured at 37°C with 5 % CO_2 to confluency as previously described (see *Section 2.5*). Bladder cells were trypsinized to resuspend them, concentrated to 4 x 10^5 /mL, and then 10 mL was dispensed into each sterile tissue culture dish (BD Falcon, 100 mm diameter x 20 mm depth) and incubated at 37°C with 5 % CO_2 for 24 h. This method below is described at a scale of 1 dish, but the infection can be readily scaled to include multiple dishes. The UTI89 transposon mutant library was routinely used to infect bladder epithelial cells, however in the initial characterisation of the up-scaled model, UTI89/pGI5 (UTI89 wild-type strain with a plasmid that has GFP constituently expressed; kindly provided by Greg Iosifidis, UTS) was used for visualisation purposes.

5.4.2.1 IBC Phase of Infection

An aliquot of the UTI89 transposon mutant library was thawed at room temperature and 100 μL was used to inoculate 10 mL of liquid LB media and grown statically overnight at

37°C. When the transposon mutant library was not used for infection, UTI89/pGI5 was streaked out onto solid LB and a single colony was used to inoculate 10 mL of liquid LB media and grown statically overnight at 37°C. The bacterial culture was then diluted down to OD_{600nm} = 1.0 in PBS and 1 mL of the dilution was used to inoculate each bladder cell culture dish. The culture dishes were then centrifuged in a swinging-bucket rotor suitable for culture plates for 5 min at 500 rcf (x g) and left to incubate statically for 2 h at 37°C with 5 % CO_2. The supernatant was then discarded, and 10 mL pre-warmed EpiLife medium supplemented with HKGS and gentamicin (100 µg/mL) was added. Culture dishes were then incubated statically for 22 h at 37°C with 5 % CO_2. Where necessary, culture dishes were imaged in phase-contrast and GFP fluorescence at 18 and 22 h using a Nikon Ti inverted microscope (Fig. 5.1).

5.4.2.2 Dispersal and Filamentation Phase of Infection

To induce the IBC-dispersal stage of infection and mimic the main lifecycle stages of UPEC in an *in vivo* UTI, the infected bladder cells were exposed to urine from one slightly dehydrated male human donor (*26*) (the urine, collected soon after waking from overnight sleep, was only used if it showed a density of at least 1.025 g/mL). Raw urine was centrifuged at 4,600 rpm for 10 min and the supernatant was then filter sterilised (with a 0.2 µm filter unit). Filtered urine samples were then pooled (a density of 1.026 g/mL was measured for the experiments described here), and stored in 40 mL aliquots at -20°C. Each aliquot was thawed at 37°C before use.

Following 22 h of static growth, as described above, the supernatant was removed and the infection surface culture was gently washed 4 times with PBS. 10 mL of the urine (supplemented with kanamycin 50 µg/mL, when infecting with the transposon library only) was added and the culture dish left to incubate at 37°C with 5 % CO_2 and gentle orbital shaking (~50 rpm, in order to resuspend bacteria released from the bladder cells).

During incubation, the overlay urine was periodically withdrawn for analysis and immediately exchanged for a new sample of pre-warmed urine, after incubation for certain

timeframes (20 or 100 min) during the next 8-20 h. The withdrawn samples were each immediately fixed with 2 % (w/v) formaldehyde (prepared from a paraformaldehyde powder stock) and analysed using a flow cytometer (BD LSRII) and microscope (Zeiss Axioplan 2) to determine cell size and shape. Flow cytometry analysis and gating was performed as previously described in *Chapter 3*. The gate applied to encompass short rod-shaped bacteria in flow cytometry data was determined to encompass 99.0 % of the inoculate culture (i.e. UTI89 grown overnight in liquid LB media). Any cells outside this 99.0 % were deemed as filamentous in all culture samples analysed (Supp. Data Fig. S6). Microscopy was performed as previously described in *Chapter 3*. The samples of collected cells were also measured on a spectrophotometer (Dynamica Halo RB-10) at A_{600nm} to estimate biomass yield (Fig. 5.2I).

5.4.3 MUTANT HARVEST FROM INFECTION FOR TRADIS ANALYSIS

Each sample replicate was made up of 3 technical culture dish replicates that were pooled together and underwent gDNA extraction and purification. Each sample was done in duplicate (2 x 3 technical replicates). Only the UTI89 transposon mutant library was used to infect bladder cells (Inoculate sample) (Fig. 5.3).

5.4.3.1 Pre-infection (inoculate) sample

In order to infect bladder cells, an aliquot of the UTI89 transposon mutant library was thawed at room temperature and 100 µL was used to inoculate 10 mL of liquid LB media and grown statically overnight at 37°C. The bacterial culture was then diluted down to an OD_{600nm} of 1.0 and 1 mL of this was used to inoculate the bladder epithelial cells, as described above. 5 mL of this same culture was also spun down at 4,600 rpm for 10 min. The supernatant was removed and the cell pellet underwent gDNA extraction and purification as previously described (see *Section 2.6*). The DNA was stored at -20°C.

5.4.3.2 IBC sample

After 22 h of static growth as detailed in *Section 5.2.2.1*, the supernatant was removed and the culture gently washed 4 times with PBS. 1 mL of lysis solution was added (see *Section 2.3,* Table 2.5) and incubated for 10 min at 37°C. The entire supernatant was then collected and spun at 4,600 rpm for 10 min. The supernatant was removed and the cell pellet underwent gDNA extraction and purification. The DNA was stored at -20°C.

5.4.3.3 Dispersed bacterial sample

Following 22 h of static growth, the supernatant was removed and the culture dish gently washed 4 times with PBS. 10 mL of concentrated human urine supplemented with kanamycin (50 μg/mL) was added and the culture dish left to incubate for 10 h at 37°C with gentle orbital shaking. Short 20 min windows of urine was collected for the next 10 h, each time re-applying fresh urine to the culture dish. The short windows of collection aimed to minimise reversal and recovery of the mutants in the urine, so that genes required for this subsequent stage would not be selected for the testing of dispersal. The collected samples were placed on ice, spun down at 4,600 rpm and the cell pellets frozen. The next day, the cell pellets underwent gDNA extraction and purification. The DNA from each sample was pooled and stored at -20°C.

5.4.3.4 Non-dispersed bacterial sample

This sample aimed to collect any bacterial cells that did not undergo dispersal (resuspension into the urine) during the 20 h urine growth and may still be inside the bladder cells within an IBC. Upon completion of the 20 min windows of urine collection in the infection samples described for the dispersal sample, 1 mL of lysis solution (Table 2.5) was added to the culture dish and incubated for 10 min at 37°C. The entire supernatant was then collected and spun at 4,600 rpm for 10 min. The supernatant was removed and the cell pellet underwent gDNA extraction and purification. The DNA was stored at -20°C.

5.4.3.5 Recovery samples

At the 10 h time point of urine exposure, urine was collected every 2 h (fresh pre-warmed urine was re-applied after each collection) and concentrated by spinning at 4,600 rpm for 10 min. The supernatant was then removed and the cell pellet was resuspended in 10 mL of liquid LB media, pooled together and left to grow for 18 h overnight at 37°C with 180 rpm shaking. 20 mL of this culture was then spun down at 4,600 rpm for 10 min and the cell pellet underwent gDNA extraction and purification. The DNA was stored at -20°C. In addition, a further 1/500 dilution into liquid LB media of the 18 h recovered culture was done, and the culture incubated for a further 24 h with 180 rpm shaking. This aimed to add an additional selection step for identifying mutants that had recovered. Following the extra 24 h growth, the culture was spun down at 4,600 rpm for 10 min and the cell pellet underwent gDNA extraction and purification.

5.4.4 TRANSPOSON-DIRECTED INSERTION-SITE SEQUENCING

Preparation of DNA for transposon-directed insertion-site sequencing (TraDIS) was performed as previously described (see *Section 4.2*). The list of primers for PCR enrichment can be found in Supp. Data Table S3H. Twenty-two DNA sequencing libraries from the 6 infection samples (Inoculate, IBC, Dispersal, Non-dispersed, Recovery and Recovery-24h in duplicate; Inoculate library in single) were pooled to equimolar concentrations and underwent SPRI-select magnetic bead clean-up (Beckman Coulter) at 0.9-0.5x left and right ratios, respectively (to select DNA of ~200-800 bp). The sequencing libraries were quantified on an Agilent Bioanalyzer and then sequenced on a paired end (PE) Illumina flow cell using an Illumina HiSeq 2500 sequencer for 250 cycles (The Garvan Institute, NSW, Australia).

5.4.5 ANALYSIS OF NUCLEOTIDE SEQUENCING DATA

Sequence reads from the data files were separated according to their unique index "barcodes" and underwent quality control analysis and filtering as previously described

(see *Section 2.6*). Due to low quality nucleotides towards the 3' end of each sequence read, the last 25 bp were trimmed to reveal 100 bp of high-quality basecalls for subsequent mapping. Each sequence read was then trimmed to retain only the last 10 bp of the transposon end sequence at the start of each read. Each read was then filtered by their match to the Tni5 and Tni7 ends of the transposon sequence (the 10 bp region), allowing 0 mismatches. The matching sequence was then trimmed from these reads and the remaining sequence was then strictly mapped to the genome of UTI89 and plasmid pUTI89 using the SMALT short read mapper as described previously in *Chapter 4* and by Barquist *et al*. (*128*) (allowing 0 mismatches). Subsequent analysis steps such as raw read counts and number of insertion sites were also performed as previously described in *Chapter 4* and by Barquist *et al*. (*128*) (Supp. Data Table S3). The sequencing depth of mapped reads and insertion sites were visualised using Artemis, version 13.0 (*131*). The frequency of mapped reads and the analysis of data for Fig. 5.4 were determined with deepTools2 software (*129*).

Using the Bioconductor package edgeR (*185*) and a previously described R protocol (*128*), we identified the UTI89 genes required for survival during infection by comparison of the read abundance across the genome of samples harvested at different time points. The raw read counts from the combined duplicate experimental samples were used as input into edgeR. For essentiality between time points, any genes that contained very low read depth ($\leq$ 5 reads) in at least one time point comparison were removed from the analysis. A criterion of P-value ≤ 0.05 and $Log_2FC < -1.0$ was used to define a list of stringent genes, which may have a reduced capacity to survive at the IBC, dispersal and recovery time points. Any genes that were determined to be required for 5 consecutive days of growth in liquid LB media were eliminated (see *Chapter 4*). The complete raw list of all genes identified through TraDIS can be found in Supp. Data Table S3.

CHAPTER 6

DISCUSSION AND CONCLUSION

6.1 General Discussion

UTIs are among the most common bacterial infections worldwide, accounting for ~150 million cases annually. A range of pathogens have been implicated in UTIs, however strains of UPEC are the predominant etiological agents. UPEC originate within the intestine, but have adapted the ability to disseminate and colonise the urinary tract. Due to the last 25 years of research, we now identify UPEC to be a highly invasive pathogen that undergoes a multi-stage intracellular infection cycle within the cells of the urinary tract. This infection cycle is a complex pathway involving epithelial cell attachment, invasion and intracellular biofilm-like proliferation, leading to the formation of sub-populations of filamentous and highly-motile cells. This accompanies bacterial dispersal and rupture of the host cell (*17, 26*) UPEC filaments have the potential to revert to bacillary rod-shaped morphology and can thereafter divide as normal bacterial cells, thereby initiating a new infection cycle.

In the last decade, it has become clear that filamentation is an essential adaptation for a range of pathogenic bacteria to survive and cause infections in humans (*31*). Reversible bacterial filamentation is now recognised as a wide-spread survival strategy to deal with stressful environments, such as the human urinary tract (*26, 58*). One inducible filamentation mechanism already evident in *E. coli* is the bacterial "SOS" response and induction of SulA (*141*). However, a *sulA*-independent mechanism of UPEC filamentation exists during infection of bladder cells (*56*). One of these pathways in UPEC likely requires the cell division gene, *damX*, which was identified by Khandige *et al.* (*56*) as being essential for the filamentation response during *in vitro* UTI. However, the exact mechanism of how this is regulated and the complete suite of genes that are concurrently involved in the pathway still remain undetermined. There also exists the possibility of both *sulA* and *damX* independent genetic pathways under different conditions.

UPEC must also adapt to new nutritional standings as they transition from the intestine to the urinary tract (*19*). Considering the different nutrient availabilities, metabolites within the urine and bladder are most likely to activate metabolic adaptive strategies in UPEC in

order to survive and cause infection (*28*). The metabolism of UPEC is tightly regulated and highly responsive to fluctuating nutrient availability. These are the physiological factors that do not directly damage the host, but nevertheless are essential for UPEC growth and survival during infections. The capacity for UPEC to utilise nutritionally-diverse environments, such as the intestines, extracellular bladder lumen and intracellular bladder space clearly plays a significant role in pathogenesis and this is only just starting to become appreciated (*28*).

In recent times, the rapid emergence of antibiotic-resistant UTIs has greatly influenced the severity of infection and financial cost burden to the health-care system, costing up to USD$6 billion on treatment each year (*5*). With the ongoing rise in antibiotic-resistant UTIs, the severity of infection and amounting burden on the health-care sector is set to escalate. This further complicates UTI therapies and highlights the urgent need to advance our understanding of the biological mechanisms and requirements underpinning bacterial survival and pathogenesis.

The general theme of this PhD thesis was the expansion of existing knowledge surrounding UPEC survival and morphological plasticity (cell shape changes). This was addressed in both pure culture conditions (*Chapter 3* and *Chapter 4*) and UPEC infection of human bladder cells (*Chapter 5*). Each Chapter in this thesis focused on distinct aspects of this overarching theme. Firstly, *Chapter 3* aimed to determine mediators of reversible filamentation in *E. coli* by high-throughput sequencing of phenotypically enriched UTI89 DNA expression libraries. Secondly, *Chapter 4* focused on constructing a transposon mutant library in UTI89 to determine the genetic requirements of survival and growth in nutrient-limited culture medium by using the transposon-directed insertion-site sequencing (TraDIS) technique. Lastly, *Chapter 5* aimed to extend and connect previous results to the UTI model by incorporating the transposon mutant library and TraDIS technique to identify genes that are required for UTI89 to progress through the distinct phases of bladder cell infection.

6.1.1 REVERSIBLE FILAMENTATION IN UPEC

To comprehensively determine the genes in UTI89 that can cause filamentation, we utilised two genome-wide screening approaches, one of which was under controlled laboratory culture conditions (*Chapter 3*) and the other during *in vitro* urinary tract infection (*Chapter 5*). In *Chapter 3*, we reported the development and utilisation of a widely applicable, high-throughput DNA sequencing approach to identify genes, or genomic fragments, that inhibit cell division after induced over-expression from a plasmid-based DNA expression library. These results, when used in combination with the TraDIS screen (*Chapter 5*), which defined the UTI89 gene set required for the dispersal phase of infection and subsequent recovery and reversal phase, will provide good candidate genes involved in the UTI-filamentation response pathway. The combination of two genome-wide screening approaches aims to identify genes with differing properties that may play a role at varying steps in the filamentation pathway (i.e. promoters and/or negative regulators of filamentation).

The results from both screening approaches identified known and novel inhibitors or regulators of cell division. Several previously uncharacterised genes were also identified that may represent new regulators of division, or be part of the UTI-associated filamentation response.

An intriguing overlap between *Chapter 3* and *Chapter 5* results was the strong identification of genes that have known roles in central carbon metabolism. Previous carbon metabolic pathways (e.g. UDP-glucose) have been implicated in regulating cell division in *E. coli* by acting on FtsZ to coordinate cell size with nutrient availability (*148*). This seems to be a common response strategy in fluctuating nutrient conditions to regulate the frequency of division, and at times even inhibit cell division transiently. The ability of *E. coli* to monitor nutritional availability and transmit this information to the division machinery is vital. This nexus of nutrient availability coupled to cell division and growth also plays an important role in *B. subtilis* (*150*). It was reported that nutrient-rich conditions (high pyruvate availability) positively regulates FtsZ assembly and division to

appropriately couple division with an increased growth rate (*150*). However, much is still unclear about these proteins acting as both metabolic enzymes and regulators of FtsZ and division. Sperber *et al*. (*199*) suggests that some of these proteins may utilise interactions with divisome proteins (i.e. FtsZ) to localise their metabolic activities to sites most favorable for their substrate to become metabolized (*199*). Despite the exact mechanism remaining unknown, understanding the regulatory cross talks that occur between metabolic pathways and cell division would provide considerable insights into bacterial physiology and plasticity. In regards to UTIs, given the UPEC morphological changes observed during infection (*17*), it seems likely that UPEC continually fine-tune their division status via multiple checkpoints and signaling pathways that frequently monitor the nutrient availability within the bladder. The most obvious cell shape response, reversible filamentation, may also be a nutritional response, but the exact trigger for this remarkable cell shape change is unclear. Many of the metabolic genes identified in our screens have not previously been implicated in metabolic-division control and could play a role in these regulatory mechanisms. Those metabolic genes that have previously been implicated in metabolic-division control in *E. coli* and *B. subtilis* (e.g. *galU*, *pdhR*) may act in a similarly regulatory manner in UPEC during bladder cell dispersal.

Another overarching result from both *Chapter 3* and *Chapter 5* screening approaches was the identification of a large number of prophage genes linked to filamentation. Several prophage genes have previously been reported to arrest cell division in *E. coli* by negatively acting on FtsZ expression and assembly (*152, 155, 200, 201*). Known prophage toxins from toxin-antitoxin pairs affecting cell division were also identified. These results may reflect a common function of prophage-associated genes to regulate bacterial cell division during lysis, which may promote bacteriophage proliferation and dissemination. At least some of the genes in remnant or defective prophages appear to have been co-opted by UPEC for regulatory purposes in stress responses leading to reversible cell division arrest (*152*). This may potentially be occurring in UTIs, and our results implicate these well-characterised prophage genes in bladder infection and shed further light on unknown prophage regulatory genes.

6.1.2 METABOLIC ADAPTATIONS AND REQUIREMENTS OF UPEC

To determine the genes in UTI89 that are required for metabolic flexibility and robustness, we created a transposon mutant library and performed TraDIS to first identify the genes important for growth in nutrient-restricted media (*Chapter 4*), and then during the intracellular growth phase of bladder infection (IBC phase; *Chapter 5*). In *Chapter 4*, we identified 60 UPEC mutants with a significant fitness defect and reduced capacity to survive in M9-glycerol minimal-media (glycerol sole carbon source) when compared to LB media (Table 4.1). Once again, these results, when used in combination with *Chapter 5*, which defined the UTI89 gene set required for the IBC phase of infection when compared to the inoculate reference culture (Table 5.1), will provide the strongest implicated list of genes involved in UPEC metabolic adaptation and response pathways. The rationale for the initial investigation into the genes required for growth in minimal-media was to first characterise the TraDIS technique, determine the robustness of TraDIS in an infection model situation and provide preliminary candidate genes that UPEC may find essential through *de novo* biosynthesis of resources that are not present in rich-complex media. When taking these parameters into account, the large number of genes and functional classes that were identified in common between *Chapter 4* and *Chapter 5* demonstrates the validity of our TraDIS method and relevance to bladder infection. A strong overlap was seen in the identification of the *pur*, *pyr*, *his* and *cys* genes involved in amino acid and nucleotide *de novo* synthesis. Similar to minimal-media, this suggests that the bladder intracellular space is limited in these metabolic precursor and biosynthesis of these resources by UPEC is essential for growth and survival. It is also noted that selective amino acid uptake and catabolism is required for UPEC growth within the bladder cell. This was supported with the verification of the *hisF* (histidine biosynthesis) gene, which we identified and verified as being required for UPEC progression to 24 h post-infection (IBC phase). This confirms that nucleotide sugars and basic amino acids are limited in the intracellular bladder niche, which force UPEC to rely on certain biosynthetic pathways in order to effectively colonise the host cell. The *de novo* synthesis of histidine by UPEC is supported by the upregulation of genes involved in histidine transport during growth in urine when compared to liquid LB medium. Our results shed new light on the importance

of UPEC to undergo *de novo* biosynthesis of amino acids and implicate *hisF* for the first time to be required for intracellular bladder colonisation.

6.1.3 FUTURE DIRECTIONS

Although the work presented in this thesis addressed each aim and produced results, which has brought to light new information and expanded existing knowledge on UPEC survival and morphological plasticity, further work can be carried out in the future.

Within *Chapter 3* of this thesis, we identified 55 genomic regions that were enriched in the replicate screen experiments (Table 3.2), of which, six genomic regions were tested further for confirmatory purposes. In the future, further genomic regions should be explored and verified to cause filamentation, while also considering that alternative ORFs, sub-fragments of an ORF, or an expressed intragenic region can cause cell division arrest (Fig. 3.4), albeit artificial (experimentally induced) cell division arrest. This was highlighted by the identification of a number of short predicted peptides, which are not expected to have evolved with the purpose of causing filamentation relevant to UPEC pathogenesis, but could be used as synthetic, artificial inhibitors of cell division. DNA sub-fragments such as these should be pursued in future work as they represent novel peptide or RNA inhibitors of cell division and potential anti-microbial agents. Synthesised peptides (as these are more stable than RNA) can initially be added to the culture media during exponential growth and analysed for any cell division defects via microscopy. Further work can involve removing translation of the cloned UTI89 gene fragment in pBAD24, observing any cell division defects and confirming mRNA fragments using Northern Blot or qPCR. It is also expected that expanded screens of larger and more diverse DNA libraries, could reveal further new and potentially useful novel inhibitors of bacterial division and also identify new sequences encoding the capacity to induce filamentation in other uropathogens. These results would build a solid framework for novel drug treatments for UTIs in the future.

Transposon mutagenesis in combination with the TraDIS technique was utilised in *Chapter 4* and *Chapter 5* to define the genes required for growth of UPEC in pure culture conditions

and during *in vitro* infection. Those mutants that were identified in common, focusing on the IBC phase of infection results (i.e. *hisF*), should be explored and verified further in the mouse model of UTI. At the completion of *Chapter 4*, we implicated *hisF* and *neuC* as a requirement for UPEC to progress to the IBC phase of infection (24 h post-infection) (Fig. 4.5), but did not rule out the possibility that the mutants could have an attenuated attachment or invasion capacity. Initial work surrounding *hisF* and *neuC* will be to investigate the mutant's ability to attach and invade bladder epithelial cells *in vitro*.

Additionally, we identified 333 and 323 significant genes that are required for UPEC during the dispersal and recovery phases, respectively (Table 5.2 and Table 5.3). Many genes within this list are cell division regulators, which should be verified for their role in the UTI-filamentation response pathway during *in vitro* and *in vivo* infections. Subsequent future work will aim to elucidate the stage of divisome assembly that is inhibited and determine the subcellular localisation of candidate genes in the filamentation pathway using immunofluorescence microscopy and/or GFP-tagged microscopy.

6.1.4 CONCLUDING REMARKS

A high-throughput sequencing-based method for genome-wide identification of genes and genomic DNA fragments that encode a capacity to induce filamentation in *E. coli* was successfully developed and applied. This revealed genes from several prophages, carbon metabolic pathways, as well as endogenous bacterial genes or loci that have known and novel roles in cell division or filamentation. Additionally, we successfully applied TraDIS to determine the genes required for growth and survival of UPEC in nutrient-restricted media, as well as determine those genes required for distinct phases of bladder cell infection. We implicated and verified *hisF* and *neuC* as a requirement for UPEC to progress to the IBC phase of infection. The catalog of genes identified through our TraDIS screens provide a near-comprehensive foundation for further characterisation of UPEC factors needed for survival and morphological plasticity. This represents an attractive framework for the development of novel treatments and therapies for UTIs in the future.

.

[blank page]

APPENDIX

SUPPLEMENTARY DATA

Supplementary Data is provided within this PhD thesis. *Chapters 3-5* have Supplementary Data Figures to support their results and these are listed below.

Chapter 3:

Supp. Data Figure S1 Additional analyses of DNA sequencing data.

Supp. Data Figure S2 *pptE* overexpression causes major effects on cell structure.

Supp. Data Figure S3 ORFs that do not cause substantial filamentation when overexpressed.

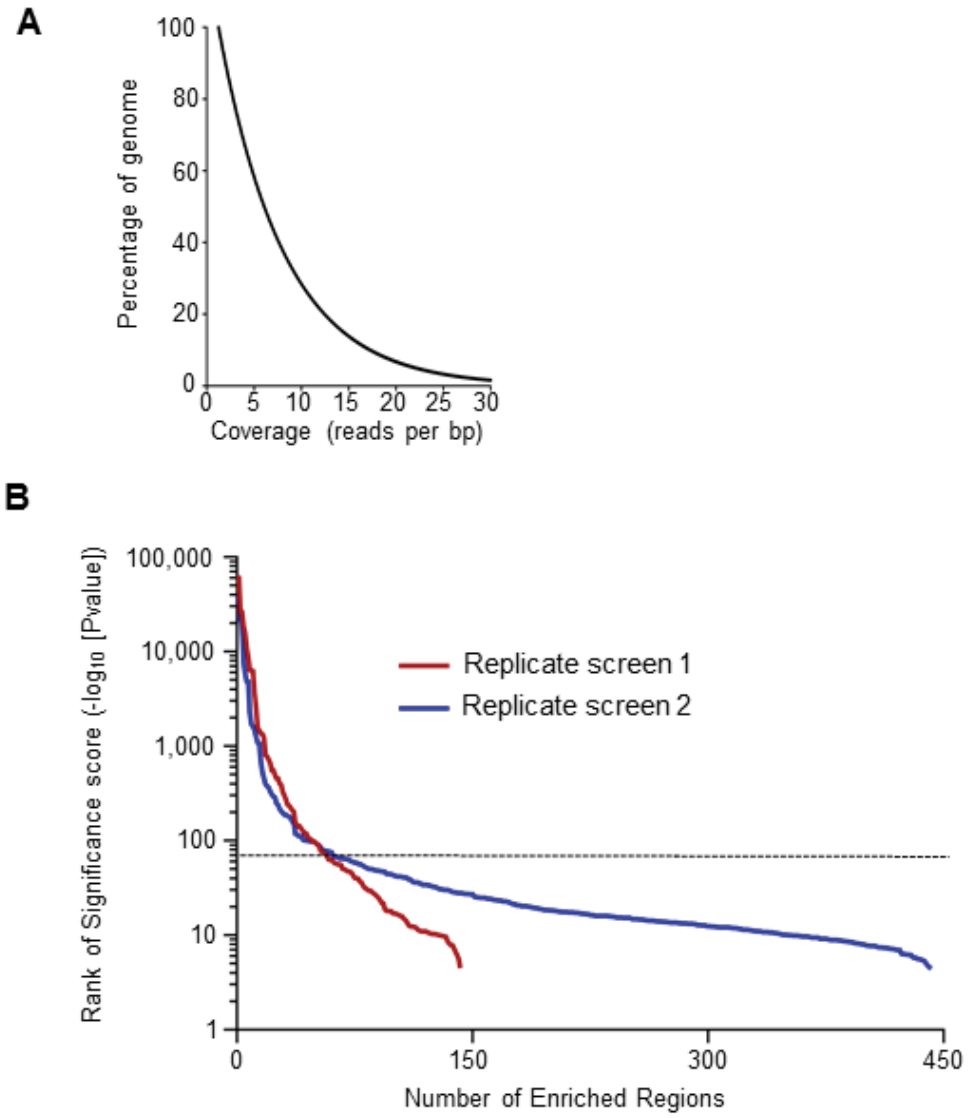
Chapter 4:

Supp. Data Figure S4 Outline of the modified TraDIS DNA sequencing protocol.

Supp. Data Figure S5 Correlation between the Tni5 and Tni7 DNA sequencing ends.

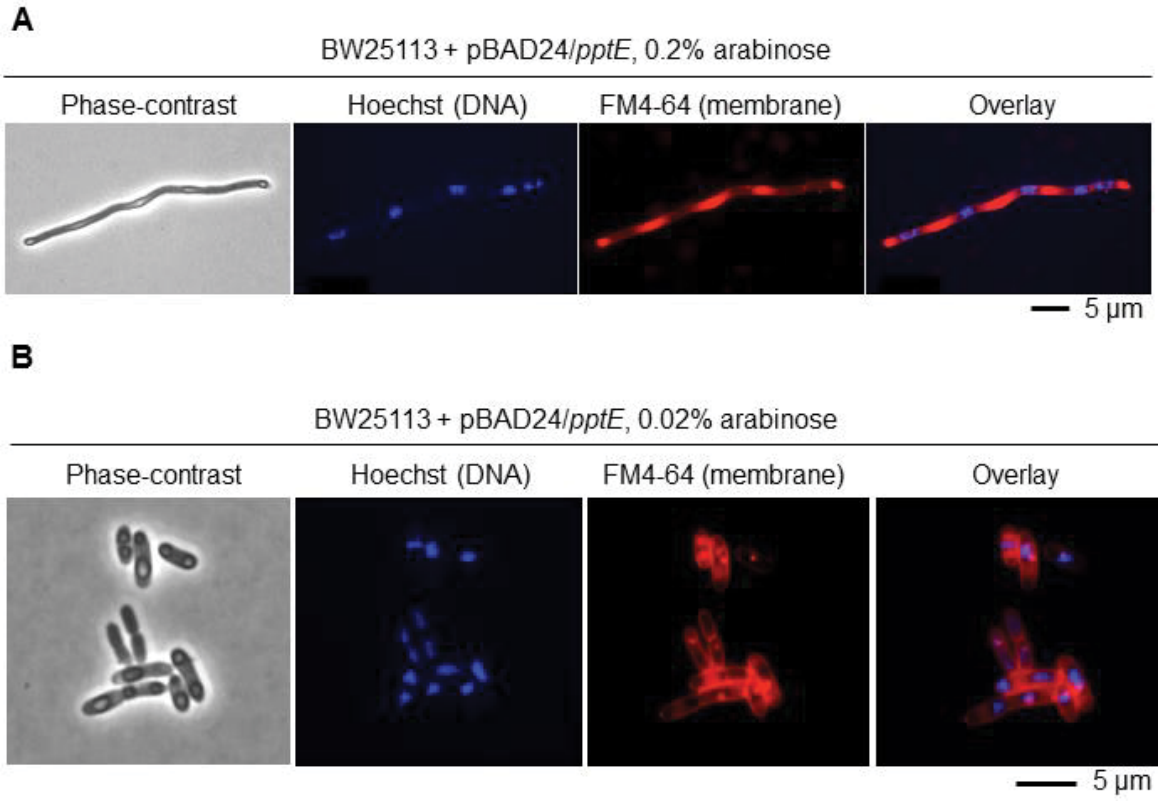
Chapter 5:

Supp. Data Figure S6 Analysis of filamentous bacteria from the dispersal phase of infection.



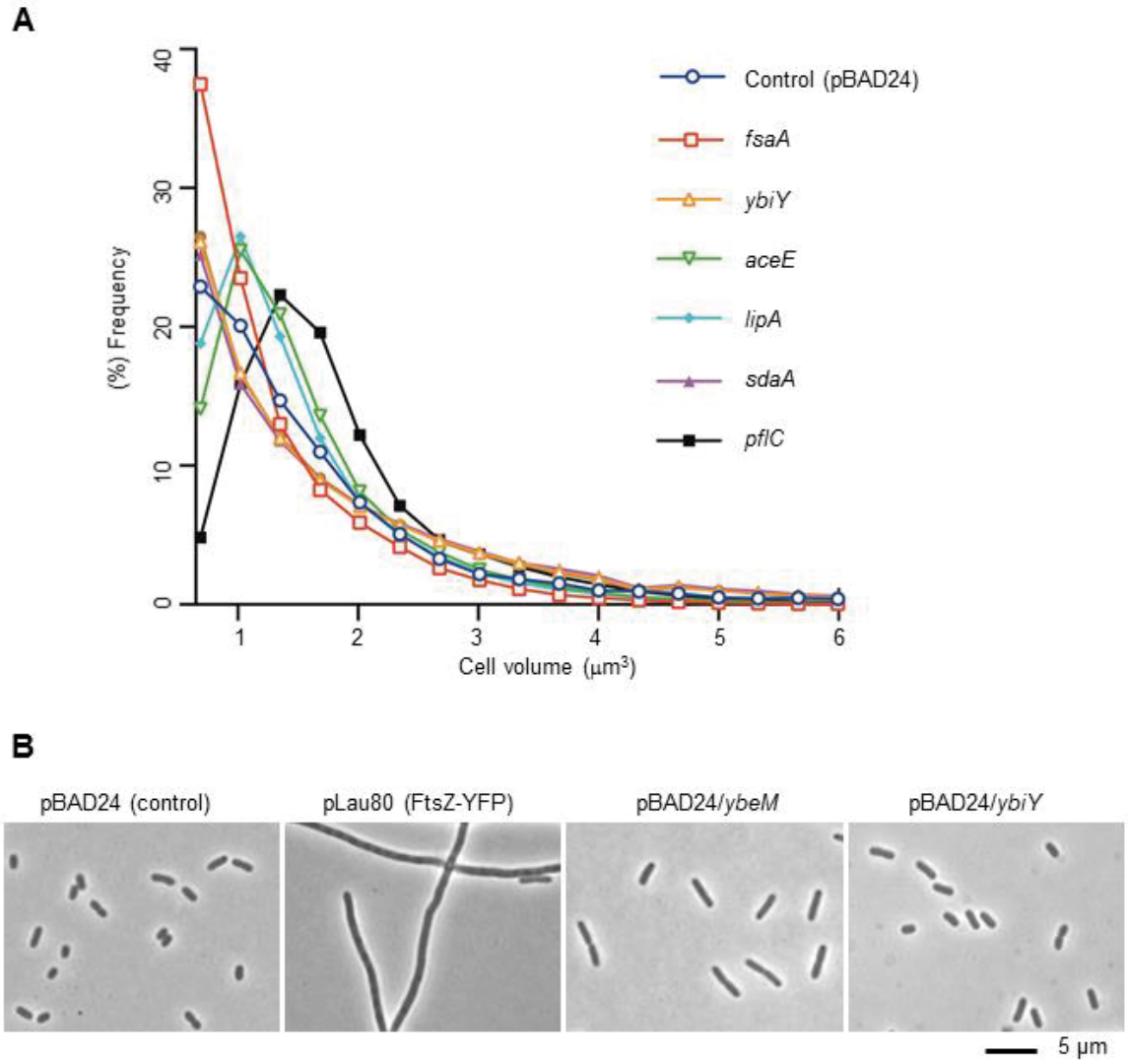
Supplementary Data Figure S1. Additional analyses of DNA sequencing data.

(A) Plot showing the global read coverage of the UTI89 chromosome of the UTI89 genomic library, based on the read data from the "reference-unsorted 2" sample. The library effectively covers the complete genome of UTI89, at a depth of 1-30 reads per bp. (B) A plot comparing the number of identified enriched regions in replicate screens 1 and 2 by the MACS peak detection software. The number of identified enriched regions were ranked by their significance score (-log_{10}[P value]), and the significance scores plotted against raw numbers of the identified regions. The dotted line represents the threshold of ≥ 70 P-value as the high-stringency criterion used to generate Table 3.2.



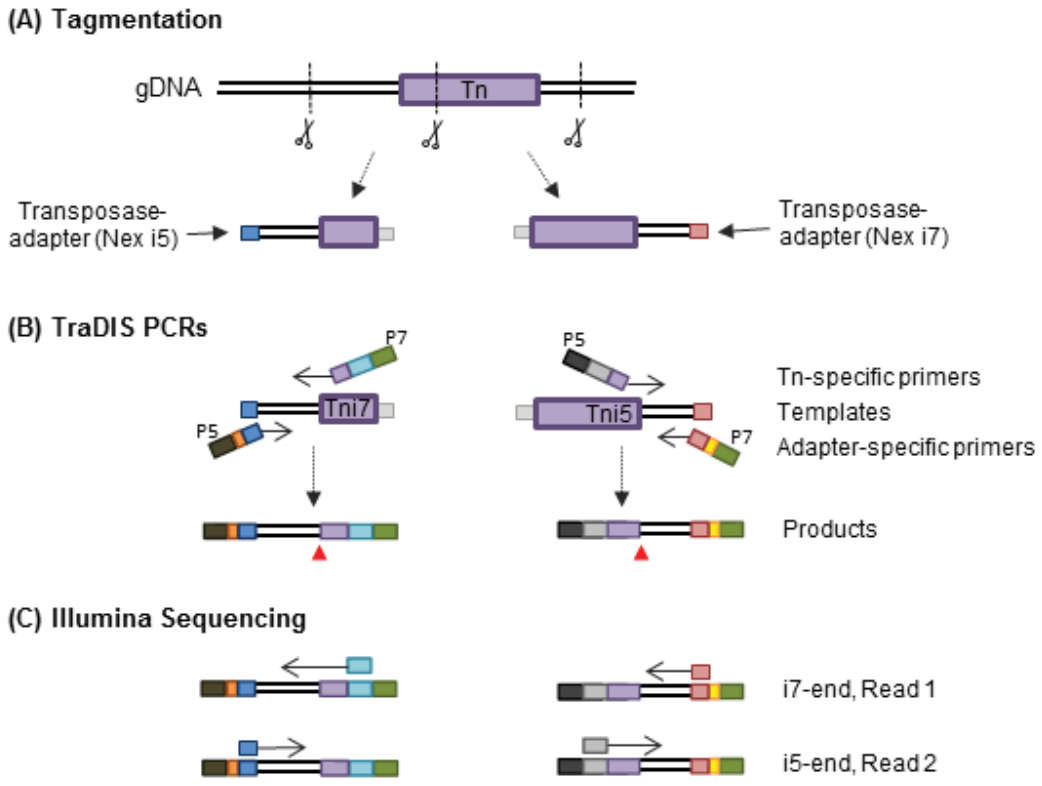
Supplementary Data Figure S2. pptE over-expression causes major effects on cell structure.

BW25113+pBAD24/pptE was grown in M9 minimal medium and induced with L-arabinose at the indicated concentrations and then fixed and stained with Hoechst 33342 (DNA) and FM4-64 (membrane) stains and then visualised by fluorescence microscopy. (A) 0.2% L-arabinose. (B) 0.02% L-arabinose. The intracellular inclusion-like structures and FM4-64 staining abnormalities were observed at both induction concentrations, whereas significant filamentation was only observed at the higher L-arabinose concentration.



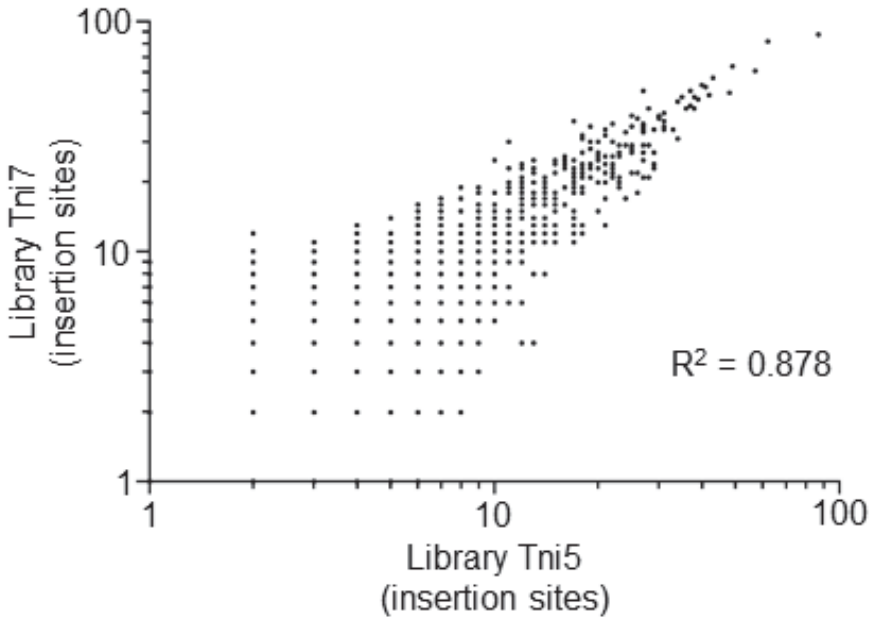
Supplementary Data Figure S3. ORFs that do not cause substantial filamentation when overexpressed.

The ORFs were expressed from pBAD24 in BW25113 by induction with 0.2% L-arabinose in M9 medium. Cells were fixed at OD_{600} = 0.8. (A) Cell volume distributions of the cell populations were determined by Coulter cytometry. The pflC expression strain shows a mild cell division defect or delay. (B) BW25113 + pBAD24/ybeM and BW25113 + pBAD24/ybiY strains examined by phase-contrast microscopy. Expression of ybeM appears to cause a minor effect on cell length, compared to the control and ybiY expression strains.



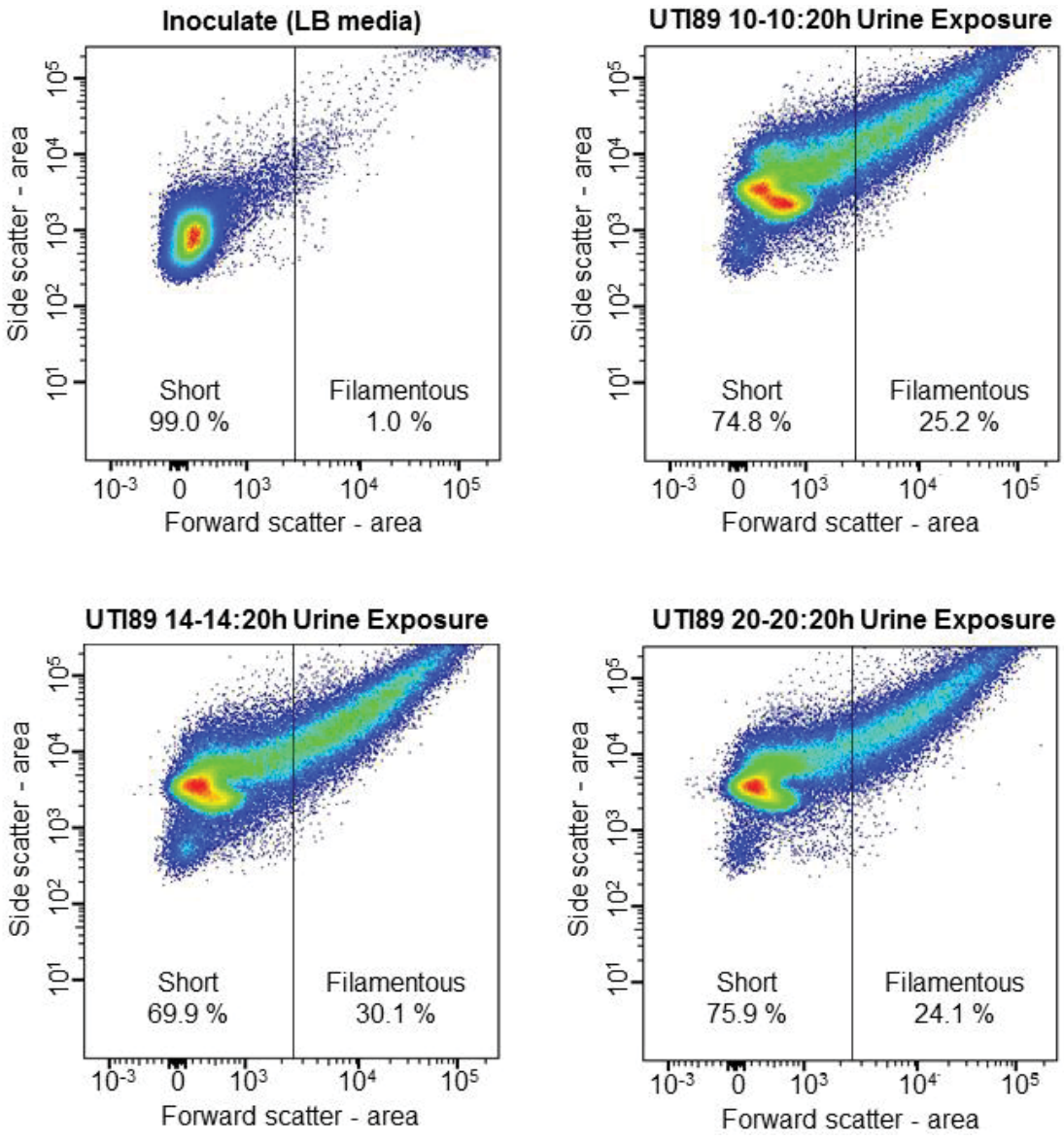
Supplementary Data Figure. S4. Outline of the modified TraDIS DNA sequencing protocol.

(A) A genomic region containing a transposon (Tn) insertion site is represented here. Genomic DNA (gDNA) is first randomly fragmented via "tagmentation". In this reaction, the transposase cuts dsDNA and attaches standard Nextera-Illumina adapters (Nex i5, blue, or Nex i7, red) to the ends of the fragmented DNA. (B) Each end of the transposon is designated ("Tni5" and "Tni7") by respective Tn-specific outward-directed primers that incorporate the appropriate 5' Illumina flow-cell adapter (P5 or P7; dark grey and dark green, respectively) and read primer binding site (read 1 – cyan, and read 2 – light grey). For PCR, the opposing primer anneals to the Nex i5 or Nex i7 adapters (blue and red, respectively). These primers also contain unique barcodes to identify each sample (orange and yellow), and the alternate flow cell binding sequences (dark grey or dark green). The TraDIS PCR amplifies the transposon-gDNA junctions (indicated by red arrow-heads in the PCR products). (C) Paired-end sequencing is then carried out with standard Illumina workflows; multiplex sequencing of samples amplified from both ends of the transposon ensures high sequence diversity during both read 1 and read 2. For data analysis, high-quality sequence reads containing the expected transposon end sequence are selected, precisely identifying the collection of insertion sites.



Supplementary Data. Figure. S5. Correlation between the Tni5 and Tni7 sequencing ends of the Library sample.

The number of insertion sites per gene in UTI89 determined from the Tni5 and Tni7 ends of the transposon. The insertion site counts on the Tni7 end was plotted against those on the Tni5 end. The R squared correlation was identified as 0.878 between Tni5 and Tni7, indicating strong correlation between both sequencing ends.



Supplementary Data. Figure. S6. Analysis of filamentous bacteria from the dispersal phase of infection.

Flow cytometry was carried out with the UTI89 inoculate culture (grown in LB media) used as the control population. Bacteria collected at 10, 14 and 20 h time points from the dispersal phase of infection (grown in urine) were analysed as previously described (Chapter 3). Dots represent single events plotted as side-scatter values of peak height and peak width, which correlates with filament length. Event density is colour coded with a heat-map of red (high density) to blue (displaying individual events). The gates for short and filamentous populations are indicated with vertical lines, and the percentages of events contained within that gate are indicated. The "short" gate was determined to encompass greater than 99% of the control bacterial population (short cells), Inoculate (LB media).

Supplementary Data is provided with this PhD thesis online. *Chapters 3-5* have Supplementary Data Tables to support their results and these are listed below.

Supp. Data Table S1 Supplementary Data for the work in Chapter 3.
Supp. Data Table S2 Supplementary Data for the work in Chapter 4.
Supp. Data Table S3 Supplementary Data for the work in Chapter 5.

References

1. B. Foxman, Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *The American journal of medicine* **113 Suppl 1A**, 5S-13S (2002).
2. W. E. Stamm, S. R. Norrby, Urinary tract infections: disease panorama and challenges. *J Infect Dis* **183 Suppl 1**, S1-4 (2001).
3. B. Foxman, Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Dis Mon* **49**, 53-70 (2003).
4. B. Foxman, P. Brown, Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. *Infect Dis Clin North Am* **17**, 227-241 (2003).
5. A. L. Flores-Mireles, J. N. Walker, M. Caparon, S. J. Hultgren, Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol* **13**, 269-284 (2015).
6. T. M. Hooton, Clinical practice. Uncomplicated urinary tract infection. *N Engl J Med* **366**, 1028-1037 (2012).
7. P. Lichtenberger, T. M. Hooton, Complicated urinary tract infections. *Curr Infect Dis Rep* **10**, 499-504 (2008).
8. F. M. Wagenlehner *et al.*, Diagnosis and management for urosepsis. *Int J Urol* **20**, 963-970 (2013).
9. M. G. Blango, M. A. Mulvey, Persistence of uropathogenic Escherichia coli in the face of multiple antibiotics. *Antimicrob Agents Chemother* **54**, 1855-1863 (2010).
10. T. Escherich, The intestinal bacteria of the neonate and breast-fed infant. 1885. *Rev Infect Dis* **11**, 352-356 (1989).
11. G. R. Nielubowicz, H. L. Mobley, Host-pathogen interactions in urinary tract infection. *Nat Rev Urol* **7**, 430-441 (2010).
12. A. Ronald, The etiology of urinary tract infection: traditional and emerging pathogens. *Am J Med* **113 Suppl 1A**, 14S-19S (2002).
13. B. Foxman, Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infect Dis Clin North Am* **28**, 1-13 (2014).
14. J. Bien, O. Sokolova, P. Bozko, Role of Uropathogenic Escherichia coli Virulence Factors in Development of Urinary Tract Infection and Kidney Damage. *Int J Nephrol* **2012**, 681473 (2012).
15. T. J. Wiles, R. R. Kulesus, M. A. Mulvey, Origins and virulence mechanisms of uropathogenic Escherichia coli. *Exp Mol Pathol* **85**, 11-19 (2008).
16. S. N. Smith, E. C. Hagan, M. C. Lane, H. L. Mobley, Dissemination and systemic colonization of uropathogenic Escherichia coli in a murine model of bacteremia. *MBio* **1**, (2010).
17. S. S. Justice *et al.*, Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1333-1338 (2004).
18. O. Gal-Mor, B. B. Finlay, Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell Microbiol* **8**, 1707-1719 (2006).
19. C. J. Alteri, H. L. Mobley, Escherichia coli physiology and metabolism dictates adaptation to diverse host microenvironments. *Current opinion in microbiology* **15**, 3-9 (2012).

20. C. J. Alteri, S. N. Smith, H. L. Mobley, Fitness of Escherichia coli during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathog* **5**, e1000448 (2009).
21. C. J. Alteri, H. L. Mobley, Metabolism and Fitness of Urinary Tract Pathogens. *Microbiol Spectr* **3**, (2015).
22. C. Svanborg *et al.*, Uropathogenic Escherichia coli as a model of host-parasite interaction. *Curr Opin Microbiol* **9**, 33-39 (2006).
23. A. K. Moller *et al.*, An Escherichia coli MG1655 lipopolysaccharide deep-rough core mutant grows and survives in mouse cecal mucus but fails to colonize the mouse large intestine. *Infect Immun* **71**, 2142-2152 (2003).
24. T. Conway, P. S. Cohen, Commensal and Pathogenic Escherichia coli Metabolism in the Gut. *Microbiol Spectr* **3**, (2015).
25. I. U. Mysorekar, S. J. Hultgren, Mechanisms of uropathogenic Escherichia coli persistence and eradication from the urinary tract. *Proc Natl Acad Sci U S A* **103**, 14170-14175 (2006).
26. T. E. Andersen *et al.*, Escherichia coli uropathogenesis in vitro: invasion, cellular escape, and secondary infection analyzed in a human bladder cell infection model. *Infection and immunity* **80**, 1858-1867 (2012).
27. J. J. Martinez, M. A. Mulvey, J. D. Schilling, J. S. Pinkner, S. J. Hultgren, Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *Embo J* **19**, 2803-2812 (2000).
28. R. Mann, D. G. Mediati, I. G. Duggin, E. J. Harry, A. L. Bottomley, Metabolic Adaptations of Uropathogenic E. coli in the Urinary Tract. *Front Cell Infect Microbiol* **7**, 241 (2017).
29. M. A. Mulvey, J. D. Schilling, S. J. Hultgren, Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. *Infection and immunity* **69**, 4572-4579 (2001).
30. D. A. Rosen, T. M. Hooton, W. E. Stamm, P. A. Humphrey, S. J. Hultgren, Detection of intracellular bacterial communities in human urinary tract infection. *PLoS medicine* **4**, e329 (2007).
31. S. S. Justice, D. A. Hunstad, L. Cegelski, S. J. Hultgren, Morphological plasticity as a bacterial survival strategy. *Nature reviews. Microbiology* **6**, 162-168 (2008).
32. C. S. Eden, L. A. Hanson, U. Jodal, U. Lindberg, A. S. Akerlund, Variable adherence to normal human urinary-tract epithelial cells of Escherichia coli strains associated with various forms of urinary-tract infection. *Lancet* **1**, 490-492 (1976).
33. S. Langermann *et al.*, Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic Escherichia coli. *J Infect Dis* **181**, 774-778 (2000).
34. I. Connell *et al.*, Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. *Proc Natl Acad Sci U S A* **93**, 9827-9832 (1996).
35. D. S. Eto, T. A. Jones, J. L. Sundsbak, M. A. Mulvey, Integrin-mediated host cell invasion by type 1-piliated uropathogenic Escherichia coli. *PLoS Pathog* **3**, e100 (2007).
36. E. V. Sokurenko, H. S. Courtney, S. N. Abraham, P. Klemm, D. L. Hasty, Functional heterogeneity of type 1 fimbriae of Escherichia coli. *Infect Immun* **60**, 4709-4719 (1992).

37. T. T. Sun, H. Zhao, J. Provet, U. Aebi, X. R. Wu, Formation of asymmetric unit membrane during urothelial differentiation. *Mol Biol Rep* **23**, 3-11 (1996).
38. M. A. Mulvey, Adhesion and entry of uropathogenic Escherichia coli. *Cell Microbiol* **4**, 257-271 (2002).
39. R. A. Welch *et al.*, Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **99**, 17020-17024 (2002).
40. J. A. Roberts *et al.*, The Gal(alpha 1-4)Gal-specific tip adhesin of Escherichia coli P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proc Natl Acad Sci U S A* **91**, 11889-11893 (1994).
41. H. Leffler, C. Svanborg-Eden, Glycolipid receptors for uropathogenic Escherichia coli on human erythrocytes and uroepithelial cells. *Infect Immun* **34**, 920-929 (1981).
42. M. A. Mulvey, J. D. Schilling, J. J. Martinez, S. J. Hultgren, Bad bugs and beleaguered bladders: interplay between uropathogenic Escherichia coli and innate host defenses. *Proc Natl Acad Sci U S A* **97**, 8829-8835 (2000).
43. M. A. Mulvey *et al.*, Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. *Science (New York, N.Y.)* **282**, 1494-1497 (1998).
44. V. Braun, F. Niedergang, Linking exocytosis and endocytosis during phagocytosis. *Biol Cell* **98**, 195-201 (2006).
45. B. Wankel *et al.*, Sequential and compartmentalized action of Rabs, SNAREs, and MAL in the apical delivery of fusiform vesicles in urothelial umbrella cells. *Mol Biol Cell* **27**, 1621-1634 (2016).
46. D. S. Eto, J. L. Sundsbak, M. A. Mulvey, Actin-gated intracellular growth and resurgence of uropathogenic Escherichia coli. *Cell Microbiol* **8**, 704-717 (2006).
47. R. E. Berry, D. J. Klumpp, A. J. Schaeffer, Urothelial cultures support intracellular bacterial community formation by uropathogenic Escherichia coli. *Infection and immunity* **77**, 2762-2772 (2009).
48. T. J. Hannan *et al.*, Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic Escherichia coli bladder infection. *FEMS Microbiol Rev* **36**, 616-648 (2012).
49. A. J. Lewis, A. C. Richards, M. A. Mulvey, Invasion of Host Cells and Tissues by Uropathogenic Bacteria. *Microbiol Spectr* **4**, (2016).
50. N. L. Barbieri *et al.*, FNR regulates expression of important virulence factors contributing to pathogenicity of uropathogenic Escherichia coli. *Infect Immun* **82**, 5086-5098 (2014).
51. R. C. Fink *et al.*, FNR is a global regulator of virulence and anaerobic metabolism in Salmonella enterica serovar Typhimurium (ATCC 14028s). *J Bacteriol* **189**, 2262-2273 (2007).
52. G. G. Anderson *et al.*, Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* **301**, 105-107 (2003).
53. D. J. Schwartz, S. L. Chen, S. J. Hultgren, P. C. Seed, Population dynamics and niche distribution of uropathogenic Escherichia coli during acute and chronic urinary tract infection. *Infection and immunity* **79**, 4250-4259 (2011).
54. F. R. Blattner *et al.*, The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462 (1997).

55. V. Hancock, R. M. Vejborg, P. Klemm, Functional genomics of probiotic Escherichia coli Nissle 1917 and 83972, and UPEC strain CFT073: comparison of transcriptomes, growth and biofilm formation. *Mol Genet Genomics* **284**, 437-454 (2010).
56. S. Khandige *et al.*, DamX Controls Reversible Cell Morphology Switching in Uropathogenic Escherichia coli. *MBio* **7**, (2016).
57. J. M. Davis, S. B. Rasmussen, A. D. O'Brien, Cytotoxic necrotizing factor type 1 production by uropathogenic Escherichia coli modulates polymorphonuclear leukocyte function. *Infect Immun* **73**, 5301-5310 (2005).
58. S. S. Justice, D. A. Hunstad, P. C. Seed, S. J. Hultgren, Filamentation by Escherichia coli subverts innate defenses during urinary tract infection. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19884-19889 (2006).
59. E. Harry, L. Monahan, L. Thompson, Bacterial cell division: the mechanism and its precison. *Int Rev Cytol* **253**, 27-94 (2006).
60. D. W. Adams, J. Errington, Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nat Rev Microbiol* **7**, 642-653 (2009).
61. E. F. Bi, J. Lutkenhaus, FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161-164 (1991).
62. E. Mileykovskaya, Q. Sun, W. Margolin, W. Dowhan, Localization and function of early cell division proteins in filamentous Escherichia coli cells lacking phosphatidylethanolamine. *J Bacteriol* **180**, 4252-4257 (1998).
63. R. B. Weart *et al.*, A metabolic sensor governing cell size in bacteria. *Cell* **130**, 335-347 (2007).
64. X. Ma, D. W. Ehrhardt, W. Margolin, Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living Escherichia coli cells by using green fluorescent protein. *Proc Natl Acad Sci U S A* **93**, 12998-13003 (1996).
65. H. C. Wang, R. C. Gayda, High-level expression of the FtsA protein inhibits cell septation in Escherichia coli K-12. *J Bacteriol* **172**, 4736-4740 (1990).
66. S. Pichoff, J. Lutkenhaus, Unique and overlapping roles for ZipA and FtsA in septal ring assembly in Escherichia coli. *Embo J* **21**, 685-693 (2002).
67. C. A. Hale, P. A. de Boer, Direct binding of FtsZ to ZipA, an essential component of the septal ring structure that mediates cell division in E. coli. *Cell* **88**, 175-185 (1997).
68. C. A. Hale, A. C. Rhee, P. A. de Boer, ZipA-induced bundling of FtsZ polymers mediated by an interaction between C-terminal domains. *J Bacteriol* **182**, 5153-5166 (2000).
69. A. Dajkovic, S. Pichoff, J. Lutkenhaus, D. Wirtz, Cross-linking FtsZ polymers into coherent Z rings. *Mol Microbiol* **78**, 651-668 (2010).
70. E. Small *et al.*, FtsZ polymer-bundling by the Escherichia coli ZapA orthologue, YgfE, involves a conformational change in bound GTP. *J Mol Biol* **369**, 210-221 (2007).
71. E. Galli, K. Gerdes, Spatial resolution of two bacterial cell division proteins: ZapA recruits ZapB to the inner face of the Z-ring. *Mol Microbiol* **76**, 1514-1526 (2010).
72. J. Buss *et al.*, In vivo organization of the FtsZ-ring by ZapA and ZapB revealed by quantitative super-resolution microscopy. *Mol Microbiol* **89**, 1099-1120 (2013).

73. E. Galli, K. Gerdes, FtsZ-ZapA-ZapB interactome of Escherichia coli. *J Bacteriol* **194**, 292-302 (2012).
74. D. P. Haeusser, W. Margolin, Splitsville: structural and functional insights into the dynamic bacterial Z ring. *Nat Rev Microbiol* **14**, 305-319 (2016).
75. N. K. Tonthat *et al.*, Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *Embo J* **30**, 154-164 (2011).
76. L. J. Wu, J. Errington, Nucleoid occlusion and bacterial cell division. *Nat Rev Microbiol* **10**, 8-12 (2011).
77. E. Mulder, C. L. Woldringh, Actively replicating nucleoids influence positioning of division sites in Escherichia coli filaments forming cells lacking DNA. *J Bacteriol* **171**, 4303-4314 (1989).
78. P. A. de Boer, R. E. Crossley, L. I. Rothfield, Roles of MinC and MinD in the site-specific septation block mediated by the MinCDE system of Escherichia coli. *J Bacteriol* **174**, 63-70 (1992).
79. K. T. Park, W. Wu, S. Lovell, J. Lutkenhaus, Mechanism of the asymmetric activation of the MinD ATPase by MinE. *Mol Microbiol* **85**, 271-281 (2012).
80. B. D. Corbin, X. C. Yu, W. Margolin, Exploring intracellular space: function of the Min system in round-shaped Escherichia coli. *Embo J* **21**, 1998-2008 (2002).
81. H. Meinhardt, P. A. de Boer, Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. *Proc Natl Acad Sci U S A* **98**, 14202-14207 (2001).
82. E. Bi, J. Lutkenhaus, Interaction between the min locus and ftsZ. *J Bacteriol* **172**, 5610-5616 (1990).
83. E. Bi, J. Lutkenhaus, Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J Bacteriol* **175**, 1118-1125 (1993).
84. O. Huisman, R. D'Ari, S. Gottesman, Cell-division control in Escherichia coli: specific induction of the SOS function SfiA protein is sufficient to block septation. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4490-4494 (1984).
85. K. Dai, J. Lutkenhaus, ftsZ is an essential cell division gene in Escherichia coli. *J Bacteriol* **173**, 3500-3506 (1991).
86. T. Henry, F. Garcia-Del Portillo, J. P. Gorvel, Identification of Salmonella functions critical for bacterial cell division within eukaryotic cells. *Mol Microbiol* **56**, 252-267 (2005).
87. S. J. Arends *et al.*, Discovery and characterization of three new Escherichia coli septal ring proteins that contain a SPOR domain: DamX, DedD, and RlpA. *J Bacteriol* **192**, 242-255 (2010).
88. C. Burke *et al.*, Harnessing Single Cell Sorting To Identify Cell Division Genes and Regulators in Bacteria. *PLoS One* **In press**, (2013).
89. G. J. Patti, O. Yanes, G. Siuzdak, Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-269 (2012).
90. R. M. Vejborg *et al.*, Identification of genes important for growth of asymptomatic bacteriuria Escherichia coli in urine. *Infect Immun* **80**, 3179-3188 (2012).
91. I. Takeda *et al.*, Understanding the human salivary metabolome. *NMR Biomed* **22**, 577-584 (2009).

92. R. Mandal *et al.*, Multi-platform characterization of the human cerebrospinal fluid metabolome: a comprehensive and quantitative update. *Genome Med* **4**, 38 (2012).
93. S. Bouatra *et al.*, The human urine metabolome. *PLoS One* **8**, e73076 (2013).
94. H. Bruckner, S. Haasmann, A. Friedrich, Quantification of D-amino acids in human urine using GC-MS and HPLC. *Amino Acids* **6**, 205-211 (1994).
95. T. A. Stamey, G. Mihara, Observations on the growth of urethral and vaginal bacteria in sterile urine. *J Urol* **124**, 461-463 (1980).
96. M. S. Conover *et al.*, Metabolic Requirements of Escherichia coli in Intracellular Bacterial Communities during Urinary Tract Infection Pathogenesis. *MBio* **7**, e00104-00116 (2016).
97. S. S. Justice, S. R. Lauer, S. J. Hultgren, D. A. Hunstad, Maturation of intracellular Escherichia coli communities requires SurA. *Infection and immunity* **74**, 4793-4800 (2006).
98. R. A. Hull, S. I. Hull, Nutritional requirements for growth of uropathogenic Escherichia coli in human urine. *Infect Immun* **65**, 1960-1961 (1997).
99. J. A. Snyder *et al.*, Transcriptome of uropathogenic Escherichia coli during urinary tract infection. *Infect Immun* **72**, 6373-6381 (2004).
100. J. P. Henderson *et al.*, Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic Escherichia coli. *PLoS Pathog* **5**, e1000305 (2009).
101. A. G. Torres, P. Redford, R. A. Welch, S. M. Payne, TonB-dependent systems of uropathogenic Escherichia coli: aerobactin and heme transport and TonB are required for virulence in the mouse. *Infect Immun* **69**, 6179-6185 (2001).
102. S. L. Chen *et al.*, Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5977-5982 (2006).
103. E. Masse, S. Gottesman, A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A* **99**, 4620-4625 (2002).
104. E. Masse, C. K. Vanderpool, S. Gottesman, Effect of RyhB small RNA on global iron use in Escherichia coli. *J Bacteriol* **187**, 6962-6971 (2005).
105. G. Porcheron *et al.*, The small RNA RyhB contributes to siderophore production and virulence of uropathogenic Escherichia coli. *Infect Immun* **82**, 5056-5068 (2014).
106. M. Miethke, M. A. Marahiel, Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* **71**, 413-451 (2007).
107. D. H. Goetz *et al.*, The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol Cell* **10**, 1033-1043 (2002).
108. M. A. Fischbach *et al.*, The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2. *Proc Natl Acad Sci U S A* **103**, 16502-16507 (2006).
109. A. Garenaux *et al.*, Avian lipocalin expression in chickens following Escherichia coli infection and inhibition of avian pathogenic Escherichia coli growth by Ex-FABP. *Vet Immunol Immunopathol* **152**, 156-167 (2013).

110. M. Hadjifrangiskou *et al.*, A central metabolic circuit controlled by QseC in pathogenic Escherichia coli. *Molecular microbiology* **80**, 1516-1529 (2011).
111. C. L. Shaffer *et al.*, Purine Biosynthesis Metabolically Constrains Intracellular Survival of Uropathogenic Escherichia coli. *Infect Immun* **85**, (2017).
112. S. Crepin *et al.*, Decreased expression of type 1 fimbriae by a pst mutant of uropathogenic Escherichia coli reduces urinary tract infection. *Infect Immun* **80**, 2802-2815 (2012).
113. J. Pak, Y. Pu, Z. T. Zhang, D. L. Hasty, X. R. Wu, Tamm-Horsfall protein binds to type 1 fimbriated Escherichia coli and prevents E. coli from binding to uroplakin Ia and Ib receptors. *J Biol Chem* **276**, 9924-9930 (2001).
114. J. Song *et al.*, A novel TLR4-mediated signaling pathway leading to IL-6 responses in human bladder epithelial cells. *PLoS Pathog* **3**, e60 (2007).
115. J. Song, B. L. Bishop, G. Li, M. J. Duncan, S. N. Abraham, TLR4-initiated and cAMP-mediated abrogation of bacterial invasion of the bladder. *Cell Host Microbe* **1**, 287-298 (2007).
116. T. H. Mogensen, Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **22**, 240-273, Table of Contents (2009).
117. Y. C. Smith, S. B. Rasmussen, K. K. Grande, R. M. Conran, A. D. O'Brien, Hemolysin of uropathogenic Escherichia coli evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice. *Infect Immun* **76**, 2978-2990 (2008).
118. R. Stenutz, A. Weintraub, G. Widmalm, The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiol Rev* **30**, 382-403 (2006).
119. C. Whitfield, I. S. Roberts, Structure, assembly and regulation of expression of capsules in Escherichia coli. *Mol Microbiol* **31**, 1307-1319 (1999).
120. T. Katayama, M. Takata, K. Sekimizu, CedA is a novel Escherichia coli protein that activates the cell division inhibited by chromosomal DNA over-replication. *Molecular microbiology* **26**, 687-697 (1997).
121. R. Byrne-Nash *et al.*, Probing the Mechanism of LAL-32, a Gold Nanoparticle-Based Antibiotic Discovered through Small Molecule Variable Ligand Display. *Bioconjug Chem* **28**, 1807-1810 (2017).
122. D. M. Guyer, I. R. Henderson, J. P. Nataro, H. L. Mobley, Identification of sat, an autotransporter toxin produced by uropathogenic Escherichia coli. *Mol Microbiol* **38**, 53-66 (2000).
123. G. C. Langridge *et al.*, Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome research* **19**, 2308-2316 (2009).
124. M. D. Phan *et al.*, The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. *PLoS Genet* **9**, e1003834 (2013).
125. T. Baba *et al.*, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
126. K. A. Datsenko, B. L. Wanner, 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**, 6640-6645 (2000).
127. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

128. L. Barquist *et al.*, The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109-1111 (2016).
129. F. Ramirez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
130. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
131. K. Rutherford *et al.*, Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).
132. P. Van De Putte, D. Van, A. Roersch, The Selection of Mutants of Escherichia Coli with Impaired Cell Division at Elevated Temperature. *Mutat Res* **106**, 121-128 (1964).
133. R. L. Lock, E. J. Harry, Cell-division inhibitors: new insights for future antibiotics. *Nature reviews. Drug discovery* **7**, 324-338 (2008).
134. S. Ma, S. Ma, The development of FtsZ inhibitors as potential antibacterial agents. *ChemMedChem* **7**, 1161-1172 (2012).
135. J. Mannik, M. W. Bailey, Spatial coordination between chromosomes and cell division proteins in Escherichia coli. *Front Microbiol* **6**, 306 (2015).
136. H. Cho, H. R. McManus, S. L. Dove, T. G. Bernhardt, Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. *Proc Natl Acad Sci U S A* **108**, 3773-3778 (2011).
137. S. Pichoff, J. Lutkenhaus, Escherichia coli division inhibitor MinCD blocks septation by preventing Z-ring formation. *J Bacteriol* **183**, 6630-6635 (2001).
138. M. W. Bailey, P. Bisicchia, B. T. Warren, D. J. Sherratt, J. Mannik, Evidence for divisome localization mechanisms independent of the Min system and SlmA in Escherichia coli. *PLoS Genet* **10**, e1004504 (2014).
139. O. Espeli *et al.*, A MatP-divisome interaction coordinates chromosome segregation with cell division in E. coli. *Embo J* **31**, 3198-3211 (2012).
140. J. Lutkenhaus, S. Du, E. coli Cell Cycle Machinery. *Subcell Biochem* **84**, 27-65 (2017).
141. M. Radman, SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci* **5A**, 355-367 (1975).
142. D. J. Horvath, Jr. *et al.*, Morphological plasticity promotes resistance to phagocyte killing of uropathogenic Escherichia coli. *Microbes and infection / Institut Pasteur* **13**, 426-437 (2011).
143. M. A. Gerding *et al.*, Self-enhanced accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR domain (DamX, DedD, and RlpA) in Escherichia coli cell constriction. *J Bacteriol* **191**, 7383-7401 (2009).
144. A. Lyngstadaas, A. Lobner-Olesen, E. Boye, Characterization of three genes in the dam-containing operon of Escherichia coli. *Molecular & general genetics : MGG* **247**, 546-554 (1995).
145. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
146. G. Ebersbach, E. Galli, J. Moller-Jensen, J. Lowe, K. Gerdes, Novel coiled-coil cell division factor ZapB stimulates Z ring assembly and cell division. *Molecular microbiology* **68**, 720-735 (2008).

147. Z. Hu, A. Mukherjee, S. Pichoff, J. Lutkenhaus, The MinC component of the division site selection system in Escherichia coli interacts with FtsZ to prevent polymerization. *Proc Natl Acad Sci U S A* **96**, 14819-14824 (1999).
148. N. S. Hill, P. J. Buske, Y. Shi, P. A. Levin, A moonlighting enzyme links Escherichia coli cell size with central metabolism. *PLoS Genet* **9**, e1003663 (2013).
149. A. K. Gohler *et al.*, More than just a metabolic regulator--elucidation and validation of new targets of PdhR in Escherichia coli. *BMC Syst Biol* **5**, 197 (2011).
150. L. G. Monahan, I. V. Hajduk, S. P. Blaber, I. G. Charles, E. J. Harry, Coordinating bacterial cell division with nutrient availability: a role for glycolysis. *MBio* **5**, e00935-00914 (2014).
151. T. Homeier, T. Semmler, L. H. Wieler, C. Ewers, The GimA locus of extraintestinal pathogenic E. coli: does reductive evolution correlate with habitat and pathotype? *PLoS One* **5**, e10877 (2010).
152. D. P. Haeusser *et al.*, The Kil peptide of bacteriophage lambda blocks Escherichia coli cytokinesis via ZipA-dependent inhibition of FtsZ assembly. *PLoS Genet* **10**, e1004217 (2014).
153. D. Balasubramanian, P. T. Ragunathan, J. Fei, C. K. Vanderpool, A Prophage-Encoded Small RNA Controls Metabolism and Cell Division in Escherichia coli. *mSystems* **1**, (2016).
154. O. N. Murashko, S. Lin-Chao, Escherichia coli responds to environmental changes using enolasic degradosomes and stabilized DicF sRNA to alter cellular morphology. *Proc Natl Acad Sci U S A* **114**, E8025-E8034 (2017).
155. R. D'Ari, O. Huisman, Novel mechanism of cell division inhibition associated with the SOS response in Escherichia coli. *J Bacteriol* **156**, 243-250 (1983).
156. D. M. Heller, M. Tavag, A. Hochschild, CbtA toxin of Escherichia coli inhibits cell division and cell elongation via direct and independent interactions with FtsZ and MreB. *PLoS Genet* **13**, e1007007 (2017).
157. K. Pedersen, K. Gerdes, Multiple hok genes on the chromosome of Escherichia coli. *Mol Microbiol* **32**, 1090-1102 (1999).
158. S. Barshishat *et al.*, OxyS small RNA induces cell cycle arrest to allow DNA damage repair. *Embo J* **37**, 413-426 (2018).
159. X. Wang *et al.*, Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**, 147 (2010).
160. H. Gueune, M. J. Durand, G. Thouand, M. S. DuBow, The ygaVP genes of Escherichia coli form a tributyltin-inducible operon. *Appl Environ Microbiol* **74**, 1954-1958 (2008).
161. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* **177**, 4121-4130 (1995).
162. I. F. Lau *et al.*, Spatial and temporal organization of replicating Escherichia coli chromosomes. *Mol Microbiol* **49**, 731-743 (2003).
163. E. Afgan *et al.*, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3-W10 (2016).
164. A. M. Feist *et al.*, A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**, 121 (2007).

165. S. Y. Gerdes *et al.*, Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *Journal of bacteriology* **185**, 5673-5684 (2003).
166. A. R. Joyce *et al.*, Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **188**, 8259-8271 (2006).
167. T. van Opijnen, A. Camilli, Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **11**, 435-442 (2013).
168. A. J. McCarthy, R. A. Stabler, P. W. Taylor, Genome-Wide Identification by Transposon Insertion Sequencing of Escherichia coli K1 Genes Essential for In Vitro Growth, Gastrointestinal Colonizing Capacity, and Survival in Serum. *J Bacteriol* **200**, (2018).
169. E. C. A. Goodall *et al.*, The Essential Genome of Escherichia coli K-12. *MBio* **9**, (2018).
170. M. Dembek *et al.*, High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. *MBio* **6**, e02383 (2015).
171. K. G. K. Goh *et al.*, Genome-Wide Discovery of Genes Required for Capsule Production by Uropathogenic Escherichia coli. *MBio* **8**, (2017).
172. J. Baranyi, T. A. Roberts, A dynamic approach to predicting bacterial growth in food. *Int J Food Microbiol* **23**, 277-294 (1994).
173. T. van Opijnen, K. L. Bodi, A. Camilli, Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**, 767-772 (2009).
174. J. D. Gawronski, S. M. Wong, G. Giannoukos, D. V. Ward, B. J. Akerley, Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci U S A* **106**, 16422-16427 (2009).
175. I. F. Bronner *et al.*, Quantitative insertion-site sequencing (QIseq) for high throughput phenotyping of transposon mutants. *Genome Res* **26**, 980-989 (2016).
176. J. A. Yethon, E. Vinogradov, M. B. Perry, C. Whitfield, Mutation of the lipopolysaccharide core glycosyltransferase encoded by waaG destabilizes the outer membrane of Escherichia coli by interfering with core phosphorylation. *J Bacteriol* **182**, 5620-5623 (2000).
177. J. D. Klena, R. S. Ashford, 2nd, C. A. Schnaitman, Role of Escherichia coli K-12 rfa genes and the rfp gene of Shigella dysenteriae 1 in generation of lipopolysaccharide core heterogeneity and attachment of O antigen. *J Bacteriol* **174**, 7297-7307 (1992).
178. C. G. Korea, R. Badouraly, M. C. Prevost, J. M. Ghigo, C. Beloin, Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol* **12**, 1957-1977 (2010).
179. F. Larsonneur *et al.*, Functional analysis of Escherichia coli Yad fimbriae reveals their potential role in environmental persistence. *Environ Microbiol* **18**, 5228-5248 (2016).
180. K. Otto, J. Norbeck, T. Larsson, K. A. Karlsson, M. Hermansson, Adhesion of type 1-fimbriated Escherichia coli to abiotic surfaces leads to altered composition of outer membrane proteins. *J Bacteriol* **183**, 2445-2453 (2001).

181. M. J. Weickert, S. Adhya, A family of bacterial regulators homologous to Gal and Lac repressors. *J Biol Chem* **267**, 15869-15874 (1992).
182. P. Gerlach, P. Valentin-Hansen, E. Bremer, Transcriptional regulation of the cytR repressor gene of Escherichia coli: autoregulation and positive control by the cAMP/CAP complex. *Mol Microbiol* **4**, 479-488 (1990).
183. G. G. Anderson, C. C. Goller, S. Justice, S. J. Hultgren, P. C. Seed, Polysaccharide capsule and sialic acid-mediated regulation promote biofilm-like intracellular bacterial communities during cystitis. *Infect Immun* **78**, 963-975 (2010).
184. W. J. Dower, J. F. Miller, C. W. Ragsdale, High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-6145 (1988).
185. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
186. M. Hadjifrangiskou *et al.*, Transposon mutagenesis identifies uropathogenic Escherichia coli biofilm factors. *Journal of bacteriology* **194**, 6195-6205 (2012).
187. C. C. Goller, P. C. Seed, Revisiting the Escherichia coli polysaccharide capsule as a virulence factor during urinary tract infection: contribution to intracellular biofilm development. *Virulence* **1**, 333-337 (2010).
188. T. F. Nicholson, K. M. Watts, D. A. Hunstad, OmpA of uropathogenic Escherichia coli promotes postinvasion pathogenesis of cystitis. *Infection and immunity* **77**, 5245-5251 (2009).
189. G. Nagy *et al.*, Loss of regulatory protein RfaH attenuates virulence of uropathogenic Escherichia coli. *Infect Immun* **70**, 4406-4413 (2002).
190. T. Porter, D. Frederick, E. Johnson, P. G. Jones, A requirement for cell elongation protein RodZ and cell division proteins FtsN and DedD to maintain the small rod morphology of Escherichia coli at growth temperatures near 8 degrees C. *J Gen Appl Microbiol* **62**, 189-198 (2016).
191. K. L. Schmidt *et al.*, A predicted ABC transporter, FtsEX, is needed for cell division in Escherichia coli. *J Bacteriol* **186**, 785-793 (2004).
192. C. N. Marbaniang, J. Gowrishankar, Role of ArgP (IciA) in lysine-mediated repression in Escherichia coli. *J Bacteriol* **193**, 5985-5996 (2011).
193. H. K. Kim, R. M. Harshey, A Diguanylate Cyclase Acts as a Cell Division Inhibitor in a Two-Step Response to Reductive and Envelope Stresses. *MBio* **7**, (2016).
194. M. A. Teleha, A. C. Miller, R. A. Larsen, Overexpression of the Escherichia coli TolQ protein leads to a null-FtsN-like division phenotype. *Microbiologyopen* **2**, 618-632 (2013).
195. M. A. Jorgenson, K. D. Young, YtfB, an OapA domain-containing protein, is a new cell division protein in Escherichia coli. *J Bacteriol*, (2018).
196. B. Withman, T. S. Gunasekera, P. Beesetty, R. Agans, O. Paliy, Transcriptional responses of uropathogenic Escherichia coli to increased environmental osmolality caused by salt or urea. *Infect Immun* **81**, 80-89 (2013).
197. F. Carballes, C. Bertrand, J. P. Bouche, K. Cam, Regulation of Escherichia coli cell division genes ftsA and ftsZ by the two-component system rcsC-rcsB. *Mol Microbiol* **34**, 442-450 (1999).

198. M. L. Bernardini, M. G. Sanna, A. Fontaine, P. J. Sansonetti, OmpC is involved in invasion of epithelial cells by Shigella flexneri. *Infect Immun* **61**, 3625-3635 (1993).
199. A. M. Sperber, J. K. Herman, Metabolism Shapes the Cell. *J Bacteriol* **199**, (2017).
200. M. Faubladier, J. P. Bouche, Division inhibition gene dicF of Escherichia coli reveals a widespread group of prophage sequences in bacterial genomes. *J Bacteriol* **176**, 1150-1156 (1994).
201. B. T. Waggoner, K. Sultana, N. Symonds, M. A. Karlok, M. L. Pato, Identification of the bacteriophage Mu kil gene. *Virology* **173**, 378-389 (1989).

[blank page]