ucc

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Using high-density mutagenesis to identify the genetic requirements for the growth of *Escherichia coli*

Finbarr James Buttimer, B.Sc

A thesis submitted for the degree of Doctor of Philosophy

National University of Ireland, Cork,
School of Microbiology,
March 2019

Head of Department: Prof. Gerald Fitzgerald
Supervisor: Dr. David Clarke

Table of Contents

Table of Figures

List of Tables

Declaration

I declare that the research presented in this thesis is my own work and that it has not been submitted for any other degree, either at University College Cork, or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concering plagiarism. Wherever contributions of others are involved, every effort has been made to indicate this clearly, by reference to the literature and by acknowledgement of collaborative research. This work was completed under the guidance of Dr. David Clarke at the School of Microbiology, University College Cork, Ireland.

Finbarr James Buttimer

July 2019

Acknowledgements

Any body of work requires the help and support of many different people and my case has been no exception. It is through the amazing support of my family, friends, and colleagues that I have been able to produce this work, and to have all these people in my life I am incredibly grateful.

First and foremost, I wish to thank my supervisor, Dr. David Clarke, for all his hard work, his constant support and guidance, and dedication to me and this project for the past four and a half years. Dave is an example of the dedicated and understanding person that a supervisor should be, and I am so grateful to have had him as a mentor.

I wish to thank, from the bottom of my heart, my parents, Noreen and Neil, and my brother and sister, Patrick and Catherine (and of course, the dog, Bran). It is difficult to describe just how much my parents and my family mean to me. They have always been there for me, always offering a listening ear, always supporting me through my ups and downs, or doing the little things like making a call or being around at home to have a chat. They were there for me throughout the hardest parts of my PhD and helped me stick with it, so if it weren't for them I wouldn't have written this thesis. Míle míle buíochas daoibh.

I would also like to dedicate this thesis to my late grandmother, Sheila O' Mahony (neé Curtin), who passed away before she could see it finished.

My PhD was funded by the Irish Research Council under a Government of Ireland Postgraduate Scholarship and co-funded by APC Microbiome Ireland. I would like to thank them for providing the financial support and facilities necessary to undertake this project. My thanks also everyone who has contributed to the work in this thesis, including Emma Smith (formerly of UCC), Francesca Short, Amy Cain, Christine Boinnett and Trevor Lawley of the Wellcome Trust Sanger Institute, and Fiona Crispie and Laura Finnegan of Teagasc Moorepark. With that I would also like to thank the technicians and staff of the School of Microbiology for all their help and guidance.

Thanks also to everyone, past and present, in office 435 for all the good food, good drink, hilarious chats, and all around brilliant craic. Also, my deepest thanks to everyone in lab 405, Alli, Marc, Ian, and Dana. The past four and a half years working with you have been an absolute blast.

Finally, I wish to thank my partner, my love, Sloane. Without Sloane there is no way I would have finished this PhD. Through thick and thin Sloane has been by my side, taking care of me, supporting and encouraging me, no matter what came our way. She has made me a better man. Je t'aime, ma chérie :)

I wrote the majority of this thesis by myself in a house in Kinsale, but my one companion was my little black and white cat, Willow. In honour of

Willow, here is an extract of a poem written by an Irish monk in the 9th century about writing with his cat:

Messe ocus Pangur bán,
cechtar nathar fria shaindán:
bíth a menmasam fri seilgg,
mu menma céin im shaincheirdd.

Caraimse fos, ferr cach clú
oc mu lebrán, léir ingnu;
ní foirmtech frimm Pangur bán:
caraid cesin a maccdán.

Ó ru biam, scél cen scís,
innar tegdais, ar n-óendís,
táithiunn, díchríchide clius,
ní fris tarddam ar n-áthius.

("Pangur Bán and I at work, adepts, equals, cat and clerk: His whole instinct is to hunt, Mine to free the meaning pent. More than loud acclaim, I love books, silence, thought, my alcove. Happy for me, Pangur Bán child-plays round some mouse's den. Truth to tell, just being here, housed alone, housed together, adds up to its own reward: concentration, stealthy art". Translation by Séamus Heaney)

List of Abbreviations

AcCoA	Acetyl coenzyme A	**IM**	Inner membrane
ACP	Acyl-carrier protein	**INSeq**	Insertion sequencing
AIEC	Adherent-invasive *E. coli*	**KDO**	2-keto-3-deoxy-octulosonate
APEC	Avian pathogenic *E. coli*	**LB**	Lysogeny broth
AR	Acid-resistance	**LEE**	Locus of enterocyte effacement
CD	Crohn's disease	**logFC**	Log2 fold-change
CDS	Protein-encoding sequence	**LPS**	Lipopolysaccharide
CI	Competitive index	**MLEE**	Multilocus enzyme electrophoresis
CoA	Coenzyme A	**NRA**	Nitrate reductase A
COG	Clusters of orthologous groups	**NRZ**	Nitrate reductase Z
DAEC	Diffusely-adherent *E. coli*	**OD**	Optical density
DSB	Disulfide bond	**OM**	Outer membrane
EAEC	Enteroaggregative *E. coli*	**OMP**	Outer membrane protein
ECA	Enterobacterial common antigen	**PEC**	Profiling of the *E. coli* chromosome
EcN	*E. coli* Nissle 1917	**PEP**	Phosphoenol pyruvate
ECOR	*E. coli* reference	**PFL**	Pyruvate-formate lyase
EHEC	Enterohaemorrhagic *E. coli*	**PG**	Peptidoglycan
EIEC	Enteroinvasive *E. coli*	**PK**	Pyruvate kinase
ETEC	Enterotoxigenic *E. coli*	**PTS**	Phosphotransferase system
ExPEC	Extraintestinal pathogenic *E. coli*	**RNS**	Reactive nitrogen species
FNR	Fumarate-nitrate reductase	**ROS**	Reactive oxygen species

	regulator
GI	Gastrointestinal
GO	Gene ontology
GSH	Glutathione
HITS	High-throughput insertion tracking by deep sequencing
HPLC	High-performace liquid chromatography
IBD	Inflammatory bowel disease
IFR	Insertion-free region
IGR	Intergenic region
IHF	Integration host factor
SPATE	Serine protease autotransporters of Enterobacteriaceae
STEC	Shiga toxin-producing *E. coli*
TLR	Toll-like receptor
Tn-seq	Transposon sequencing
TraDIS	Transposon-directed insertion site sequencing
UPEC	Uropathogenic *E. coli*
UTI	Urinary tract infection

Abstract

Escherichia coli is highly adapted to life within the mammalian gastrointestinal (GI) tract, capable of adapting to multiple environments en route to colonising the intestine. Moreover, the means by which the species copes with changes in the microenvironments of the GI tract strongly influences the nature of *E. coli*'s relationship with the host i.e. whether it exists as a commensal or pathogen. However, the response of *E. coli* to many of these conditions is complex, often employing a whole-cell response. This necessitates the use of high-throughput approaches in order to fully understand factors the bacterium requires for growth under these conditions. This work outlines the use of transposon-directed insertion site sequencing (TraDIS) to describe genetic requirements for fitness of *E. coli* K-12 MG1655 during growth in the presence of bile and under anaerobic conditions in the presence of nitrate, representative of bile exposure in the duodenum and conditions of inflammation in the intestine. TraDIS reveals, in detail, genetic requirements for growth under these conditions, revealing new roles for many genes with no prior association with growth under these conditions. This work will, therefore, contribute to future studies of *E. coli* colonisation of the GI tract by identifying candidate genes required for fitness under these growth conditions.

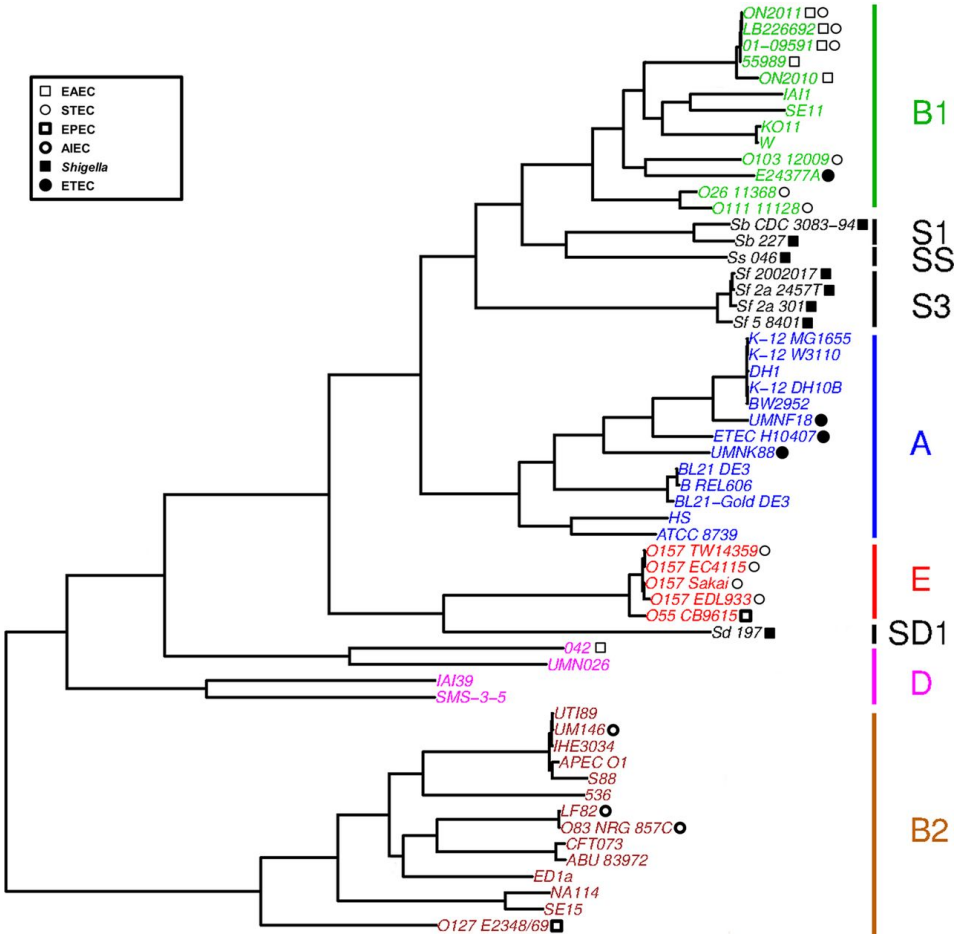
Chapter 1 Introduction

1.1. Escherichia coli: role in gut health and disease

1.1.1. Biology and population structure

Escherichia coli is a Gram-negative, facultatively anaerobic, non-spore forming bacterium of the family Enterobacteriaceae (Scheutz & Strockbine, 2015). *E. coli* is the model organism of molecular biology and one of the best characterised organisms. Its primary habitat is the gastrointestinal (GI) tract of warm-blooded mammals and reptiles, with a secondary habitat in water, soil, and sediment (Berg, 1996; Gordon, 2004; Savageau, 1983). *E. coli* typically colonises the colonic mucosa where it exists as one of the most abundant facultative anaerobes in the human gut microbiota (Berg, 1996).

E. coli is a highly diverse species. Multiple methods exist for the classification of the population structure of *E. coli*, ranging from pathogenic phenotype (Kaper *et al.*, 2004), serotyping lipopolysaccharide (O), flagellar (H), and capsule (K) antigens (Evans Doyle J. & Evans, 1983), to whole genome sequencing (Sims & Kim, 2011). However, the population structure described by Ochman and Selander using the *E. coli* reference (ECOR) collection and multilocus enzyme electrophoresis (MLEE) forms the basis for current studies on the population structure of *E. coli*, with several refinements offered in the interim (Leimbach *et al.*, 2013; Ochman & Selander, 1984; Sims & Kim, 2011). Currently, *E. coli* is divided into five major phylogroups, A, B1, B2, D, and E (Figure 1.1).



Matthew A. Croxen et al. Clin. Microbiol. Rev. 2013; doi: 10.1128/CMR.00022-13

Figure 1.1. Phylogenetic tree of intestinal pathogenic *E. coli*. *E. coli* strains can be grouped into 5 main phylogenetic groups: A (blue), B1 (green), B2 (brown), D (pink), and E (red). Shigella/EIEC also form additional phylogroups (black). Pathotypes do not always group together in the same phylogroup. The hybrid EAEC and STEC strains are denoted with both an open square and open circle. Unmarked strains are either commensal, extraintestinal pathogenic *E. coli* (ExPEC), or avian-pathogenic *E. coli* (APEC). ETEC strains are isolated from both humans and animals, while DAEC is not represented in the phylogenetic tree. Adapted from Croxen *et al*., 2013.

Groups A and B1 are the youngest lineages and consist primarily of non-pathogenic commensal strains, however B1 contains more of a mix of pathogens and commensals, including non-O157:H7 enterohaemorrhagic *E. coli* (EHEC). Phylogroup B2 is the most diverse lineage, comprised of many extraintestinal pathogenic *E. coli* (ExPEC) and adherent-invasive *E. coli* (AIEC) strains. Phylogroup D is not monophyletic and splits into subgroups D1 and D2 (Figure 1.1). Group D1 is composed of uropathogenic *E. coli* (UPEC) and enteroaggregative *E. coli* (EAEC) isolates and clusters towards groups A, B1, and E, whereas group D2, consisting of ExPEC and environmental strains, clusters more towards group B2. Finally, group E contains O157:H7 EHEC and enteropathogenic *E. coli* (EPEC) strains. Group E is close to the '*E. coli* pathotype' *Shigella* outgroup, retained as a genus for historical reasons (Pettengill *et al.*, 2016). Genotypic variation in *E. coli* is not due to a clonal population structure, but rather the genome of *E. coli* is highly dynamic, undergoing extensive horizontal gene transfer and recombination between strains. Indeed, two strains can differ by as much as a megabase in genome size (Ochman & Jones, 2000). Nevertheless, each strain contains a shared 'core' genome including essential housekeeping genes and other indispensible functions, supplemented with an accessory genome, required for expressing various phenotypes and environmental adaptation (Rasko *et al.*, 2008). The size of the core genome has been estimated in a number of different studies, ranging from approximately 1,400 genes to 2,800 genes (Fukiya *et al.*, 2004; Kaas *et al.*, 2012;

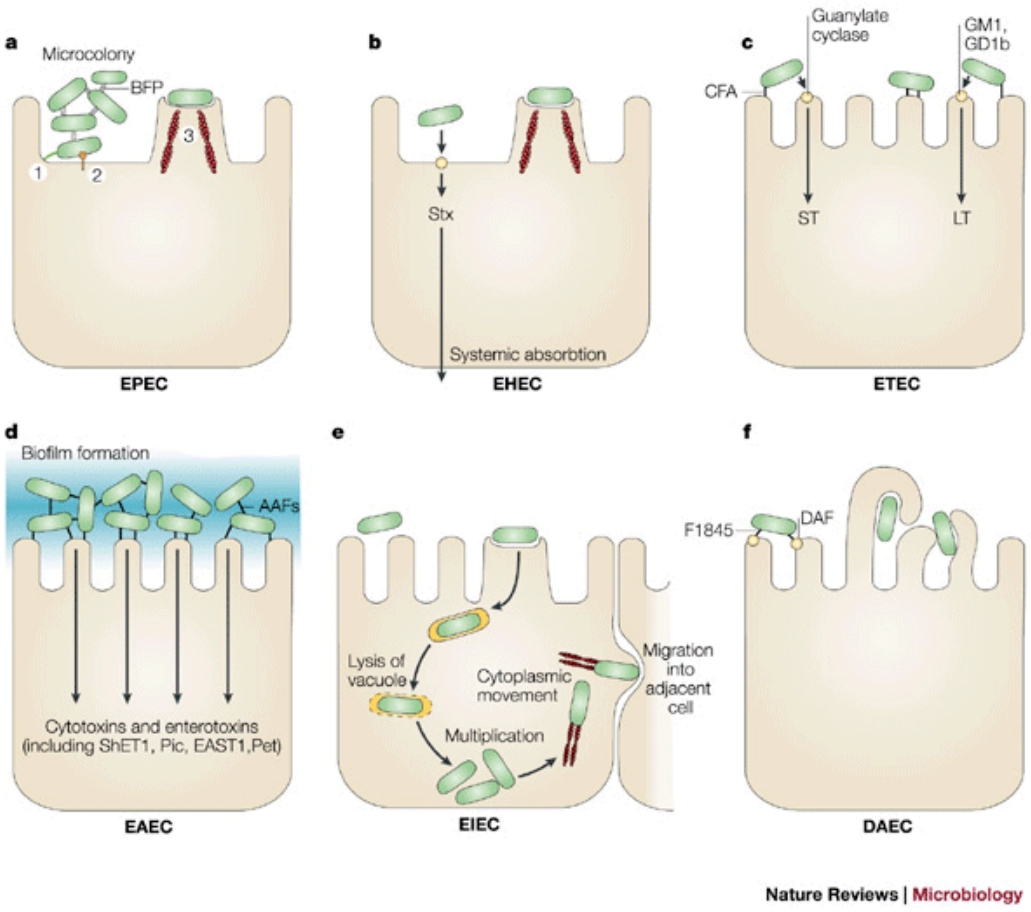
Lukjancenko *et al.*, 2010; Vieira *et al.*, 2011). Therefore, the core genome has been subdivided into 'soft-core' and strict core', i.e. genes present in at least 95% of all strains and 100% of all strains, respectively (Kaas *et al.*, 2012). A study of all publicly available *E. coli* strains estimated the soft-core genome to consist of 3,051 genes, and the strict-core genome at 1,702 genes (Kaas *et al.*, 2012). However, all core and accessory genes, known as the pangenome, are thought to consist of between 18,000 to 43,000 genes, a number thought to increase with the increasing availability of fully sequenced genomes. (Leimbach *et al.*, 2013; Rasko *et al.*, 2008; Snipen *et al.*, 2009). This extensive pangenome reflects the high diversity of the *E. coli* species.

E. coli is a microbe that has a complex relationship with its host. Over 90% of the human population carry strains of *E. coli*, in greater quantities than domestic or wild animals, with an intimate association between the types of strains hosted and environment, geography, or lifestyle (Gordon, 2004; Tenaillon *et al.*, 2010). The majority of strains exist as commensals; however, depending on host and microbe genetics and gut environmental conditions, many *E. coli* can also act as pathogens. In this way, *E. coli* represents a model organism for studying the paradigm of the commensal-to-pathogen switch (Leimbach *et al.*, 2013).

1.1.2. Pathogenic and commensal E. coli

E. coli are globally significant pathogens; diarrhoeal diseases alone caused by pathogenic *E. coli* led to approximately 420 million cases of

diarrhoea and over 250,000 deaths in 2010, a significant proportion of which occurred in children under 5 years old (Pires *et al.*, 2015). However, *E. coli* are also a common cause of, or contributor to, sepsis, meningitis, urinary tract infections (UTIs), intra-abdominal infections, and nosocomial infections (Peleg & Hooper, 2010; Tenaillon *et al.*, 2010). Pathogenic *E. coli* are typically divided into pathotypes, i.e. based on their mode of colonisation and infection (Figure 1.2). The majority of *E. coli* exist as commensals within the human gut, and indeed play multiple roles as members of the normal GI tract microbiota (Tenaillon *et al.*, 2010). In the GI tract, commensal *E. coli* reside in the caecum and colon, colonising the mucus layer that covers the GI epithelium, and the bacteria are shed into the lumen and faeces with the degradation of mucus (Poulsen *et al.*, 1994). Commensal *E. coli* colonise the neonatal colon soon after birth, helping to deoxygenate the colon, which is thought to facilitate the establishment of an anaerobic niche for beneficial obligate anaerobes such as *Bifidobacterium, Bacteroides,* and *Clostridium* (Bettelheim & Lennox-King, 1976; Guaraldi & Salvatori, 2012).



g

Figure 1.2. *E. coli* pathotypes and modes of interaction with intestinal enterocytes. **a.** Enteropathogenic *E. coli* (EPEC) adhere to small bowel enterocytes, but destroy the normal microvillar architecture, inducing the characteristic attaching and effacing (A/E) lesion. Cytoskeletal rearrangements are accompanied by an inflammatory response and diarrhoea. 1. Initial adhesion, 2. Protein translocation by type III secretion, 3. Pedestal formation. **b.** Enterohaemhorrhagic *E. coli* (EHEC) also induce the A/E lesion, but in the colon. The distinguishing feature of EHEC is the elaboration of Shiga toxin (Stx), the systemic absorption of which leads to potentially life-threatening complications. **c.** Similarly, Enterotoxigenic *E. coli* (ETEC) adhere to small bowel enterocytes and induce watery diarrhoea by the secretion of heat-labile (LT) and/or heat-

stable (ST) enterotoxins. **d.** Enteroaggregative *E. coli* (EAEC) adheres to small and large bowel epithelia in a thick biofilm and elaborates secretory enterotoxins and cytotoxins. **e.** Enteroinvasive *E. coli* (EIEC) invades the colonic epithelial cell, lyses the phagosome and moves through the cell by nucleating actin microfilaments. The bacteria might move laterally through the epithelium by direct cell-to-cell spread or might exit and re-enter the baso-lateral plasma membrane. **f.** Diffusely-adherent *E. coli* (DAEC) elicits a characteristic signal transduction effect in small bowel enterocytes that manifests as the growth of long finger-like cellular projections, which wrap around the bacteria. AAF, aggregative adherence fimbriae; BFP, bundle-forming pilus; CFA, colonization factor antigen; DAF, decay-accelerating factor; EAST1, enteroaggregative E. coli ST1; LT, heat-labile enterotoxin; ShET1, Shigella enterotoxin 1; ST, heat-stable enterotoxin. Figure and text adapted from Kaper *et al*., 2004.

Commensals can also provide a barrier to invasion by pathogenic strains, in a phenomenon known as colonisation resistance (Apperloo-Renkema *et al.*, 1990). However, *E. coli* strains that are normally commensal or beneficial can, under certain circumstances, be detrimental to the host. For example, commensal *E. coli* are well documented reservoirs of antibiotic resistance determinants, which are thought to be shared between commensals and pathogens, or between humans and animals (Bailey *et al.*, 2010; Karami *et al.*, 2007; Skurnik *et al.*, 2006; Stecher *et al.*, 2012), although this has recently been shown not to occur between humans and livestock animals, necessitating further investigation (Ludden *et al.*, 2019). Furthermore, commensals can change their behavior in response to changing environmental conditions, eliciting more pathogenic phenotypes. This is particularly noteworthy in the case of UPECs, where strains typically exist as commensals within the gut but cause disease when they invade the urinary tract (Sabaté *et al.*, 2006). Also of note is the common observation that humans with inflammatory bowel disease (IBD), particularly Crohn's disease (CD), contain greater numbers of *Enterobacteriaceae* in their gut microbiota, particularly *E. coli* (Lopez-Siles *et al.*, 2014; Martinez-Medina *et al.*, 2006; Willing *et al.*, 2009). One line of research suggests that this may be due to *E. coli*'s ability to use anaerobic terminal electron acceptors, such as nitrate, produced by the host during inflammation for anaerobic respiration (Winter *et al.*, 2013).

The boundary between commensal and pathogenic *E. coli* is blurred, and often as a result of the differing responses of strains to

changes in microecology or environmental challenge (Leimbach *et al.*, 2013). For example, virulence factors such as P fimbriae of UPEC strains exacerbate the course of UTI in allowing *E. coli* to colonise the kidneys, but these fimbriae also aid in colonisation of the gut (Le Gall *et al.*, 2007; Wold *et al.*, 1992). Another example occurs during gut inflammation. The ability to anaerobically respire nitrate allows *E. coli* and other Enterobacteriaceae to utilise the metabolic end-products of fermentation by Bacteroidia and Clostridia (e.g. formate and hydrogen) as electron donors, allowing *E. coli* to establish itself within its primary niche (Conway & Cohen, 2015; Faber & Bäumler, 2014; Jones *et al.*, 2011). However, excessive nitrate production as a result of chronic gut inflammation favours *E. coli*'s ability to respire using nitrate as a terminal electron acceptor, allowing it to outgrow other members of the microbiota and potentially exacerbating the symptoms of gut inflammatory diseases (Faber & Bäumler, 2014). Furthermore, certain pathotypes of *E. coli*, including adherent-invasive *E. coli* (AIEC), DAEC, and ExPEC, are thought to influence the symptoms of IBD in humans with genetic or immune defects (Mirsepasi-Lauridsen *et al.*, 2019). These examples have led to the designation of certain commensal strains of *E. coli* as 'pathobionts', or resident microbes with pathogenic potential (Mirsepasi-Lauridsen *et al.*, 2019). This highlights the importance of developing a greater understanding of the effects of environmental change on the behaviour of commensals to enable an accurate prediction of the commensal-to-pathogen switch. To do this, a detailed understanding of

the environmental conditions *E. coli* faces during colonisation of the GI tract is required.

1.2. Colonisation of the gastrointestinal tract by E. coli

All strains of *E. coli*, pathogen or commensal, must first colonise the gut, defined as achieving and maintaining a stable population without reintroduction (Meador *et al.*, 2014). Successful colonisation requires mechanisms to survive environmental stresses, but also mechanisms to compete against other microbes (typically other *E. coli* strains), in order to occupy a distinct niche. *E. coli* harbours a wide array of mechanisms that allow it to address these different environmental challenges i.e. stresses that it faces during colonisation of the human gut (Figure 1.3).

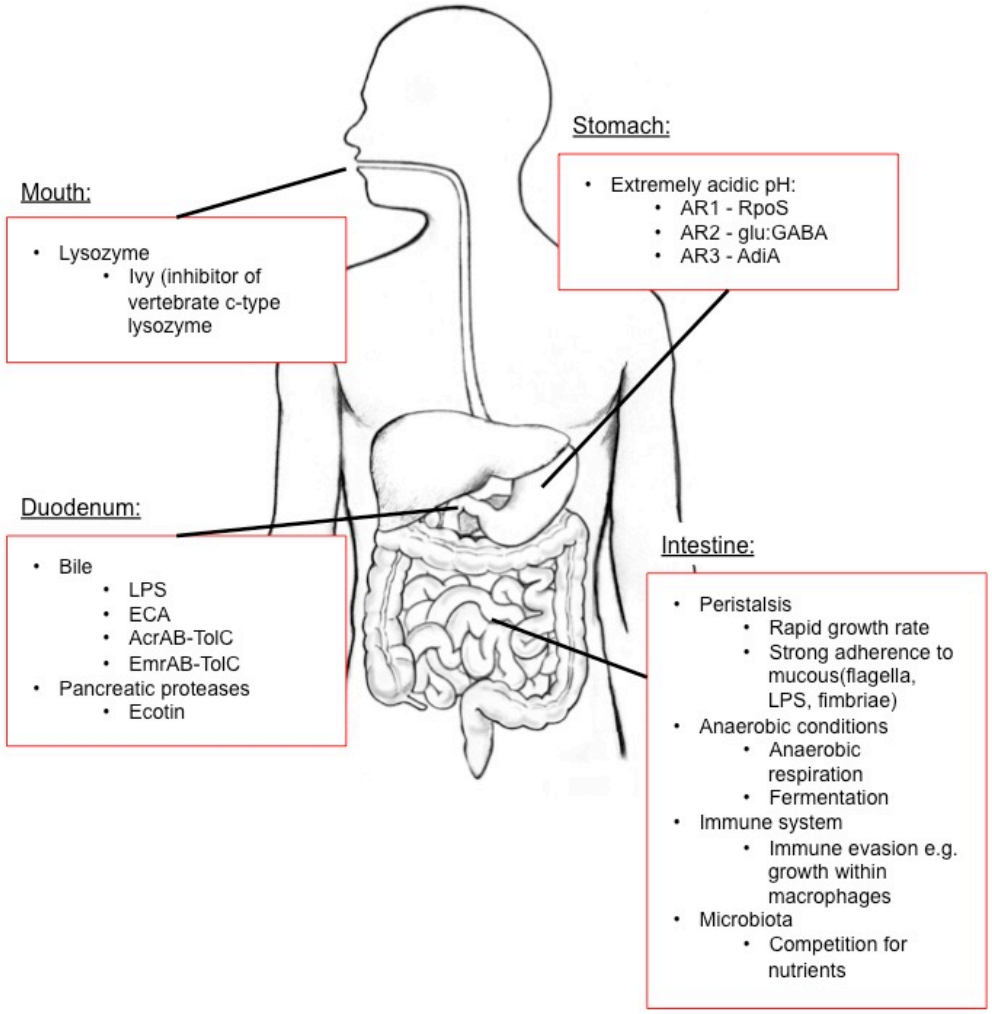


Figure 1.3. Stresses faced by *E. coli* in different regions of the GI tract. Stresses highlighted with bullet points, with adaptation mechanisms described in indented bullet points. Image adapted from NIDDK image library (https://catalog.niddk.nih.gov/Catalog/ImageLibrary/searchresults.cfm?keyword=93&type=keyword accessed 26 Nov 2018).

1.2.1. Survival within the mouth, oesophagus, and gastric juice

E. coli is not thought to be a prominent member of the healthy oral microbiome (Aas *et al.*, 2005; Dewhirst *et al.*, 2010). In fact, the presence of *E. coli* in the mouth has been used as an indicator of fecal contamination (Baydaş *et al.*, 2007; Oliveira *et al.*, 2012). Nevertheless, *E. coli* does need to passage through the mouth in order to colonise the gut, and as such contains mechanisms to survive the potential stresses in this environment. The principal innate anti-microbial mechanism contained within saliva is lysozyme, an enzyme that breaks the β-(1,4)-glycosidic bond between the alternating N-acetylmuramic acid and N-acetylglucosamine residues in bacterial peptidoglycan, resulting in cell wall damage and lysis. *E. coli* harbours mechanisms that contribute to lysozyme resistance, including the innate protection provided by the outer membrane, but also specific mechanisms such as the production of Ivy, an inhibitor of vertebrate C-type lysozyme (Deckers *et al.*, 2008; Monchois *et al.*, 2001). Similarly, *E. coli* is considered a non-permanent member of the oesophagus, with Enterobacteriaceae only prominent in disease states such as oesophagitis or Barrett's oesophagus (Amir *et al.*, 2013; Di Pilato *et al.*, 2016). However, the case of the neonatal meningitis-associated strain, *E. coli* K1, offers a notable example of the commensal-to-pathogen switch involving the oesophagus. K1 strains can be present in the stools of healthy infants, children, and adults, but are responsible for up to 80% of cases of meningitis in neonates, primarily derived from direct transfer of the pathogen from mother to infant at birth (Glode *et al.*, 1977). With the aid of virulence factors such as Hek, K1

strains can invade and transcytose epithelial surfaces, access intravascular space, and then survive within the bloodstream, providing a route for access through the blood brain barrier (Fagan *et al.*, 2008; Kim, 2003). The susceptibility of neonates specifically is thought to be due to K1's ability to colonise the immature neonatal mucous barrier and other innate defenses (Birchenough *et al.*, 2013). The GI tract is thought to be the primary route of initial colonisation, and importantly, it has been shown that K1 can enter systemic circulation via the oesophagus in susceptible neonatal rat pups (Sarff *et al.*, 1975; Witcomb *et al.*, 2015).

Gastric acid of the stomach is the most inhospitable environment in the mammalian anatomy, with pH values as low as 1.5 - 2.5 (Martinsen *et al.*, 2005). Unlike other related Enterobacteriaceae such as *Salmonella* spp. (which can tolerate acid stress to pH 3), *E. coli* can tolerate extreme acid stress (approx. pH 2) for several hours (Audia *et al.*, 2001). Indeed, *E. coli* is as tolerant to acidic conditions as the notable stomach symbiont, *Helicobacter pylori* (Foster, 2004). *E. coli* bears three acid-resistance (AR) mechanisms, AR1, AR2, and AR3 (Figure 1.4; (Foster, 2004)).

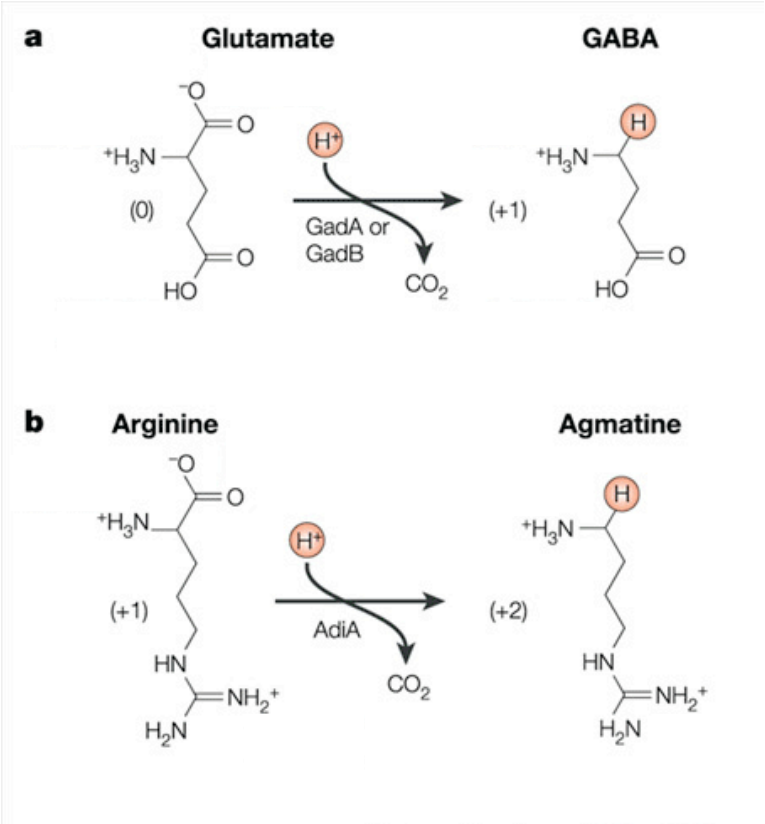


Figure 1.4. Consumption of protons during decarboxylation of **a)** glutamate (AR2) and **b)** arginine (AR3). GadA and GadB are glutamate decarboxylases involved in AR2, AdiA is the arginine decarboxylase. Adapted from Foster, 2004.

AR2 and AR3 are both decarboxylase/antiporter-dependent acid resistance mechanisms, each using decarboxylases to replace α-carboxyl groups on their respective amino acid substrates (glutamate for AR2 and arginine for AR3) with a proton recruited from the cytoplasm, producing CO_2 and γ-amino butyric acid (AR2) or agmatine (AR3) as end-products. The consumption of the proton results in an increase in internal pH (Foster, 2004). Until recently, the mechanisms underpinning AR1 were largely unknown. AR1 is active under conditions where cells grown to stationary phase in LB broth buffered at relatively low acidity (pH 5.5),

allowing cells to survive dilution into more highly acidic media (pH 2.5). In contrast, cells buffered at a higher pH are killed when diluted into the more acidic media. Similarly, cells grown under these conditions in the presence of glucose and not supplemented with glutamate or arginine are killed (Lin *et al.*, 1995). Recently, it was shown that AR1 utilises the same decarboxylase enzymes of AR2, but using an internally derived source of glutamate (Aquino *et al.*, 2017). Moreover, it was shown that there is a complex regulatory network underpinning AR1 and AR2, coordinated with carbon and nitrogen metabolism (Aquino *et al.*, 2017).

In a comparison with the highly acid-resistant EHEC strain O157:H7, commensal strains were shown to survive equally well in simulated gastric juice at pH 1.5, thus highlighting the universality of the acid stress response mechanism in *E. coli* (Foster, 2004; Lin *et al.*, 1996). However, the pH of gastric juice is not constant, and when elevated by the buffering effect of food or in conditions such as hypochlorhydria or achlorhydria, the risk of gastroenteritis caused by the outgrowth of foodborne bacteria increases (Martinsen *et al.*, 2005; Nalin *et al.*, 1978; Sarker & Gyr, 1992; Waterman & Small, 1998). While acid is considered to be the primary mechanism for antimicrobial control in gastric juice, other components such as the protease, pepsin, contribute to an antimicrobial effect (Zhu *et al.*, 2006). It is not clear whether or not *E. coli* harbours specific response mechanisms to pepsin, but the stomach environment remains a crucial barrier to colonisation.

1.2.2. The duodenum and bile resistance

1.2.2.1. The duodenum

The duodenum is an important junction that (i) connects the stomach and the jejunum, (ii) is linked to the liver and pancreas, and (iii) is largely responsible for the breakdown of food by the small intestine and regulates the emptying rate of the stomach. Significant environmental stresses are exerted within the duodenum, including the release of digestive enzymes from the pancreas such as trypsin and lipase, which can damage bacterial proteins and the cell envelope. In addition, bile is secreted into the duodenum from the gallbladder, which serves a dual role in the breakdown of fats and acting as an antimicrobial agent (Begley *et al.*, 2005). The duodenum contains a lower microbial load compared to other regions in the GI tract in humans, yet it has been shown to harbour a highly diverse microbiota, including a detectable level of *Escherichia/Shigella* (Li *et al.*, 2015). Interestingly, a study of commensal *E. coli* in pigs showed that some commensal strains of *E. coli* (from phylogroup A) were more likely to be found in the duodenum/ileum (both regions differed little in dispersion of strains) than in the colon/faeces (Dixit *et al.*, 2004). *E. coli* contains specific mechanisms to withstand the selective pressures within the duodenum. For example, ecotin is a serine protease inhibitor produced by some strains of *E. coli* which allows for resistance to trypsin, but also other pancreatic-derived proteases, such as chymotrypsin and kallikrein (McGrath *et al.*, 1995). Moreover, *E. coli* contains mechanisms to prevent degradation of its lipopolysaccharide (LPS) by lipases, including modification of acyl chains of the integral LPS

component, lipid A (Raetz *et al.*, 2007). The response to bile, however, has been studied in far more detail (Table 1.1).

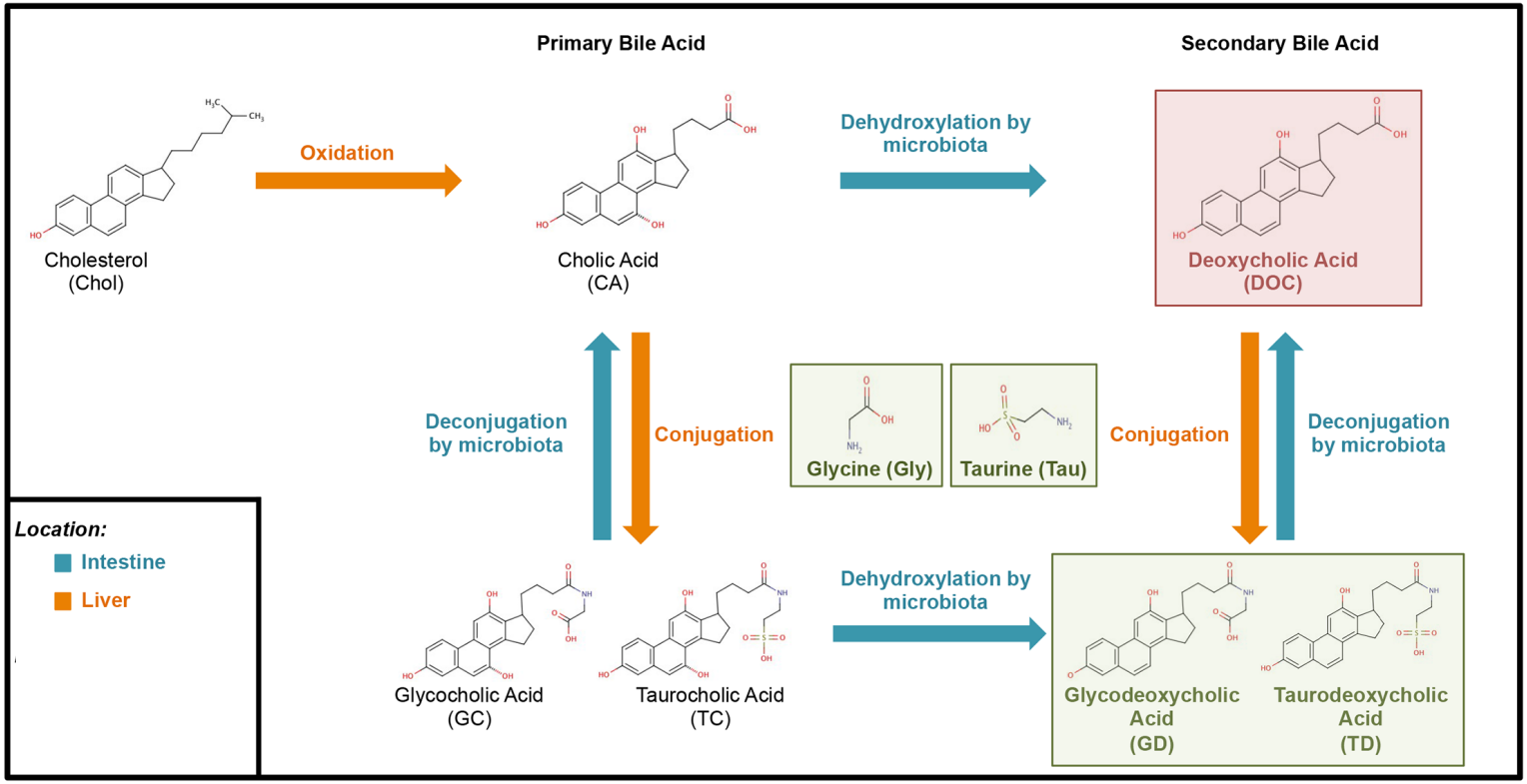
Gene/Function	Function	Reference
Exclusion of bile acids		
LPS	Slows diffusion of bile acids across	(Begley *et al.*, 2005)
OmpC	outer membrane	(Nikaido, 2003)
MicF	Narrower porin inhibits bile acid entry	(Bernstein *et al.*, 1999)
	Negatively regulates the wider porin, OmpF	
Exclusion of bile/efflux		
AcrAB-TolC	Actively effluxes bile acids	(Thanassi *et al.*, 1997)
EmrAB-TolC	Actively effluxes bile acids	(Lee *et al.*, 2000)
MdtABC-TolC	Increases resistance to bile acids	(Baranova & Nikaido, 2002)
MdtM	Works in concert with AcrAB-TolC to efflux bile acids	(Paul *et al.*, 2014)
Repair and defense against damage		
dinD	DNA damage-inducible protein	(Bernstein *et al.*, 1999)
impB	Error-prone DNA repair	(Foster, 2007)
hupAB	Controls DNA supercoiling	(Begley *et al.*, 2005)
osmY	Response to hyperosmotic stress	(Begley *et al.*, 2005)
Modulation of virulence		
Alter motility expression		(Hamner *et al.*, 2013)
Alter expression of iron acquisition genes		(Hamner *et al.*, 2013)

Table 1.1. Bile resistance mechanisms in *E. coli*.

1.2.2.2. E. coli growth in the presence of bile

Bile is a complex mixture of bile acids, cholesterol, fatty acids, phospholipids, biliverdin, and multiple other substances. Bile is synthesised in pericentral liver hepatocyte cells and is released into the duodenum from the gall bladder via the common hepatic duct (Figure 1.5B). However, bile acids are absorbed along the entire length of the gut, conserving bile acid concentrations under normal conditions. Approximately 50% of the organic components of bile consist of bile acids, whose core structure consists of a steroid nucleus, which can be conjugated via a peptide bond to either glycine or taurine (Figure 1.5A). Primary bile acids are synthesised directly from cholesterol in the liver and include cholic acid and chenodeoxycholic acid. These primary bile acids may then be modified by bacteria in the intestine e.g. via deconjugation, dehyrdroxylation, and dehydrogenation to produce secondary bile acids, such as deoxycholic acid and lithocholic acid (Figure 1.5A; (Begley *et al.*, 2005)).

A


Primary Bile Acid
Secondary Bile Acid
Cholesterol
(Chol)
Oxidation
Cholic Acid
(CA)
Dehydroxylation by
microbiota
Deoxycholic Acid
(DOC)
Deconjugation
by microbiota
Conjugation
Glycine (Gly)
Taurine (Tau)
Conjugation
Deconjugation
by microbiota
Location:
Intestine
Liver
Glycocholic Acid
(GC)
Taurocholic Acid
(TC)
Dehydroxylation by
microbiota
Glycodeoxycholic
Acid
(GD)
Taurodeoxycholic
Acid
(TD)

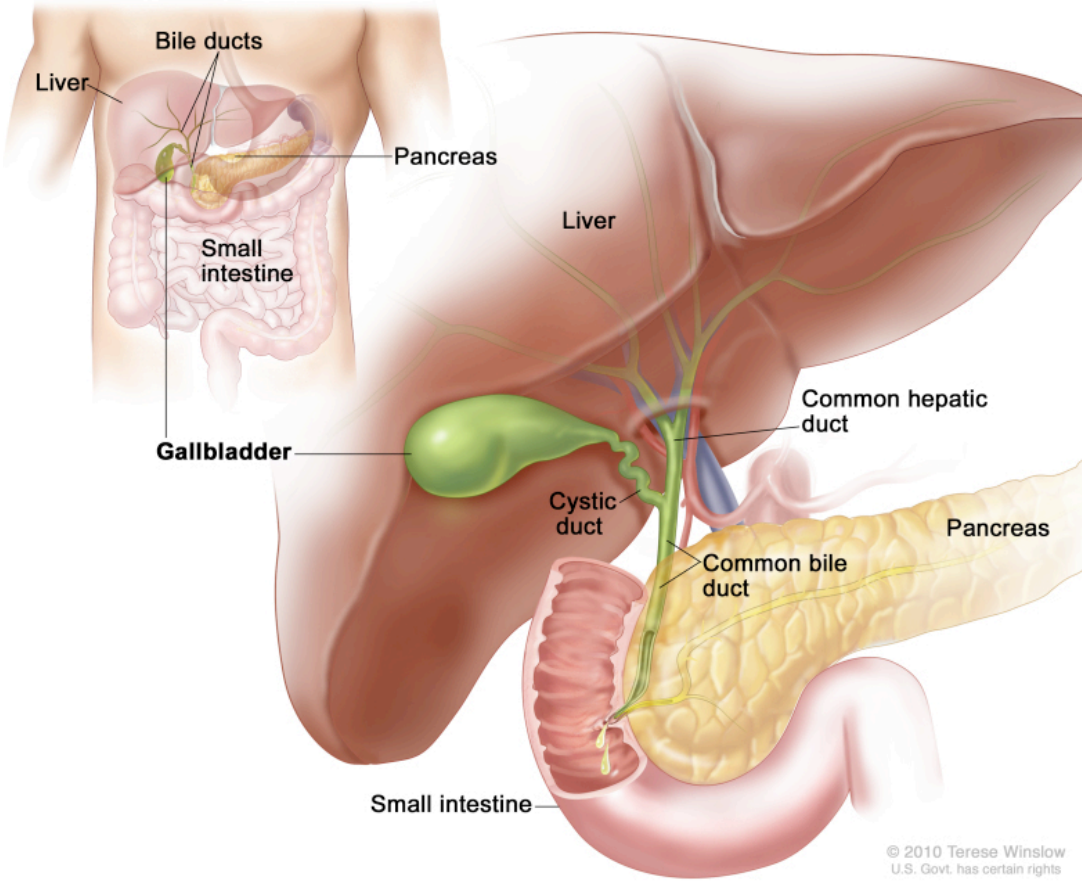


Figure 1.5. (A) Bile acid metabolism (adapted from (Bachmann *et al.*, 2015)). **(B)** Anatomy of the gallbladder (from https://www.ncbi.nlm.nih.gov/books/NBK65963/figure/CDR0000270720__198/ (accessed 25 Feb 2018).

The primary role of bile is to emulsify and solubilse fats from ingested food, allowing for their absorption in the small intestine (Russell, 2003). However, bile also acts as one of the most potent innate barriers to microbial colonisation of the GI tract, having a profound impact upon the ecology of the gut (Sarker & Gyr, 1992). Several studies have shown a role for bile in controlling microbial load in the gut, either by the increase in growth when bile is restricted (Berg, 1995; Inagaki *et al.*, 2006), or reduction in bacterial overgrowth and translocation when bile acids were administered (Ding *et al.*, 1993; Lorenzo-Zúñiga *et al.*, 2003). Moreover, as the concentration of bile salts decreases further away from the duodenum and into the intestine, the microbial load increases (Hay & Zhu, 2016). Furthermore, bile acids can alter the community structure of the microbiome, as has been shown in rats, where significant increases in phylum *Firmicutes* and specific classes such as Clostridia were observed when the rats were fed increased levels of the bile acid, cholic acid (Islam *et al.*, 2011; Ridlon *et al.*, 2014). Bile can also determine the spatial distribution of different strains of the same species, as was shown in germ-free mice colonised with different *E. coli* mutants, whereby spatial distribution of the different mutants was determined by the tradeoff between the ability of those strains to withstand bile acids and their ability to compete for nutrients (De Paepe *et al.*, 2011).

Bile is a potent antimicrobial agent for multiple reasons. Bile acts as a signal to other cells and systems within the GI tract, promoting an immune response to infection e.g. via the promotion of cathelicidin production in the hepatobiliary tract, or the activation of genes in the

ileum involved in enteroprotection via the farnesoid X receptor (D'Aldebert *et al.*, 2009; Inagaki *et al.*, 2006). Bile also displays physiochemical properties that are antimicrobial, primarily due to the lipid-dissolving action of bile acids resulting in bacterial cell lysis, or the presence of immobilising immunoglobulin A and mucous within bile (Begley *et al.*, 2005; Hofmann, 1999). However, bile also elicits secondary effects, such as DNA damage, denaturing of proteins, induction of oxidative stress, pH stress, osmotic stress, or even the lowering of available concentrations of calcium and iron (Begley *et al.*, 2005).

Bacteria employ multiple systems in order to survive in the presence of bile (Table 1.1). The different responses can be summarized into four main strategies: exclusion of bile, extrusion of bile, repair and defense against damage, and modulation of virulence (Hay & Zhu, 2016). In Gram-negative bacteria, bile acids are either prevented from diffusing, or diffuse far more slowly, through the Gram-negative outer membrane due to the presence of the thick LPS layer (Nikaido, 2003). This is highlighted by the hypersensitivity to bile acids reported for LPS deep-rough core mutants of *E. coli* K-12 substr. MG1655 (Møller *et al.*, 2003). Moreover, *pmrA* mutants, showing aberrant modification of LPS, show increased sensitivity to the bile acid derivative, deoxycholate (Froelich *et al.*, 2006). Porins are also important in excluding entry of bile acids, and both *E. coli* and *Salmonella* favour the narrower β-barrel porin, OmpC, over the wider-pored OmpF, in order to increase tolerance to bile acids (Nikaido, 2003; Thanassi *et al.*, 1997). Once past the outer membrane,

bile acids can reach, insert into, or pass through, the inner membrane into the cytoplasm. As a protective measure against this scenario, *E. coli* can actively efflux bile acids from the cytoplasm. The best-characterised and most important efflux system involved in bile efflux is the AcrAB-TolC efflux pump (Eicher *et al.*, 2009; Thanassi *et al.*, 1997). The AcrAB-TolC pump recognises multiple substrates; however, it has a specific affinity for bile acids, demonstrating that a specific response to bile is elicited by the cell. Other auxiliary efflux mechanisms are employed by *E. coli*, including: the EmrAB-TolC efflux pump which actively effluxes bile salts and works in parallel with AcrAB-TolC (Lee *et al.*, 2000; Thanassi *et al.*, 1997); the major facilitator superfamily transporter, MdtM, which acts synergistically with AcrAB-TolC to efflux bile salts (Paul *et al.*, 2014); and MdtABC-TolC that, when overproduced, increases resistance to a number of bile salts (Baranova & Nikaido, 2002). If efflux mechanisms fail, bile acids can damage proteins, DNA, membranes, and cause protein aggregation (Begley *et al.*, 2005). In response to this damage, *E. coli* does induce the SOS response and multiple DNA repair mechanisms, as shown by the upregulation of several stress response genes required for the activation of the SOS response, DNA repair, and oxidative stress response mechanisms (Bernstein *et al.*, 1999; Foster, 2007; Merritt & Donaldson, 2009). Finally, pathogens can respond in specific ways to the presence of bile by modulating their virulence properties. For example, *Salmonella* species downregulate the expression of their pathogenicity islands, which allows for increased expression of unique, *Salmonella*-associated genes involved in stress response and survival in the presence of bile

(Hernández *et al.*, 2012; Prouty *et al.*, 2006). Bile salts have also been shown to modulate the expression of virulence genes in the EHEC strain, O157:H7, altering expression of flagellar and iron acquisition genes (Hamner *et al.*, 2013).

Enterobacteriaceae, including *E. coli*, show high resilience to the stresses induced by bile (Kramer *et al.*, 1984). For example, resistance to detergents has been shown to be widespread across the *E. coli* phylogroups (D'Mello & Yotis, 1987; Jacobsen *et al.*, 2009). Indeed, this resilience does pose a health risk. For example, *Salmonella enterica* Serovar Typhi can reside within the gallbladder, which can act as a reservoir in asymptomatic carriers (Dougan & Baker, 2014). Moreover, *E. coli* is one of the most commonly isolated microorganisms from bile in patients with community-acquired cholangitis (inflammation of the biliary tract) and choledocholithiasis (gallstones within the bile duct (Kaya *et al.*, 2012; Razaghi *et al.*, 2017)). However, while pathogenic *E. coli* do display some specialised responses to bile in terms of controlling the expression of some virulence genes, the major bile resistance mechanisms described above are shared between pathogens and commensals (Sistrunk *et al.*, 2016). Indeed, one study has shown that many of *E. coli* strains involved in biliary tract infections were not associated with any of the pathotypes mentioned previously, indicating that they may be commensal or pathobiont strains (Razaghi *et al.*, 2017).

1.2.3. Colonisation and growth within the intestine

1.2.3.1. E. coli attachment and growth

E. coli can be found in substantial numbers in the duodenum, jejunum, and ileum (Conway & Cohen, 2015). Multiple factors influence the numbers and distribution of microorganisms in these habitats (Donaldson *et al.*, 2016). Owing to the different roles of the small intestine and the large intestine, namely absorption of the products of digestion (vitamins, carbohydrates, proteins, and lipids) and the absorption of water and inorganic salts, respectively, the physical and histological structures of these environments differ (Figure 1.6).

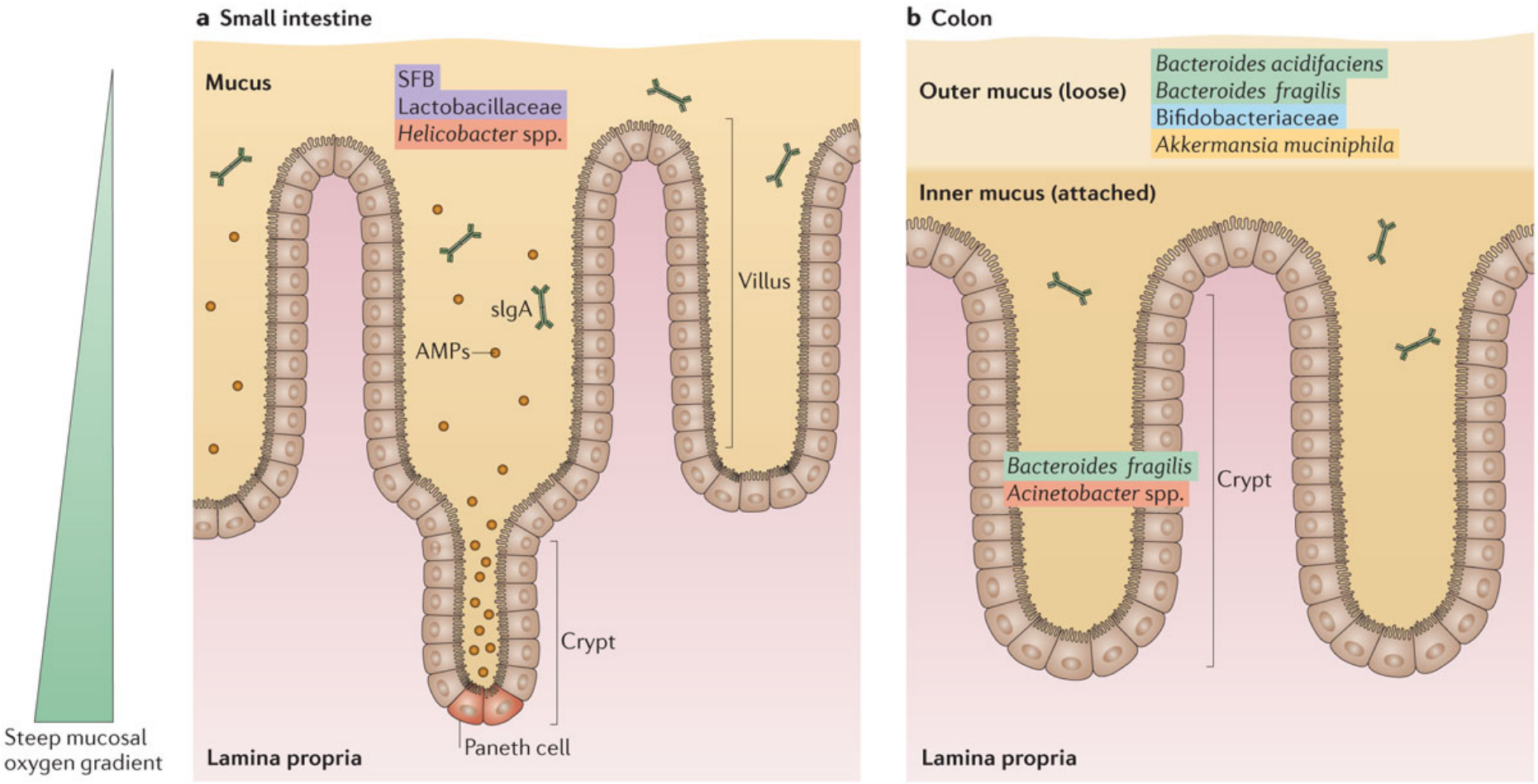


Figure 1.6. The mucosal epithelium of the (A) small intestine and (B) colon. From Donaldson *et al*., 2016.

Four microhabitats exist within the intestine: the surface of the epithelial cells, the deep mucous layer of the crypts, the mucous layer, and the lumen (Conway *et al.*, 2013). To control excessive and aberrant overgrowth of the microbiota in each habitat, the intestine and mucosa provide both physical barriers to growth and precise immune surveillance and processing mechanisms. Peristalsis, required for moving luminal contents through the digestive tract, has a significant effect on bacterial ecology by controlling the rate of flow and mixing of luminal contents; excessive peristalsis would lead to washout of the microbiota in mucous, whilst too little peristalsis would result in insufficient mixing and microbial overgrowth (Cremer *et al.*, 2016; Kim *et al.*, 2016). Related to this is the rapid and high rate of turnover of the epithelium that is required to maintain tissue homeostasis. The continuous sloughing off of epithelial cells facilitates pathogen expulsion and can localise and confine inflammation or infection (Kim *et al.*, 2010). Therefore, the growth rate of *E. coli* is set to maintain its population level ahead of the rate of mucous turnover (Conway & Cohen, 2015).

E. coli must embed itself within mucous in order to colonise efficiently and this is achieved using motility, fimbriae, LPS, and/or capsules (Conway *et al.*, 2013). Serine protease autotransporters of Enterobacteriaceae (SPATE) proteins are another colonisation mechanism widespread in *E. coli*, facilitating colonisation of mucous via a variety of functions such as immunoglobulin binding, mucous degradation, or utilising mucous as a nutrient source (Dautin, 2010). Pathogenic *E. coli* can also utilise specific colonisation mechanisms, for

example EPEC strains can directly attach to enterocytes using the bundle-forming pilus (Kaper *et al.*, 2004). Furthermore, mucous contains high concentrations of antimicrobial agents such as IgA, defensins, and cathelicidins, against which *E. coli* harbours multiple resistance or adaptation strategies. For example, *E. coli* may use Type 1 fimbriae to attach to mannosylated secretory IgA to enhance colonisation or increase turnover of fimbriated cells, and many strains produce the outer membrane (OM) protease, OmpT, which can degrade or otherwise process human antimicrobial peptides, including the human cathelicidin, LL-37 (Friman *et al.*, 1996; Gruenheid & Le Moual, 2012; Mason & Huffnagle, 2009; Thomassin *et al.*, 2012). Finally, specialised immune cells, such as M cells, Paneth cells, or dendritic cells, constantly sample the microbiota and their byproducts throughout the mucosa, presenting antigens to the adaptive immune system in specialised lymphoid tissue known as Peyer's Patches (Kim *et al.*, 2010).

Pathogenic and pathobiont *E. coli* strains elicit interesting mechanisms for evading or exploiting these immune mechanisms. For example, AIEC can utilise *de novo* pyrimidine biosynthesis to survive and replicate within macrophages in Crohn's disease patients (Thompson *et al.*, 2016). Some UPEC and diarrhoeagenic *E. coli*, such as K1, K4, or K5 strains, express capsules that mimic host tissue, allowing for the evasion or mitigation of phagocytosis and complement mediated killing (Cress *et al.*, 2014; Miajlovic *et al.*, 2014). *E. coli* can also escape phagocytosis by inhibiting immune signaling, such as the dampening of TLR-mediated cytokine release via the expression of Tir by EPEC strains, or the

inhibition of macrophage phagocytosis by manipulation of the FcγRIII-FcRγ complex (Van Avondt *et al.*, 2015). Commensal strains can also manipulate the immune system. For example, *E. coli* Nissle 1917 (EnN) can stimulate the production of β-defensin and chemokines in epithelial cell lines, resulting in the promotion of immune cell recruitment and the strengthening of the epithelial cell barrier (Sassone-Corsi & Raffatellu, 2015). Moreover, early-life colonisation by commensal *E. coli* and *Bifidobacterium infantis* has been associated with increased numbers of mature $CD20^+$ B cells, indicating a potential interaction between early-life colonising *E. coli* and the developing immune system, resulting in colonisation resistance to pathogens (Lundell *et al.*, 2012; Sassone-Corsi & Raffatellu, 2015).

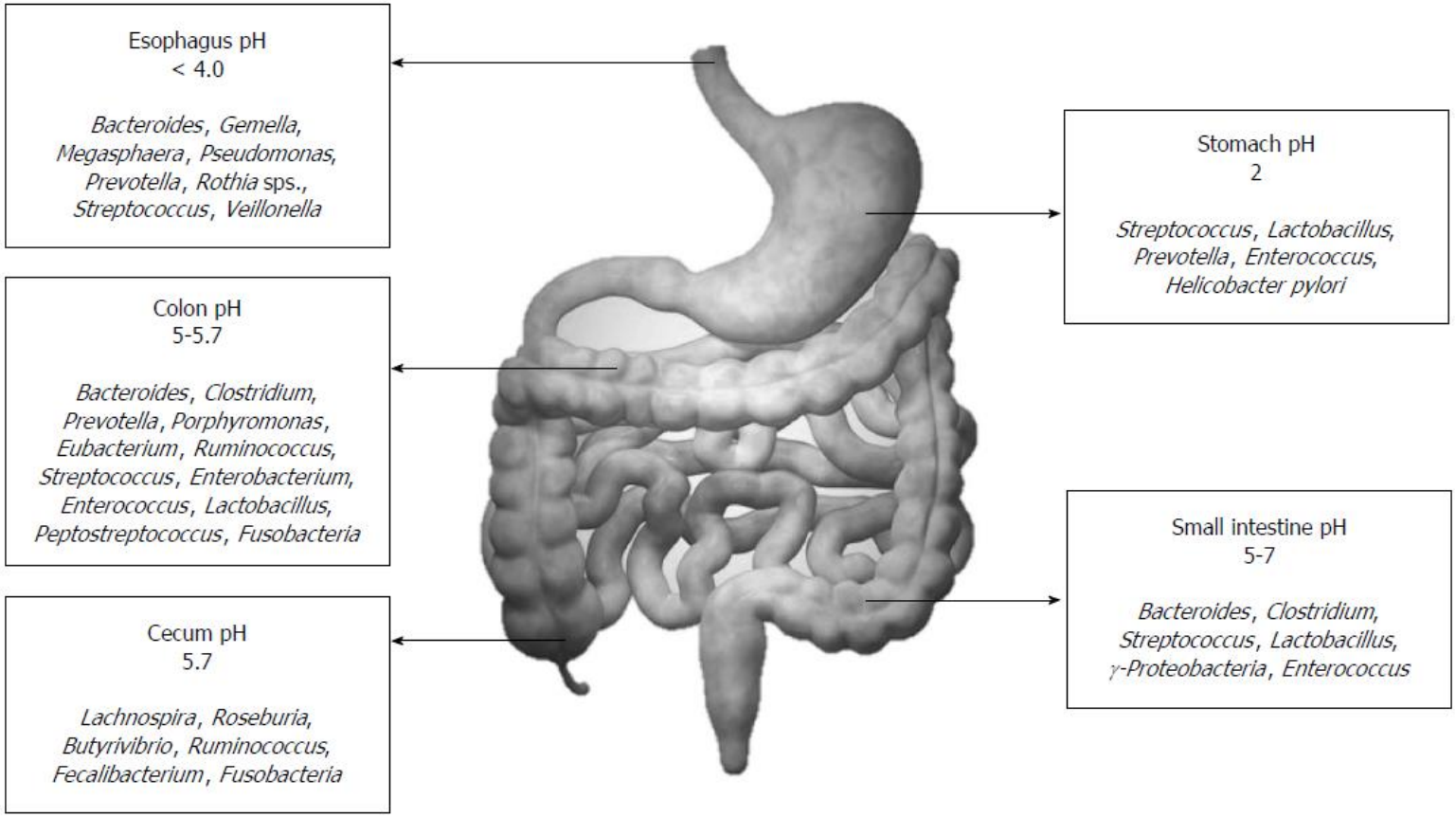


Figure 1.7. Distribution and composition of the healthy human gut microbiota. Adapted from Jandhyala *et al*., 2015.

1.2.3.2. Nutrient availability and the Restaurant Hypothesis

The environmental conditions of the intestine promotes the growth of huge numbers of microorganisms (Figure 1.7). Microbial load in the duodenum is as low as 10^3 cells per gram luminal content, increasing to 10^7 cells per gram in the ileum, to as many as 10^{12} cells per gram of luminal content in the colon (Sekirov *et al.*, 2010). The microbiota differs between the small and large intestines: the microbiota of the small intestine is not as well characterised as other regions, yet it is thought to contain predominantly *Streptococcus*, whereas the large intestine contains a much greater microbial load and diversity, dominated by anaerobes from the phyla Firmicutes (e.g. *Clostridia*) and Bacteroidetes (e.g. *Bacteroides*) (Hollister *et al.*, 2014; Jandhyala *et al.*, 2015). However, *E. coli* accounts for only 0.1% of the healthy intestinal microbiota, indicating that it does not utilise available nutrients within the large intestine as effectively as other species (Human Microbiome Project Consortium, 2012).

E. coli most commonly exists as a minority member of large, mixed-species biofilms within the gut mucosa (Conway & Cohen, 2015). In fact, the composition of the microbiota exerts the greatest influence on the engrafting ability of an *E. coli* strain, since it determines the nutritional niches that are available to the incoming strain (Conway & Cohen, 2015). In the colon, the mono- and di-saccharides that *E. coli* requires for growth are not provided directly from ingested food, but via the breakdown of complex polysaccharides in mucous or dietary fibre by anaerobes (Conway & Cohen, 2015). Therefore, it has been proposed that the

population of anaerobes within the mixed-species biofilm determines the profiles of sugars available for use by *E. coli* for growth (Conway & Cohen, 2015). The mixed-species biofilms are known as 'Restaurants', and the Restaurant Hypothesis states that the collection of strains of *E. coli* that reside within the intestine are those that can best assimilate the nutrients available within each Restaurant (Conway & Cohen, 2015). Studies using the streptomycin-treated mouse model support the hypothesis, showing that the effective use of nutrients is the primary determinant of colonisation by *E. coli* (Fabich *et al.*, 2008; Maltby *et al.*, 2013).

1.2.3.3. E. coli adaptation to changes in the intestine: nitrate respiration and gut inflammation

The GI tract is not a steady-state environment, with factors such as diet, antibiotic use, disease, age, and inflammation constantly altering environmental conditions within the gut (Conlon & Bird, 2014; Faber & Bäumler, 2014). Inflammation is an important mechanism to control the growth of microbes and resolve infections. However, a consequence of inflammation is that is can alter the gut environment in two significant ways: (i) by limiting the availability of trace elements such as iron and zinc, and (ii) through the generation of byproducts that alter the redox environment (Figure 1.8).

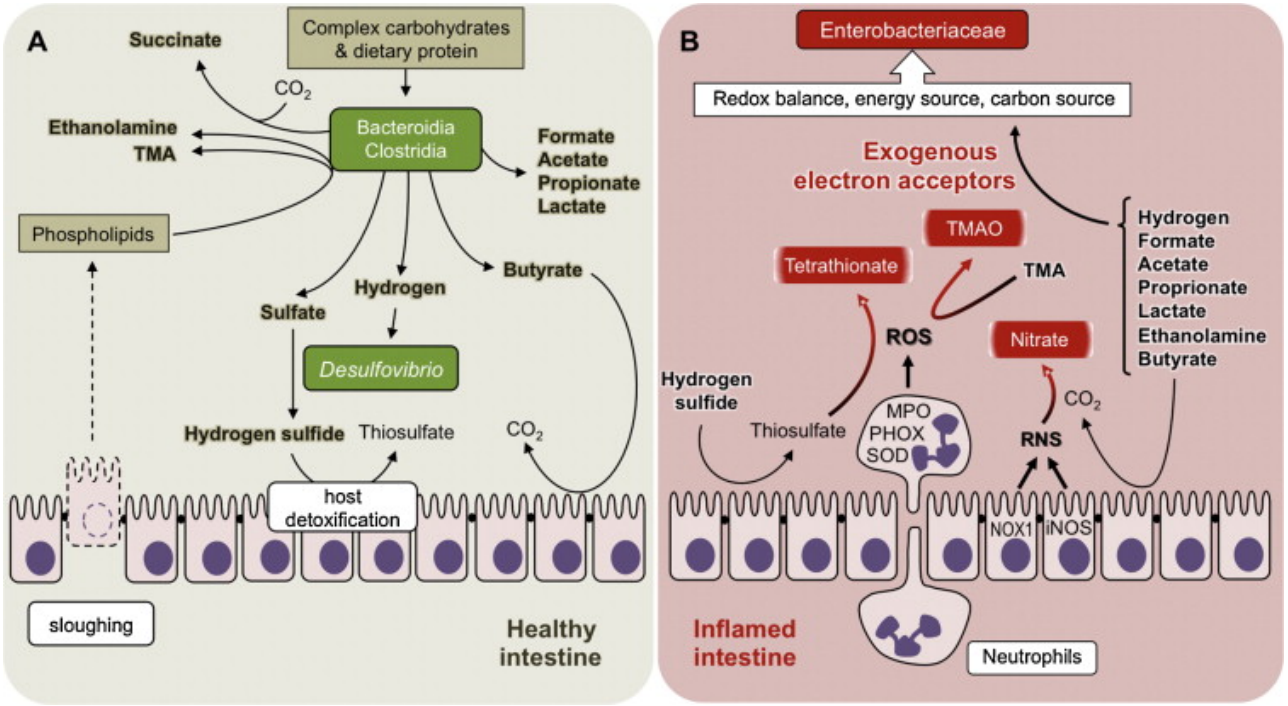


Figure 1.8. The inflammatory host response creates a new metabolic niche in the intestine. **A** The healthy intestine. **B** The inflamed intestine. Adapted from Faber and Bäumler, 2014.

Alterations in the redox environment are largely due to changes in the expression of host genes such as NOX1 (NADPH oxidase 1), DUOX2 (dual-function NADPH oxidase 2), and iNOS (inducible nitric oxide synthase). Together, the proteins produced by these genes produce reactive oxygen species (ROS) and reactive nitrogen species (RNS), which create a hostile environment at the mucosal surface in order to limit microbial growth (Faber & Bäumler, 2014). However, the interaction of these various byproducts can also produce a variety of oxidised compounds that can be used as terminal electron acceptors by some bacteria. For example, iNOS generates nitric oxide that can react with superoxide radicals produced by NOX1 and DUOX2 to generate peroxynitrate ($ONOO^-$), which itself can be further converted to nitrate (Figure 1.9). Thus, intestinal inflammation promotes the production of nitrate, a molecule that can be used as a terminal electron acceptor during anaerobic growth by facultative anaerobes such as *E. coli* (Faber & Bäumler, 2014; Winter *et al.*, 2013).

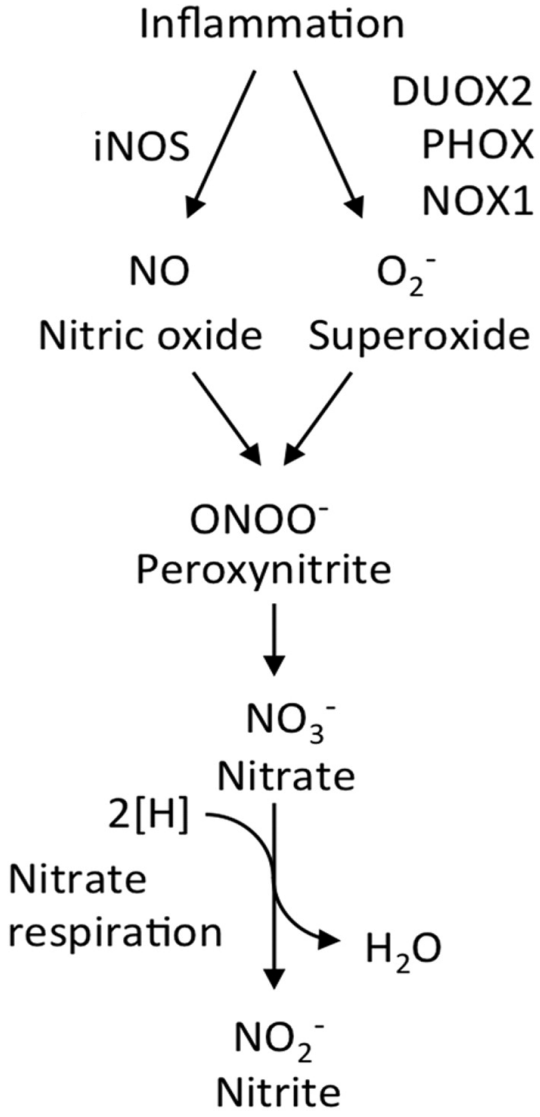


Figure 1.9. The generation of nitrate during inflammation in the intestine (adapted from (Lopez *et al.*, 2012)).

Anaerobic nitrate respiration allows *E. coli* to both enhance its metabolic capacity and to boost its growth in the gut. Respiration allows for the maintenance of redox balance by transferring electrons from NADH to terminal electron acceptors such as nitrate, while at the same time preserving phosphoenolpyruvate (PEP) for anabolic reactions (Faber & Bäumler, 2014). Respiration also allows *E. coli* to use a broader range of carbon sources, such as the non-fermentable sugar, glycerol, or formate and hydrogen, as electron donors in the anaerobic respiratory electron transport chain (Cole & Richardson, 2013). This metabolic flexibility greatly enhances the competitiveness of *E. coli*, allowing the population of *E. coli* to expand in the gut during inflammation, i.e. resulting in an enterobacterial bloom (Jones *et al.*, 2011; Winter *et al.*, 2013).

During periods of inflammation in the gut, blood plasma nitrate levels have been shown to increase (Dykhuizen *et al.*, 1996). Moreover, increased levels of nitrate can be detected in the ceacal mucous of mice treated with dextran sulfate sodium (DSS), a trigger for gut inflammation (Dykhuizen *et al.*, 1996; Winter *et al.*, 2013). Humans with chronic gut inflammation characteristically have an increased abundance of Enterobacteriaceae, which may contribute to the pathogenesis of IBD (Alhagamhmad *et al.*, 2016; Kotlowski *et al.*, 2007). Moreover, nitrate produced in the inflamed mouse intestine has been shown to directly boost the growth of *E. coli* (Winter *et al.*, 2013). Therefore, it has been proposed that chronic inflammation alters the environment of the gut to favour the growth of *E. coli*, in particular via the provision of increased

amounts of nitrate which allow the bacterium to outgrow other members of the microbiota by anaerobic respiration. This may then lead to an increased abundance of *E. coli* under chronic inflammation, which, in turn, may promote or worsen the symptoms of IBD.

1.3. Functional genomics and transposon sequencing

1.3.1. Functional genomics

Functional genomics aims to describe gene functions and interactions using a wide array of genomic and transcriptomic technologies. Mutagenesis is a core component of the functional genomics toolkit, allowing for 'loss-of-function' annotations of genes, i.e. disruption or deletion of a genes results in reduced growth or physiological changes of the mutant that provide clues as to the function of the gene being investigated. Among the most widely used techniques for gene deletion or disruption in *E. coli* include λ Red recombinase-based gene deletion and transposon mutagenesis (Datsenko & Wanner, 2000; Goryshin *et al.*, 2000). These techniques have been used to generate libraries representing mutants in the majority of genes in *E. coli*, allowing for comprehensive analyses of genetic requirements for growth under different conditions (Baba *et al.*, 2006; Joyce *et al.*, 2006). Until relatively recently, this has most commonly been conducted by screening libraries of defined knockout mutants, cultured separately in arrayed mutant pools. A notable example of this is the Keio collection, a library of 3864 single-gene deletion mutants constructed in the *E. coli* K-12 strain, BW25113, which has been used to broaden our understanding of growth

requirements of *E. coli in vitro* (Joyce *et al.*, 2006; Long & Antoniewicz, 2014). However, this approach has some limitations, including being labour-intensive to construct and screen, and only providing resolution to the gene level (Barquist *et al.*, 2013). More recently, approaches have been developed to overcome these issues, including what will be collectively termed here as 'transposon sequencing'.

1.3.2. Transposon sequencing

Transposon sequencing combines random transposon mutant libraries with massively parallel sequencing (Barquist *et al.*, 2013; van Opijnen & Camilli, 2013). These methods require transposon mutant libraries which are pooled together to allow for simultaneous and rapid screening of all mutants at the same time. Once screened, libraries are then processed, sequenced to identify transposon-genomic DNA junctions, and analysed, to precisely locate and quantify transposon insertions, allowing for the determination of genetic requirements for growth under a particular condition (Figure 1.10).

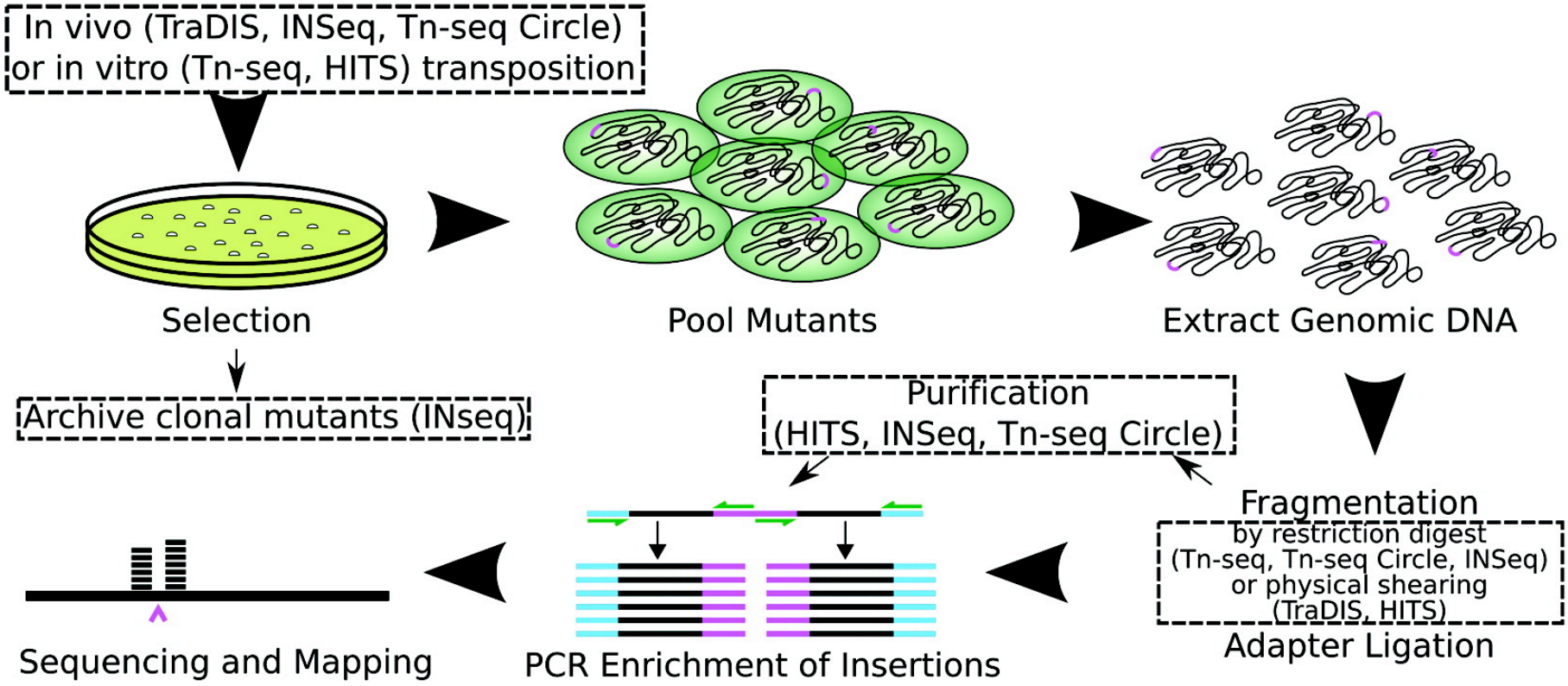


Figure 1.10. An illustration of the typical transposon sequencing protocols. Adapted from Barquist *et al*., 2013.

Transposon sequencing offers several advantages over defined mutant libraries. Transposon mutagenesis is a more rapid mutagenesis method compared to targeted gene deletions, and since transposon sequencing uses pooled mutant libraries, library screening is much more rapid. Massively parallel sequencing allows for extremely high depth of coverage, and since the precise location of transposon insertions can be determined, it can provide high-level resolution on requirements for protein domains, promoters, non-coding RNA, and intergenic regions. Moreover, multiplexing allows for several libraries to be sequenced simultaneously, allowing for much greater scalability of genetic screens to incorporate several growth conditions (Gray *et al.*, 2015).

Several iterations of transposon sequencing, including 'Tn-seq' (transposon sequencing), 'HITS' (high-throughput insertion tracking by deep sequencing), 'INSeq' (insertion sequencing), and 'TraDIS' (transposon-directed insertion site sequencing), have been developed and, whilst their guiding principles are similar, there are minor variations associated with each method (see Figure 1.10; (Barquist *et al.*, 2013; van Opijnen & Camilli, 2013)). Transposon sequencing is typically applied to mutant libraries constructed using either the Tn5 or Himar1 *Mariner* transposons, since both transposons can be used in broad range of species, and insert with low bias for any particular region or sequence within the genome (Goryshin *et al.*, 2000; Lampe *et al.*, 1998). Tn-seq and INSeq utilise the *Mariner* transposon specifically, whereas HITS and TraDIS can use any transposon or insertional method. Once the mutant

library has been constructed by growing mutants on agar plates, the mutant colonies are pooled to create a library of mutated cells that are incubated under selective growth conditions. Genomic DNA is then extracted from the mutant pool, fragmented, and transposon insertion-genomic DNA junctions are enriched (Figure 1.10). Tn-seq and INSeq fragment DNA using the type II restriction enzyme, *MmeI*, which cleaves 20 bp downstream from its recognition site incorporated near the terminal repeats of the *Mariner* transposon. On the other hand, during TraDIS and HITS, DNA is fragmented physically by shearing and then DNA fragments of a specific size are selected by, for example, agarose gel electrophoresis (van Opijnen & Camilli, 2013). In all methods, sequencing adaptors are ligated to fragmented DNA ends, allowing for amplification of transposon-gDNA junctions using adaptor-specific and transposon-specific primers. To ensure greater amplification of transposon-containing DNA fragments only, TraDIS employs splinkerette adaptors, which only allow hybridisation of the reverse strand primer when the forward strand primer has generated a complementary strand (see Figure 2.2, Materials and Methods) (Barquist *et al.*, 2016). Finally, following purification of PCR products, transposon-enriched DNA is subject to high-throughput sequencing using standard sequencing primers in conjunction with transposon-specific primers. This generates reads, or short sequences, corresponding to the transposon-gDNA junction, which are digitally counted to determine both the precise location of transposon insertions in

the genome, as well as the frequency of those insertions in the mutant library.

The design of a transposon sequencing experiment and the means by which the sequencing reads are analysed influence how genetic requirements for growth under a particular condition are determined (Figure 1.11). For example, a gene may be nominated **qualitatively** as 'essential' or 'non-essential' for survival, depending on whether or not transposons are absent from that gene to a sufficient degree following growth under a particular condition. For instance, TraDIS studies have used the 'insertion index', the number of transposon insertions within a protein coding sequence (CDS) as a qualitative measure of gene essentiality (Goodall *et al.*, 2018; Langridge *et al.*, 2009). However, following growth under a selective condition, transposons may not be entirely absent from a gene but altered in read counts to a statistically significant degree. This allows for genetic requirements for growth to be determined **quantitatively**, typically calculated as the ratio of observed reads in an input (control) mutant pool compared to an output (test) pool (Figure 1.11). These scores are usually expressed as $\log_2$ fold change (logFC) or as a 'fitness value'. To help refine these measures of essentiality or fitness, transposon insertions and read counts can be visualised using software such as Artemis, allowing for the determination of essentiality of protein domains or noncoding regions where transposon insertions would be absent (for example, see Figure 1.12 (Goodall *et al.*, 2018)).

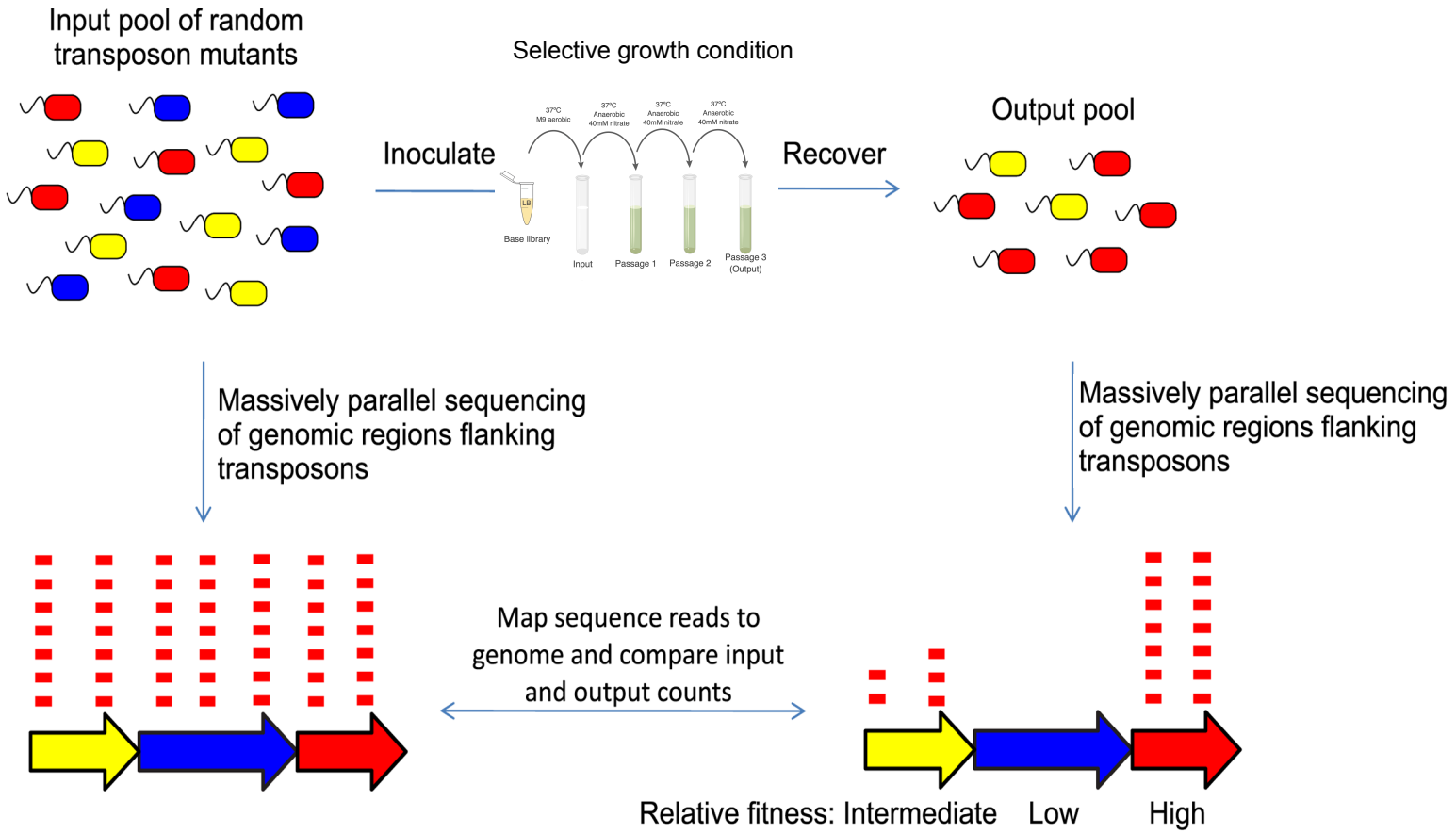

Input pool of random transposon mutants
Selective growth condition
Inoculate
37°C M9 aerobic
37°C Anaerobic 40mM nitrate
37°C Anaerobic 40mM nitrate
37°C Anaerobic 40mM nitrate
LB
Base library
Input
Passage 1
Passage 2
Passage 3 (Output)
Recover
Output pool
Massively parallel sequencing of genomic regions flanking transposons
Massively parallel sequencing of genomic regions flanking transposons
Map sequence reads to genome and compare input and output counts
Relative fitness: Intermediate
Low
High

Figure 1.11. Experimental strategy for TraDIS mutant screens. An input pool of random transposon insertion mutants is generated, and used to inoculate *in vitro* experimental conditions. Output pools of bacteria that are capable of survival and growth in each condition are harvested and their gDNA isolated. Massively parallel sequencing of the regions flanking each transposon allow the disrupted genes to be identified. Genetic requirements for growth can be measured qualitatively based on the presence or absence of transposon insertions (e.g. the non-essential yellow and red genes - represented by arrows - contain transposon insertions, unlike the essential blue gene which contains no insertions). Comparison of the sequence counts derived from the input and output pools can also allow the relative fitness of each mutant to be assessed. Adapted from (Chaudhuri *et al.*, 2013).

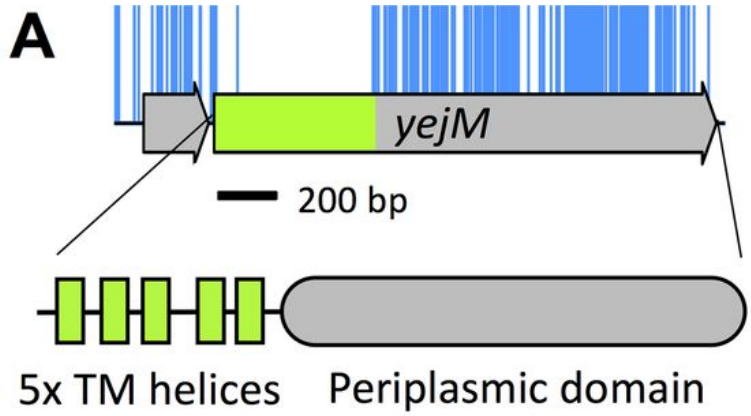


Figure 1.12. Additional features identified through detailed analysis of high-resolution insertion data. Insertions within the *yejM* protein coding sequence (CDS) are localised to a nonessential periplasmic domain. The 5′ end of the CDS has no insertions and corresponds to the five essential transmembrane (TM) domains of YejM. Blue lines indicate transposon insertion sites. Grey arrows indicate non-essential genes/domains, while green arrows indicate essential genes/domains. Adapted from Goodall *et al*., 2018.

1.3.3. TraDIS studies of E. coli

TraDIS has been applied to investigate important biological questions in a number of different bacteria, including *E. coli* (some published TraDIS studies are outlined in Table 1.2). The application of TraDIS has led to the discovery of new gene functions in *E. coli* that may have not been revealed using other methods. For example, TraDIS has been used to define the complement of genes required for capsule biosynthesis in the

K1-producing UPEC strain, PA45B, uncovering previously unidentified roles for two regulators, *mprA* and *lrhA* (Goh *et al.*, 2017). Interestingly, a role for *lrhA* was identified as a result of the enrichment of transposon insertions oriented within an intergenic region (IGR) upstream of *lrhA*, suggesting that increased transcription of the gene significantly impacted capsule production (Goh *et al.*, 2017). In addition, TraDIS was used to study motility of the fluoroquinolone-resistant sequence type, ST131, an emerging group of multidrug-resistant *E. coli* associated with disease, demonstrating that mutations in 30 genes induced hypermotility, including 8 IGRs (Kakkanat *et al.*, 2017). TraDIS has also been used to study GI tract colonisation, invasion, and systemic survival of *E. coli* K1 in rat pups (McCarthy *et al.*, 2018). This study identified 167 gene products required for GI tract colonisation, and 97 genes required for survival in human serum (McCarthy *et al.*, 2018). Finally, a TraDIS-based approach was retrospectively applied in a study of O157:H7 colonisation of the bovine GI tract (Eckert *et al.*, 2011). In this work, TraDIS was applied to a relatively small mutant pool of 1,805 mutants, and, when comparing their results to a previous signature-tagged mutagenesis screen (using the same mutant library), they were able to improve the assignment of fitness scores from 4.4% of the mutants analysed to 91.1%, including the identification of 41 additional attenuating insertions in the locus of enterocyte effacement (LEE) (Eckert *et al.*, 2011). Furthermore, no further animal use was required in this study - TraDIS being applicable to use on

preserved cultures from the initial colonisation screen - which represents a significant cost and material benefit to *in vivo* screens.

Finally, as previously mentioned, TraDIS can offer an extremely high depth of coverage, allowing the determination of essentiality to the sub-gene or protein domain level. An excellent example of this has been demonstrated recently by Goodall *et al.*, which is discussed later in this thesis (see Chapter 3) (Goodall *et al.*, 2018).

Study	Reference
E. coli K1 *in vitro* growth, GI tract colonisation, and survival in serum	(McCarthy *et al.*, 2018)
Capsule production in K1 UPEC strain PA45B	(Goh *et al.*, 2017)
Motility in *E. coli* ST131	(Kakkanat *et al.*, 2017)
The essential genome of *E. coli* K-12	(Goodall *et al.*, 2018)
Serum resistance in *E. coli* ST131	(Phan *et al.*, 2013)
Susceptibility to T4 and T7 phage in O157:H7	(Cowley *et al.*, 2018)
Retrospective study of O157:H7 screened in cattle	(Eckert *et al.*, 2011)
S. Typhi adaptation to survival in water	(Kingsley *et al.*, 2018)
Twitching motility-mediated biofilm formation in *Pseudomonas aeruginosa*	(Nolan *et al.*, 2018)
The essential gene set of *Yersinia pseudotuberculosis* IP32953	(Willcocks *et al.*, 2018)

Table 1.2. TraDIS studies of *E. coli* and other Enterobacteriaceae.

1.4. Objectives of this study

The overall objective of this work is to use TraDIS to characterise the genetic requirements for growth under conditions relevant to *E. coli* colonisation of the GI tract. Some aspects of *E. coli* colonisation remain poorly understood, therefore TraDIS should offer comprehensive and novel insights into colonisation mechanisms. To achieve this, a transposon mutant library of MG1655 was firstly subjected to analysis by TraDIS in order to understand the genetic requirements for growth in LB medium. The mutant library was then subject to growth in the presence of bile to allow a TraDIS-based analysis of the genetic requirements underpinning bile resistance in *E. coli*. Similarly, TraDIS was applied to the mutant library following anaerobic growth in the presence of the alternative terminal electron acceptor, nitrate, to understand global genetic requirements during anaerobic nitrate respiration.

Chapter 2 Materials and Methods

2.1. Strains and growth conditions

Strains and oligonucleotides used in this study are listed in Table 2.1 and Table 2.2. *E. coli* were routinely cultured in Lysogeny Broth (LB) (5 g/L Yeast Extract (Merck), 10 g/L NaCl (Sigma), 10 g/L Tryptone (Merck)). *E. coli* were cultured under anaerobic conditions in M9 minimal medium (33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 5 g/L NH_4Cl, 2.5 g/L NaCl, 1 g/L $MgSO_4$ 0.4% (w/v) glucose, 10 µg/ml thiamine and 25 µg/ml uridine adjusted to pH 7.4). For solid media, 1.5% (w/v) agar (Merck) was added. Bacteria were plated onto solid media, grown overnight at 37°C, and stored at 4°C until needed. Overnight cultures were inoculated with a single colony from solid agar and grown overnight in LB shaking at 37°C. Antibiotics were included in solid and liquid medium where appropriate at the following concentrations: Kanamycin (Sigma-Aldrich) 50 µg/ml, Chloramphenicol (Sigma-Aldrich) 20 µg/ml, Ampicillin (Sigma-Aldrich) 100 µg/ml.

Strain	Characteristics	Source
E. coli K-12 *substr*. MG1655	F-, λ-, *ilvG*-, *rfb*-50, *rph*-1	Prof. Ian Henderson, University of Birmingham, UK
E. coli K-12 *substr*. BW25113	F^-, DE(*araD*-*araB*)567, *lacZ*4787(del)::*rrnB*-3, LAM^-, *rph*-1, DE(*rhaD*-*rhaB*)568, *hsdR*514	*E. coli* genetic stock centre (CGSC), Yale, USA.

Table 2.1. Strains used in this study.

Primer	Sequence (5' - 3')	Tm (°C)	Description
SplA5_top	G*AGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T	N/A	Splinkerette adaptor - top strand sequence * indicates a phosphorothioate group
SplA5_bottom	/5Phos/G*ATCGGAAGAGCGGTTCAGCAGGttttttttttcaaaaaaa*a	N/A	Splinkerette adaptor - bottom strand sequence. * indicates a phosphorothioate group
SplAP5.1	CAAGCAGAAGACGGCATACGAGATAACGTGATGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence ATCACGTTAT
SplAP5.2	CAAGCAGAAGACGGCATACGAGATAAACATCGGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence CGATGTTTAT
SplAP5.3	CAAGCAGAAGACGGCATACGAGATATGCCTAAGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence TTAGGCATAT
SplAP5.4	CAAGCAGAAGACGGCATACGAGATAGTGGTCAGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence TGACCACTAT
SplAP5.5	CAAGCAGAAGACGGCATACGAGATACCACTGTGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence ACAGTGGTAT
SplAP5.6	CAAGCAGAAGACGGCATACGAGATACATTGGCGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence GCCAATGTAT
SplAP5.7	CAAGCAGAAGACGGCATACGAGATCAGATCTGGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence CAGATCTGAT
SplAP5.8	CAAGCAGAAGACGGCATACGAGATCATCAAGTGAGATCGGTCTCGGCATTCC	65	Splinkerette-specific primer with bardcoding sequence ACTTGATGAT
qPCR2.1	AATGATACGGCGACCACCGAG	70	P7-specific primer for library quantification
qPCR2.2	CAAGCAGAAGACGGCATACGA	67	P5-specific primer for library quantification
Ez-Tn5	AATGATACGGCGACCACCGAGATCTACACATGATGATATATTTTTATCTTGTGC AATGTAACATCAGAG	>75.0	Transposon-specific sequencing primer
iPCRtagSeq	AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC	>75.0	Index read sequencing primer
Illumina_Read_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	N/A	Illumina Read 1 sequencing primer

Table 2.2. Primers and other oligonucleotides used in this study.

2.2. Transposon mutant library construction

The MG1655 transposon mutant library used in this work was originally constructed by Dr. Emma Smith in the School of Microbiology, UCC.

2.2.1. Preparation of electrocompetent E. coli

Electrocompetent *E. coli* were prepared to allow for transformation with Tn5 transposomes (Epicentre). A single colony of *E. coli* K-12 MG1655 was suspended in 5 ml of LB and grown overnight at 37°C. 200 µl of overnight culture was diluted into 100 ml fresh LB and grown, shaking, at 37°C, until the culture reached an OD_{600} = 0.3 - 0.5. The culture was cooled on ice for 30 minutes, before being centrifuged at 3000 x *g* for 10 minutes at 4°C. The supernatant was discarded and pellets were resuspended in 100 ml of sterile, ice-cold, deionized water (dH_2O), followed by centrifugation. This step was repeated once more, resulting in two wash steps. The cell pellet remaining following the second centrifugation was resuspended in 3 ml sterile ice-cold dH_2O, centrifuged again, and then resuspended in 180 µl sterile, ice-cold, 20% (v/v) glycerol. This mixture was aliquoted into 50 µl volumes and used immediately or stored at -80°C for later use.

2.2.2. Transposon mutagenesis

Aliquots of electrocompetent MG1655 were transformed with the EZ-Tn5™ <KAN-2>Tnp Transposome™ (Epicentre) by electroporation, following EZ-Tn5™ kit instructions. Briefly, 1 µl volumes of transposome (containing the transposon DNA fragment and transposase enzyme) were

incubated with electrocompetent cells on ice for 1 min. The mixture was transferred to a 50 x 2 mm universal fit eletroporation cuvette (Cell Projects) and then subject to electroporation. Cells were immediately resuspended in 950 µl super optimal broth with catabolite repression (SOC; 20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) before incubating at 37°C, shaking for 1 h. Recovered cells were plated onto QTrays (Genetix) containing LB kanamycin agar and incubated for 24 h at 37°C. Colonies were counted, pooled, and resuspended into sterile LB 10% (w/v) glycerol before storage at -80°C. Transposon mutagenesis was conducted six times, creating six independent pools of transposon mutants (Table 2.3)

Transposon mutant pool	**Numer of colonies (approx.)**
1	179,900
2	175,500
3	240,000
4	90,600
5	101,000
6	132,000
Total	919,000

Table 2.3. Numbers of kanamycin-resistant colonies within each mutant pool.

2.2.3. Base library

Aliquots of all 6 mutant pools were pooled together to generate a 'base library' composed of approx. 919,000 mutants (see Table 2.3). Mutant pools were resuspended to an OD_{600} of 1 in sterile LB broth with 10% (v/v) glycerol and aliquoted into 1 ml volumes. The 1 ml aliquots were pooled together, mixed, and aliquoted into further 1 ml volumes before storage at -80°C.

2.3. Transposon library screening

2.3.1. Assay for growth in the presence of bile

To investigate genetic requirements for growth in the presence of bile, the base library was screened in LB broth supplemented with 2% (w/v) and 10% (w/v) ox bile as described previously (Langridge *et al.*, 2009). A base library aliquot was grown in 500 ml ml sterile LB kanamycin broth overnight, shaking at 37°C. gDNA was isolated from 5 ml of culture using a Genomic-tip 100/G kit, and was thereafter denoted as the INPUT library. An equivalent of 3 x 10^7 cfu from the input culture was transferred to 1 ml sterile LB broth containing 0.02% (w/v) Ox-bile (Sigma-Aldrich) and incubated at 37°C for 50 minutes. The full volume of culture was then transferred to 50 ml sterile LB broth containing 2% (w/v) Ox-bile and incubated overnight, shaking at 37°C. 1 ml of this culture was transferred to 50 ml LB containing 10% (w/v) Ox-bile and incubated for 24h, shaking at 37°C. This culture was thereafter denoted as the OUTPUT library. gDNA was isolated from 5 ml of this library using a Genomic-tip 100/G kit.

2.3.2. Assay for anaerobic growth in the presence of nitrate

100 µl aliquots (duplicates) of the base library was inoculated into 10 ml of sterile M9 broth and grown overnight, shaking aerobically, at 37°C. This was denoted the INPUT library and DNA was isolated as described. Cells were pelleted by centrifugation at 3000 x *g* for 10 minutes and then resuspended to an OD_{600} = 1 in sterile M9 salts solution (M9 salts in water only). 8 universal tubes were filled to the brim (approximately 33ml) with sterile M9 kanamycin broth, 4 cultures were each supplemented with 0.4% (w/v) glucose or 0.4% (w/v) glycerol. At this point, 40mM sodium nitrate was added to 2 tubes containing glucose or glycerol, in total giving 2 tubes with M9 and glucose only, 2 tubes with M9, glucose, and nitrate, and 2 tubes with M9, glycerol, and nitrate. In an anaerobic chamber, each universal tube was inoculated with input library culture to an OD_{600} = 0.05. Tubes were mixed gently and approximately 2 ml of liquid was removed from each tube (to create a headspace to allow cells to mix) before sealing tightly. Universal tubes were incubated at 37°C, shaking (to keep the *E. coli* cells suspended and cultures homogenous), for 24h, in an anaerobic jar. Culturing in universal tubes was repeated, using culture from the previous passage as inoculum (inoculum level at 1%) for a total of three passages. The third passage of each condition was denoted the OUTPUT snd gDNA was isolated from input and outputs cultures using a Genomic-tip 100/G.

2.4. TraDIS

I acknowledge and thank Francesca Short, Christine Boinett, and Amy Cain of the the Wellcome Trust Sanger Institute for carrying out the initial TraDIS sequencing and some preliminary analyses of the base library and mutant library following growth under anaerobic conditions (as per section 2.3.3), and Fiona Crispie and Laura Finnegan of Teagasc Moorepark for TraDIS sequencing of the mutant library following growth in bile (section 2.3.2). All analyses of TraDIS sequencing data discussed in this thesis were carried out by myself.

Standard TraDIS analysis was conducted as per previously established protocols (Barquist *et al.*, 2016). In order to carry out TraDIS, gDNA was used to generate sequencing libraries suitable for Illumina® sequencing technology (see Figure 2.1).

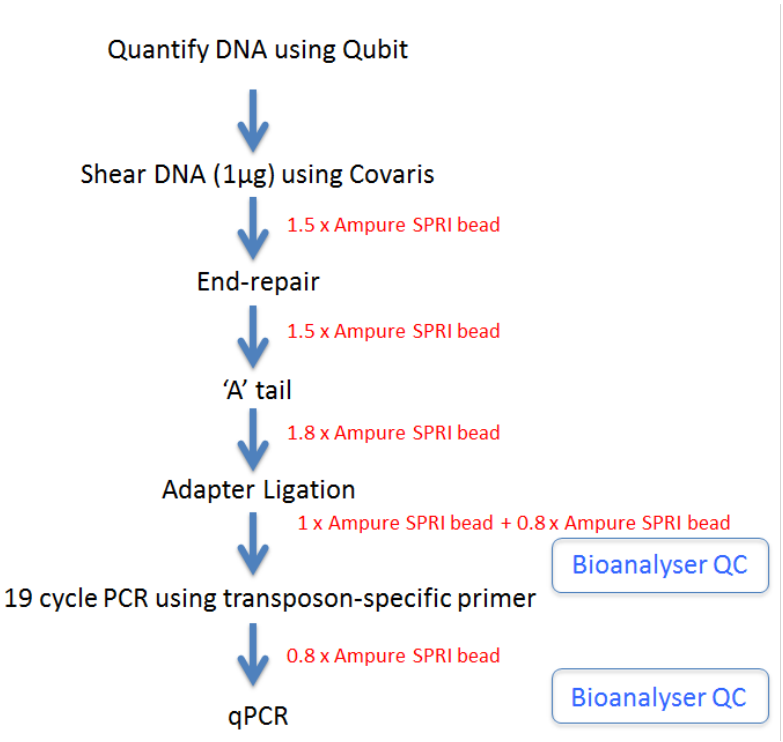


Figure 2.1. TraDIS sequencing library preparation workflow. From Barquist *et al.* 2016.

Briefly, this involved physical shearing gDNA into uniform fragments, then end-repairing and dA-tailing those fragments in order to allow the attachment of specialised adaptors (see Figure A1.2). These adaptors facilitated the amplification specifically of transposon-containing DNA fragments by PCR. qPCR was conducted on the PCR-amplified library in order to quantify the amount of transposon-containing DNA within the library, before sequencing the library following standard protocols for Illumina® sequencing.

Library preparation was conducted using a custom protocol (courtesy of Francesca Short) adapted for use with the NEBNext® DNA Library Prep Reagent Set for Illumina® (NEB; see Figure 2.1). Appropriate working guidelines were followed throughout the protocol to ensure minimal contamination and sample loss, including use of low-bind tubes, separation of pre-PCR and post-PCR preparation areas and reagents, and use of filtered micropipette tips.

Transposon sequencing data was processed, mapped, and analysed using the Bio::TraDIS pipeline of command-line software utilities (Barquist *et al.*, 2016). This analysis generated lists of genes and statistical analyses as a measure of the numbers of transposon mutants within each TraDIS library.

2.4.1. DNA preparation and shearing

gDNA was quantified and subject to a quality check prior to TraDIS library construction using a Qubit fluorometer (Invitrogen). At least 1 µg high quality DNA was used for each library preparation. DNA was made up into 120 µl with elution buffer (EB; 10 mM Tris-Cl, pH 8.5) in a sterile microcentrifuge tube and mixed thoroughly. Resuspended DNA was transferred to a fresh, labeled microTUBE (Covaris; contains an AFA filament), and inserted into the holder of an M220 Focused-ultrasonicator (Covaris). DNA was sheared into 250 - 300 bp fragments using a with the following parameters: Peak Power 140; Duty Factor 10%; Cycles Per

Burst 200; Time 80 s. When sheared, samples were stored in tubes or transferred to a sterile, labeled microcentrifuge tube.

2.4.2. AMPure XP cleanup

Sheared DNA was purified using AMPure XP beads (Beckman Coulter). Prior to purification, magnetic beads were mixed and allowed to acclimate to room temperature for 30 mins. In a fume hood, 180 µl AMPure XP beads were added to 120 µl sheared DNA and mixed fully by pipetting up and down 10 times. The mixture was incubated at room temperature for 5 mins, before placing onto a magnetic rack for approximately 4 mins until the supernatant cleared. The supernatant was removed and transferred to a clearly labeled microcentrifuge tube, which was retained for quality control checks. 300 µl 80% ethanol was added to the tube containing the magnetic beads, ensuring the beads were not disturbed, and left for 30 s. The ethanol was removed and discarded, again ensuring that the beads were not disturbed. Ethanol was added and removed in this manner again for a total of 2 washes. The tube containing the magnetic beads was centrifuged briefly to collect any remaining ethanol at the bottom of the tubes, which was carefully removed by pipetting. Beads were air-dried for 2 - 3 mins until the magnetic bead pellet had the appearance of wet paint. Once dry, tubes were removed from the magnet and 52 µl EB was added to resuspend the magnetic beads. Tubes were vortexed thoroughly to resuspend the magnetic beads and then incubated for 5 mins at room temperature. Tubes were placed on a magnetic rack for 2 -

3 minutes until the supernatant was clear and 50 µl of this cleared solution was transferred to a fresh, labeled microcentrifuge tube.

2.4.3. End-repair, dA-tailing, adaptor ligation

In order to sequence DNA, specialised sequencing adaptors have to be ligated to either end of the purified fragments. This required an initial repair of any overhanging breaks using T4 polynucleotide kinase and the DNA polymerase I Large (Klenow) fragment (included in the NEBNext® DNA Library Prep Reagent Set), followed by dA-tailing of the end-repaired DNA to allow the adaptor to anneal to the DNA fragments.

2.4.3.1 End-repair

The following components were mixed in a sterile 0.2 ml PCR tube (all components indicated apart from the sheared DNA are derived from the NEBNext® DNA Library Prep Reagent Set): purified sheared DNA (75 µl); phosphorylation reaction buffer (10 µl); T4 DNA polymerase (5 µl); T4 polynucleotide kinase (5 µl); dNTPs (4 µl); DNA polymerase I, Large (Klenow; 1 µl); sterile dH_2O (to 100 µl). The mixture was incubated in a thermal cycler for 30 mins at 20°C and purified using 160 µl AMPure XP beads. The mixture was eluted in 32 µl EB into a fresh, sterile PCR tube.

2.4.3.2. dA-tailing

End-repaired DNA was dA-tailed by adding the following components into a sterile PCR tube (all components indicated apart from the end-repaired DNA are derived from the NEBNext® DNA Library Prep Reagent Set):

end-repaired, blunt DNA (32 µl); NEBuffer 2, 10X (5 µl); deoxyadenosine 5'-triphosphate (10 µl); Klenow fragment (3' → 5' exo-; 3 µl). The mixture was incubated in a thermal cycler for 30 mins at 37°C and purified using 90 µl AMPure XP beads. The mixture was eluted into 10 µl EB in a fresh, sterile PCR tube.

2.4.3.3. Adaptor ligation

Specialised 'Splinkerette' indexed adaptors (see Figure 2.2) were ligated to dA-tailed DNA by adding the following components into a sterile PCR tube (all components apart from the dA-tailed DNA and Spl5 Adaptor are derived from the NEBNext® DNA Library Prep Reagent Set): dA-tailed DNA (10 µl); Quick Ligation reaction buffer, 2X (25 µl); Spl5 Adaptor, 10 µM (10 µl); Quick T4 DNA ligase 5 µl). The mixture was incubated in a thermal cycler for 15 mins at 20°C. 3 µl of USER Enzyme Mix was added and mixed by pipetting up and down, before incubating again in a thermal cycler at 37°C for 15 mins. The mixture was purified using 90 µl AMPure XP beads and eluted in 100 µl EB into a fresh, sterile PCR tube. To verify successful ligation of adaptors, a 1:10 dilution of purified ligated DNA was made in EB and 1 µl was added to an Aglient High Sensitivity Chip (Aglient), to be analysed in an Agilent 2100 Bioanalyzer (Aglient). Successful ligation of adaptors was verified by the production of a characteristic molecular weight profile (Figure 2.3).

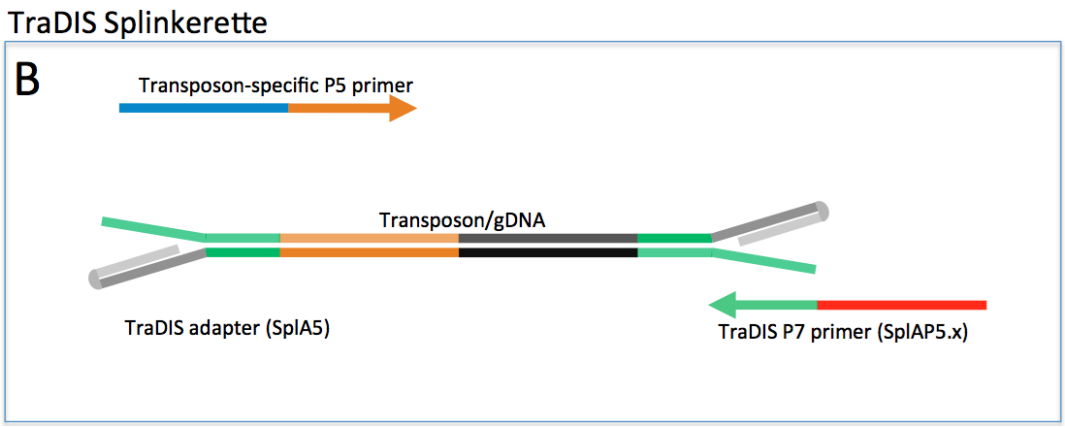


Figure 2.2. TraDIS splinkerette adaptors. Splinkerettes (green DNA strands with grey loops) are attached to either end of the transposon/gDNA fragment. Amplification of both strands can only occur if the transposon-specific primer (indicated) hybridises to the transposon-specific sequence first. Adapted from Barquist *et al*., 2016.

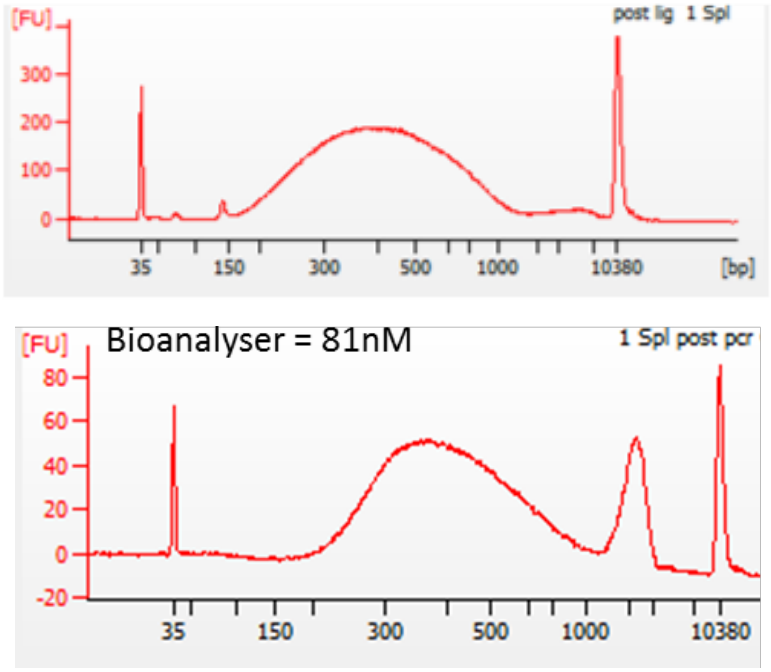


Figure 2.3. Typical High Sensitivity Agilent Bioanalyzer traces of DNA fragments (1:10 dilution) post-ligation (top graph) and post-PCR (bottom graph) using TraDIS adaptors and primers. Adapted from Barquist *et al*. 2016.

2.4.4. Bead-based size selection and PCR enrichment of adaptor-ligated DNA

2.4.4.1. Bead-based size selection

Adaptor-ligated DNA was selected by bead-based size selection. This involved following the AMPure XP protocol to select for fragments of size 370 bp, approximately the same size as DNA fragments (250-300bp) with ligated adaptors (43 bp each). The protocol described in section 2.4.2. was followed, except starting with 100 µl adaptor-ligated DNA and 70 µl AMPure XP beads and eluting into 15 µl Tris-HCl or Tris-EDTA (TE). All supernatants were retained for quality-control analysis.

2.4.4.2. Enrichment of adaptor-ligated DNA by PCR

The following components were added into a sterile PCR tube: DNA (15 µl); sterile H_2O (5 µl); KAPA Biosystems HiFi HotStart ReadyMix, 2X (25 µl); 5' transposon-specific primer (0.5 µl); SPIAP5.x barcoded primer (see Table 2.2), 10 µM (5 µl). The mixture was incubated in a thermal cycler with the following parameters: 95°C 3 mins; 19 cycles of 98°C (20 s), 65°C (30 s), 72°C (30 s); 72°C 5 mins; hold at 4°C. The mixture was purified twice using AMPure XP beads, first with 40 µl beads and eluting into 50 µl EB, then with 40 µl beads and eluting into 30 µl EB. To confirm PCR purification of adaptor-ligated DNA, 1 µl of a 1:10 dilution of purified PCR product was run on a Bioanalyzer as described above (see Figure A1.2 for an example of a post-PCR Bioanalyzer profile).

2.4.5. qPCR

The concentration of transposon-containing DNA fragments in each library was measured by quantitative PCR (qPCR). qPCR was conducted using the KAPA SYBR® FAST qPCR Master Mix (2X) Kit (KAPA biosystems). Two qPCR reactions were conducted per library: A. using Illumina library quantification DNA standards (KAPA Biosystems) and primers qPCR2.1 and qPCR2.2 (see Table 2.2), which target the P7 and P5 regions of the ligated adaptors, respectively; B. using the P5-targeting primer (qPCR 2.2) and the transposon-sequencing primer (see Table 2.2). Reaction A. is used to quantify the total concentration of DNA in each library, while reaction B. is used to quantify the percentage of DNA containing transposon-gDNA junctions. The two reactions were set up as follows: for reaction A. the following were added to a sterile PCR tube: purified adaptor-ligated DNA, diluted 1:5000 (13 µl); KAPA qPCR master mix **with** Illumina standards (39 µl); sterile dH_2O (13 µl). For reaction B. the following were added to a sterile PCR tube: purified adaptor-ligated DNA, diluted 1:5000 (13 µl); KAPA qPCR master mix, **without** Illumina primers (32.5 µl); transposon-specific sequencing primer, 10 µM (1.3 µl); qPCR primer 2.2, 10 µM (1.3 µl); sterile dH_2O (16.9 µl). Each reaction mix was (total 65 µl) divided into three wells of a white Lightcycler® 480 96 well plate (Roche) such that 20 µl was in each well. This way reactions were conducted in triplicate. The plate was then incubated in a Lightcycler® 480 Instrument (Roche) under the following conditions: 95°C (5 min), and 35 cycles of 95°C (30 s) and 65°C (45 s). The concentration of the undiluted libraries was then calculated using the quantification

template spreadsheet, available from KAPA at https://www.kapabiosystems.com/document/kapa-library-quantification-data-analysis-template/?dl=1 . The transposon-specific product typically made up to between 50% and 80% of the total DNA concentration. Libraries can be sequenced at a defined set of concentrations (4 nM, 2 nM, 1 nM, and 0.5 nM) and libraries were diluted to the nearest appropriate concentration in EB in sterile low-bind tubes prior to sequencing.

2.4.6. Illumina sequencing

The TraDIS library was sequenced on a MiSeq™ (Illumina) sequencer. The library was denatured and loaded according to standard Illumina protocols. Briefly, 4 µl of 100 µM transposon-specific sequencing primer and 4 µl 100 µM Illumina Read 1 sequencing primer (see Table 2.2) were added to 600 µl buffer HT1 and added to port 19 of the MiSeq reagent cartridge. 4 µl of 100 µM Index read primer (iPCRtagseq; see Table 2.2) was added to 600 µl HT1 buffer and this mixture was added to port 19.

Sequencing was conducted using a custom TraDIS recipe as outlined in the sample sheet (Table 2.4). The sample sheet provided the MiSeq™ machine instructions for setup, performance, and analysis of the sequencing run. This recipe incorporates 'dark' cycles, whereby 10 sequencing cycles are conducted with no imaging, which optimises sequencing across the monotemplate sequence of the transposon during Read 1 ((Barquist *et al.*, 2016); Figure 2.4). Following this, the transposon

sequence was generated as a separate 10 bp sequence following Read 1 and before the Index read.

[Header]					
Investigator Name		Tradis			
Project Name					
Experiment Name		Tradis_Rerun_May2018			
Date					
Workflow		GenerateFASTQ			
Chemistry		Transposon10			
[Reads]					
	42				
[Settings]					
[Manifests]					
[Data]					
Sample_ID		Sample_Name	GenomeFolder	Index	Index2
	1			TAAGAGACAG	CGATGTTTAT
	2			TAAGAGACAG	TGACCACTAT
	3			TAAGAGACAG	GCCAATGTAT
	4			TAAGAGACAG	ACTTGATGAT

Table 2.4. Layout of a typical sample sheet used in this study.

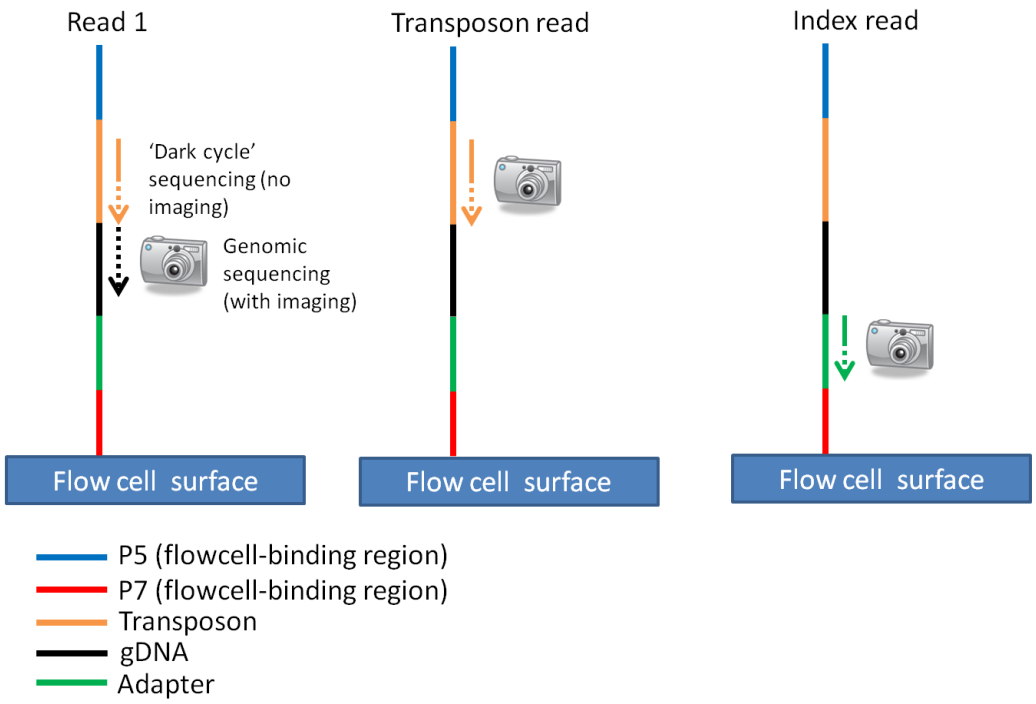


Figure 2.4. HiSeq and MiSeq TraDIS recipes allow for 'dark' sequencing across the difficult monotemplate sequence of the transposon. The transposon-specific sequencing primer hybridises to the known transposon sequence 10 bp upstream of the junction with gDNA. Sequencing takes place with no imaging for 10 or 12 cycles (10 or 12 for MiSeq, 12 for Hiseq2500) and continues with imaging for 42 cycles. The transposon sequence is generated as a separate 10bp (MiSeq) or 12bp (HiSeq2500) read following read 1 and before the index read. From Barquist *et al*. 2016.

2.5. TraDIS data analysis

2.5.1. Bio::TraDIS pipeline

The Bio::TraDIS pipeline of command-line software utilities was employed to process, map, and analyse transposon sequencing data. A detailed protocol is available in Barquist *et al.* 2016, however some modifications

were employed for this study. Sequencing data was mapped to the *E. coli* K.12 *substr.* MG1655 genome sequence .embl file, GenBank accession no. U00096.3 (Benson *et al.*, 2013).

2.5.1.1. Appending transposon reads to Read 1

Following some sequencing runs, transposon reads were generated in the header of the .fastq file as opposed to the sequence, which prevented use of the `add_tradis_tags` script at the beginning of the pipeline (Figure 2.5).

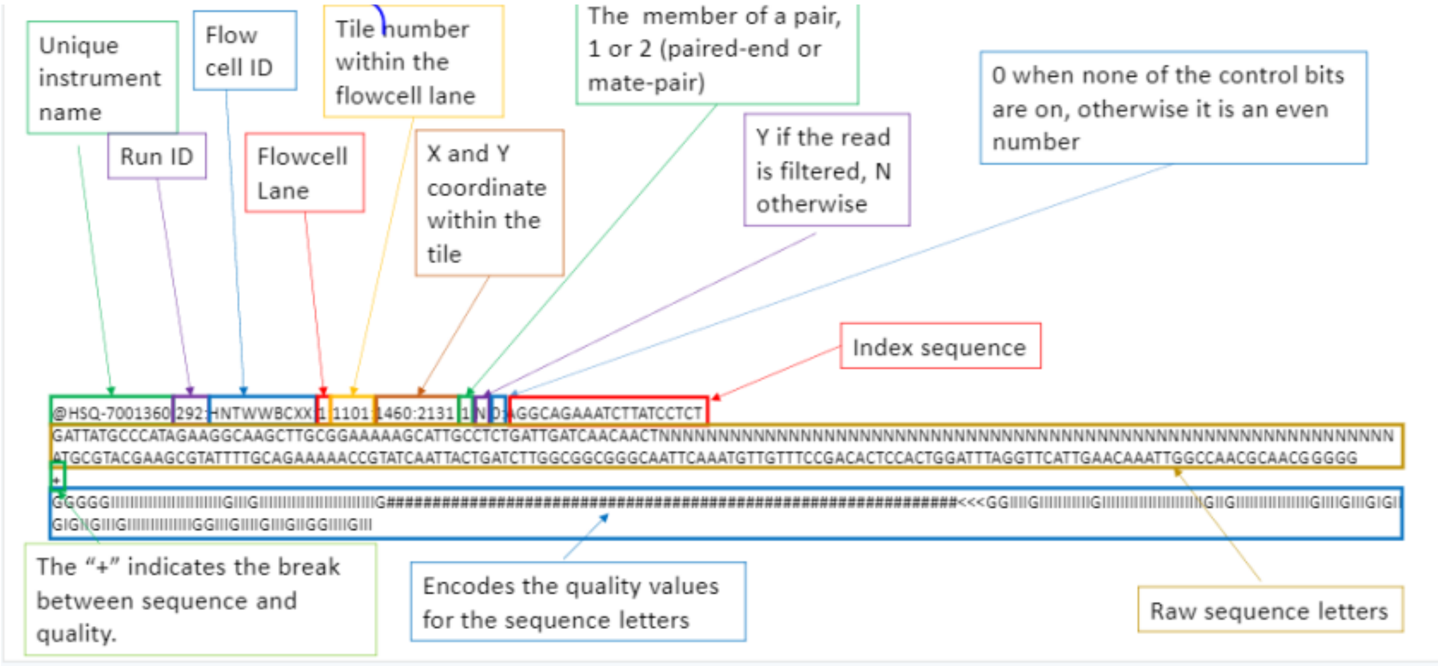


Figure 2.5. The FASTQ file format. In the Illumina FASTQ format, transposon indices were often parsed into the Index sequence section of the header (red box) instead of the raw sequence (blue box). Adapted from https://kscbioinformatics.wordpress.com/2017/02/03/raw-illumina-sequence-data-files-for-dummies-part-1/ Accessed: 30 Jan 2019.

To output the different reads (i.e. Read 1, Transposon Read, and Index Read - see Figure 2.4) into separate .fastq files, the bcl2fastq Conversion software v2.19 (Illumina) was run on raw .bcl files obtained from the sequencing run under the following parameters:

```
nohup /usr/local/bin/bcl2fastq --runfolder-dir <RunFolder> --output-dir <BaseCalls> --sample-sheet /path/to/SampleSheet.csv --create-fastq-for-index-reads
```

In Galaxy v2.0.1.1 (https://usegalaxy.org/), the FASTQ joiner program (Blankenberg *et al.*, 2010) was used to append Index 1 (i.e. transposon) reads to Read 1 (i.e. chromosomal reads). Index 1 read files were entered into 'Left-hand Reads' and Read 1 read files into 'Right-hand Reads'. FASTQ Header Style was set to 'new' and no bases were entered between reads. The output .fastq files were gzipped and then brought forward for analysis by the Bio::TraDIS pipeline as outlined in the software protocols (Barquist *et al.*, 2016).

2.5.1.2. Assessing gene essentiality

Essentiality and/or fitness data generated by `tradis_essentiality.R` and `tradis_compaison.R`. in the form of .csv files were analysed in Microsoft Excel (2011). Statistical thresholds were intrinsically applied by `tradis_essentiality.R`, therefore no further statistical analysis was

conducted on essential genes identified by this script. However, for fitness values (logFCs) generated by `tradis_compaison.R` , a p value cutoff of p < 0.05 was applied. q value thresholds were applied such that no more than 1 gene in the list of genes with a p value < 0.05 could be considered a false positive. This q value threshold was unique to each experiment. To verify essentiality and/or fitness values, genetic requirements for fitness were also assessed manually using transposon insertion plots. In Artemis (Carver *et al.*, 2012), gzipped plot files (e.g. '[library_name.replicon.name].insert_site_plot.gz) generated by the `bacteria_tradis` script were overlaid on the U00096.3 sequence using the 'Graph -> Add User Plot...' option. This generated a viewable version of transposon insert sites and read densities.

2.5.2 Gene set analysis

Lists of genes within statistical thresholds were grouped together by gene set analysis to infer biological meaning from the data. In this work, genes were primarily grouped by Clusters of Orthologous Groups (COGs), Gene Ontology (GO), or manually ((Kristensen *et al.*, 2010), http://geneontology.org/). To group genes by COG, the gene list was entered into Ecogene's cross reference mapping and download tool (http://www.ecogene.org/ecodownload/crossref; (Zhou & Rudd, 2012)), selecting the 'oldCOG' option to output a list of genes with their associated COG values. Genes were grouped by GO using the Panther Classification System (http://www.pantherdb.org/; (Mi *et al.*, 2017)). Manual assessment was conducted based on information available in

EcoCyc (https://ecocyc.org/), EcoGene (http://ecogene.org/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed), and other online databases (Keseler *et al.*, 2017; The UniProt Consortium, 2017; Zhou & Rudd, 2012).

2.5.3. Data visualisation

A global overview of transposon insertion sites in a circular map of the MG1655 genome was generated by inputting transposon insertion counts for each gene into CiVi (http://civi.cmbi.ru.nl/; (Overmars *et al.*, 2015)). To measure correlation between transposon insertion sites and readcounts between library replicates, Microsoft Excel's RSQ function and the `ggplot` suite of tools in R were used to generate R^2 values and scatterplots, respectively (https://cran.r-project.org/web/packages/ggplot2/index.html).

2.6. Validation of TraDIS and other analyses

2.6.1. End-point growth analysis of Keio library mutants

To validate mutant phenotypes detected during TraDIS, single gene knockout mutants from the Keio collection were grown in triplicate under conditions mimicking those of the transposon library screen and their final OD values measured.

When analysing the growth of mutants in the presence of bile, collections of mutants were grown in well plates containing LB kanamycin supplemented with different concentrations of LB ox bile as per section

2.3.1. Final OD_{595nm} values of mutants following growth in LB 10% bile were measured on a Tecan GENios spectrophotometer. Changes in mutant growth were deemed statistically significant if the average mutant growth displayed a one standard deviation difference in value compared to the average growth of the plate as a whole.

When analysing the growth of mutants under anaerobic conditions, mutants and wild-type were grown in well plates containing M9 minimal medium supplemeted with different carbon sources and/or nitrate as per section 2.3.2. Cultures were grown under anaerobic conditions for one passage, and significant changes in mutant growth was determined by comparison to the wild-type. End-point growth was measured as previously described.

2.6.2. Competition assays

5 ml of overnight cultures (in triplicate) of mutant and wild type were centrifuged at 3,000 g for 10 minutes and cell pellets were resuspended in sterile PBS to an OD_{600} = 1. Equal volumes of wild type and mutant cells were added to 5 ml culture broth to an OD_{600} = 0.05 and incubated overnight at 37°C. Resuspended cells were serially diluted in PBS to a dilution of 10^{-7}, and 100 µl of dilutions 10^{-6}, 10^{-7}, and 10^{-8} were plated onto LB (wild type) and LB kanamycin (mutant) agar, before incubation overnight. These plates were used to enumerate the concentration of cells in the input cultures. Following incubation, cultures were serially diluted in PBS to dilution 10^{-7}, and 100 µl of dilutions 10^{-5}, 10^{-6}, and 10^{-7}

were plated onto LB agar (to enumerate mutant + wild type cell count), and LB kanamycin agar (to enumerate mutant cell count) and incubated overnight. These plates were used to enumerate total cell count in the output culture. The competitive index (CI) of mutants was calculated as follows: $w = (ln_o/ln_i[mutant])/ln_o/ln_i[wild\ type])$, where w = fitness, i = input cfu/ml, o = output cfu/ml.

2.6.3. Anaerobic growth curves

The growth patterns of wild-type MG1655 grown under anaerobic conditions were characterised from cultures grown for 24 h in 96 well plates containing M9 minimal medium supplemented with different carbon sources and/or nitrate as per section 2.3.2. Overnight cultures of MG1655 were washed once in PBS and resuspended to an OD_{600} of 1 prior to inoculation. Each washed overnight culture was inoculated into three wells containing 150 µl of M9 medium such that each well contained an estimated starting OD_{600} equivalent to 0.01. The 96 well plate was incubated at 37ºC for 24 h in a microplate spectrometer which was placed in an anaerobic fume hood. Cultures were grown under anaerobic conditions for 24 h and OD_{600} readings were taken every 15 minutes. Growth rates were determined based on the average slope of the curve for the duration of the exponential phase of growth. Total growth was calculated as the area under the curve.

2.6.4. Cell-free supernatant analysis

5 ml of transposon mutant library cultures grown under anaerobic conditions (see section 3.3.) were centrifuged at 12,000 x *g* for 2 mins and supernatants transferred to fresh sterile microcentrifuge tubes. These supernatants were subject to further analyses to characterise the culture conditions following library screening.

Cell-free supernatants were analysed for succinic acid, lactic acid, acetic acid, formic acid, and ethanol by high performance liquid chromatography (HPLC) with a refractive index detector (Agilent 1200 HPLC system). An Agilent HiPlex H 300 x 7.7 mm column was used with 0.01 N H_2SO_4 as the elution fluid, at a flow rate of 0.6 ml min^{-1}. 20 µl of each sample was injected for analysis and the temperature of the column was maintained at 65°C.

Production of nitrite in cultures grown in the presence of nitrate indicates the activity of nitrate reductases. Nitrites were deteced using the Griess method (Griess & Bemerkungen, 1879). Three drops of Griess reagent (equal volumes 1 mg/ml *N*-(1-Naphthyl)ethylenediamine and 10 mg/ml sulfalinic acid) were added to 100 µl of cell-free supernatants on a white porcelain tile. The presence of nitrites was confirmed by the emergence of a red pink colour.

2.6.5. Computer software

Growth curves, competition assays, and other data analysis was performed using Microsoft Excel (2011), Microsoft Powerpoint (2011), and GraphPad Prism (ver. 6). Statistical analyses were conducted using GraphPad Prism. Command line tools and R scripts were employed using Terminal (Apple Inc., ver. 2.5.3) and R for Mac OS X (ver. 3.5.3).

Chapter 3 TraDIS analysis of a pooled transposon mutant library of *E. coli* MG1655

Introduction

Transposon sequencing technologies, such as TraDIS, allow for powerful and in-depth functional genomic analysis, connecting mutant phenotypes to gene function. Transposon sequencing and other methods such as defined mutant library analysis require that the composition of the starting mutant library be known, since changes in phenotypes during library screening must be traced back to their associated mutants. In other words, knowing which mutants are present or absent in the mutant library prior to screening is necessary.

TraDIS, as well as other transposon sequencing methods, uses pooled transposon mutant libraries which offers several advantages (Barquist *et al.*, 2013). Firstly, it allows for rapid and simultaneous screening of entire mutant libraries. For instance, mutant pools ranging from 10,000 to millions can be screened at once, as opposed to the 3864 mutants of the Keio library, each of which must be screened separately (Baba *et al.*, 2006; Barquist *et al.*, 2013; Eckert *et al.*, 2011; Goodall *et al.*, 2018; Langridge *et al.*, 2009). Secondly, the means by which mutants are identified in a pooled library allows for a highly detailed measure of gene essentiality. Locations of transposon insertions are identified by sequencing of transposon-gDNA junctions, allowing transposons to be located to their precise insertion site on the genome. This allows for the identification of essentiality to the level of protein domain and non-protein-encoding regions (such as promoters) (Barquist *et al.*, 2013). Finally, characterising the relative numbers of mutants in a pooled library results

in quantitative measures of gene fitness. The numbers of reads mapping to transposon insertions in the sequencing output can represent the relative proportion of those mutants within the total mutant population. The change in both transposon insertion frequency and number of reads can be represented as a log fold-change (logFC) score, which allows for a measure of mutant fitness even for non-essential genes during growth under selective conditions (Barquist *et al.*, 2016).

However, the use of pooled transposon mutant libraries can also be associated with various potential disadvantages. Firstly, pooling mutants will select against mutants that have a general slow growth phenotype or mutants present at low titres in the initial population (Grenov & Gerdes, 2008). Secondly, downstream analysis of pooled libraries is more complicated than for arrayed mutant libraries. For example, mutants in arrayed or defined mutant libraries are more easily traced e.g. to a coordinate of a multi-well plate, as opposed to requiring sequencing to locate insertions. Thirdly, transposon mutagenesis can lead to an over- or under-estimation of essential genes. This can be due to multiple factors, including structural features within DNA inhibiting transposon insertions (e.g. regions of extreme structure, DNA binding proteins occluding regions of DNA, proximity of a gene to the replication terminus), or secondary effects of mutations in nonessential genes i.e. polarity effects (Goodall *et al.*, 2018; Grenov & Gerdes, 2008).

Part of the optimal strategy for understanding genetic requirements for growth under any one particular condition should therefore couple transposon sequencing with defined mutant library screening. Previous

studies of gene essentiality in *E. coli* support this hypothesis. For instance, the numbers of essential genes identified during construction of the Keio collection of mutants, where essentiality was determined based on growth of mutants on selective agar, was 300 (Baba *et al.*, 2006). On the other hand, while 302 genes are nominated as essential by the profiling of the *E. coli* chromosome (PEC) database (which collates essentiality data from single-gene essentiality studies and large-deletion mutants (Yamazaki *et al.*, 2008)), not all of the identified genes are shared with the essential genes as defined by the Keio collection (Goodall *et al.*, 2018). Furthermore, a recent comparison between these studies and a TraDIS-based approach of essentiality in *E. coli* K-12 showed that only 248 genes were shared between all 3 studies (Goodall *et al.*, 2018). However, the differences in gene lists produced by these 3 studies could be explained by differences in methodology, as opposed to differences in absolute physiological or metabolic requirement for those genes (Goodall *et al.*, 2018). Therefore, comparing different mutant library analyses allowed the building of a more robust consensus on the true nature of genetic requirements for growth in LB.

The aim of this chapter was to characterise a TraDIS mutant library constructed in *E. coli* MG1655, a domesticated strain of *E. coli* K-12, and to compare results from this TraDIS with previous studies of essential genes in *E. coli* (Goodall *et al.*, 2018). This was done to identify genes essential and/or important for growth in LB.

Results & Discussion

3.1. TraDIS

3.1.1. Sequencing results

A saturated transposon mutant library was constructed in MG1655 as described in Materials and Methods. Approximately 919,000 mutants were pooled as colonies from LB agar plates and prepared for TraDIS sequencing (as described in Materials and Methods). The transposon mutant library was constructed by Dr. Emma Smith and the TraDIS sequencing was carried out by the Wellcome Trust Sanger Institute. All subsequent analysis of the sequencing reads was carried out during this thesis. Sequencing statistics are listed in Table 3.1.

Total Reads	Reads Mapped	% Mapped	Unique Insertion Sites
1111496	906223	97.352	193495

Table 3.1. TraDIS sequencing statistics

A total of 1,111,496 reads were obtained from the MiSeq run, of which approximately 84% contained the Tn5 transposon tag, with 97% of these reads mapping to the MG1655 reference genome (Genbank accession number U00096.2). In total there were 193,495 unique insertion sites, representing a density of approximately 1 insertion per 24 bp. Therefore, this mutant library has a similar insertion density to other high resolution transposon mutant libraries (Barquist *et al.*, 2013). Analysis of the sequencing reads revealed that the insertions were spread evenly across

the genome (Figure 3.1) and in roughly equal proportions between positive and negative strands. For example, total insertions within genes (not including intragenic regions or other non-gene regions) on the positive and negative strands equaled 73,325 and 79,732 respectively.

An insertion index, defined as the number of insertions in a gene divided by gene length, was calculated for each gene. This number can act as a comparable measure of essentiality, since not all genes are of the same length. Insertions in the extreme 3' end of genes (equivalent to the final 10% of the gene) were excluded from analysis, since such mutations can still result in functional gene products (Barquist *et al.*, 2016). A histogram of insertion index values showed a bimodal distribution, indicating that both essential and non-essential genes could be clearly delineated based on insertion index values (Figure 3.2).

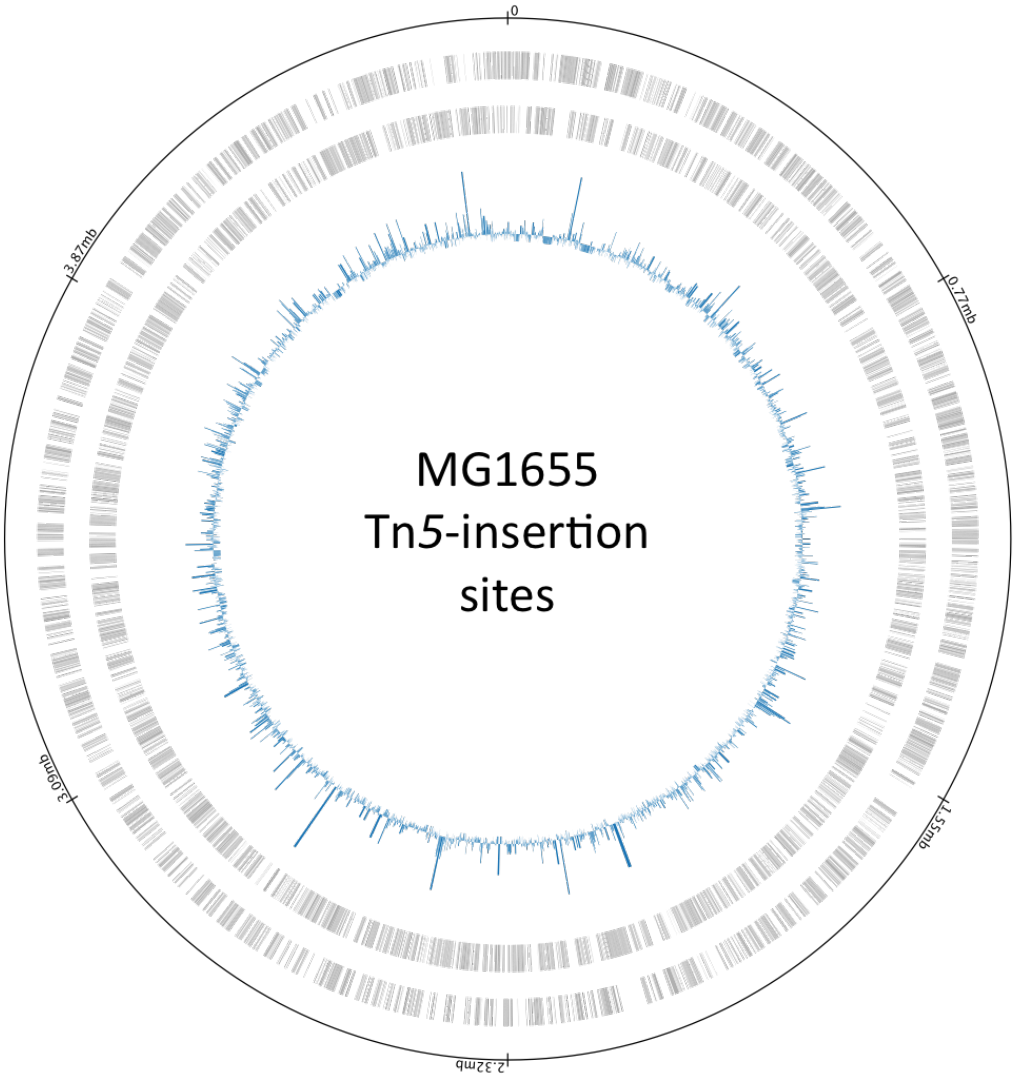


Figure 3.1. Transposon insertions mapped to the MG1655 genome (Genbank accession no. U000096). The outer ring represents base pair numbering. The next two rings of black bars represent genes transcribed from the positive and negative strands, respectively. The blue lines represent the number of transposon insertions within genes, the height of each bar representing transposon insertion density within each gene.

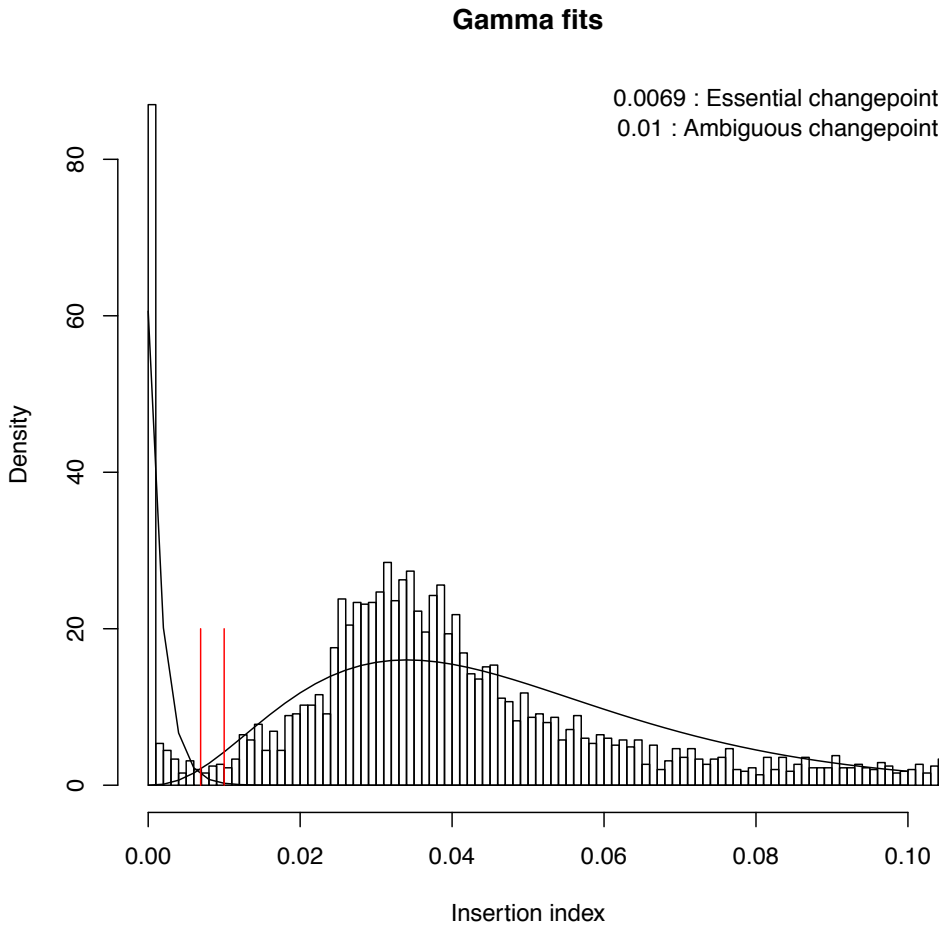


Figure 3.2. Bimodal distribution of insertion indices following TraDIS analysis of the saturated mutant library constructed in MG1655. Clear bars represent the number of genes of a particular insertion index (x-axis). Ambiguously assigned genes lie between the red lines, indicating insertion indices of 0.0069 (left-hand line) and 0.01 (right-hand line). Essential genes displayed insertion indices <0.0069 (to the left of the left-hand red line), while non-essential genes displayed insertion indices >0.01 (right of the right-hand red line).

An exponential distribution was fitted to the left node, used as an indication of whether a gene was essential, whereas a gamma distribution was fitted to the right node, used as an indication of whether a gene was non-essential (Figure 3.2). The probability of whether a gene belonged to each node was calculated, and the ratio of these values was used to calculate a log likelihood ratio. This showed that genes with an insertion index of less than 0.0069 were likely to be essential, an insertion index of greater than 0.01 indicated a non-essential gene, and an insertion index between these two values indicated an ambiguous gene (Figure 3.2). Determining whether a gene is essential should not rely on this technique alone, since scoring methods are hugely influential in determining essentiality (Grenov & Gerdes, 2008). Moreover, many genes considered essential did contain transposon insertions within their coding sequence (CDS), raising the prospect that these genes may be falsely annotated as essential. One technique used to refine the scoring technique is to treat DNA as one contiguous sequence, as opposed to a range of discrete sequences i.e. genes. Using a refined statistical model for TraDIS, Goodall *et al*. were able to define the minimum length of DNA that could be considered as essential in *E. coli* K-12, known as insertion-free regions, or IFRs (Goodall *et al.*, 2018). IFRs were calculated as being a minimum of 75 bp when considering the genome as a whole, or 47 bp within a gene. This offered both the ability to define essentiality within genes i.e. to the protein domain level (Figure 3.3), but also to non-coding regions. That said, manual assessment of transposon insertion plots formed the primary basis for determining essentiality of genes or

IFRs. Therefore, the previously calculated IFR values (Goodall *et al.*, 2018) in addition to manual inspection of transposon insertion plots, were considered when assessing gene essentiality as determined by insertion index. Applying these different thresholds and filters resulted in 428 genes being defined as essential in MG1655, with 32 given an ambiguous allocation (see Table A1, Appendix).

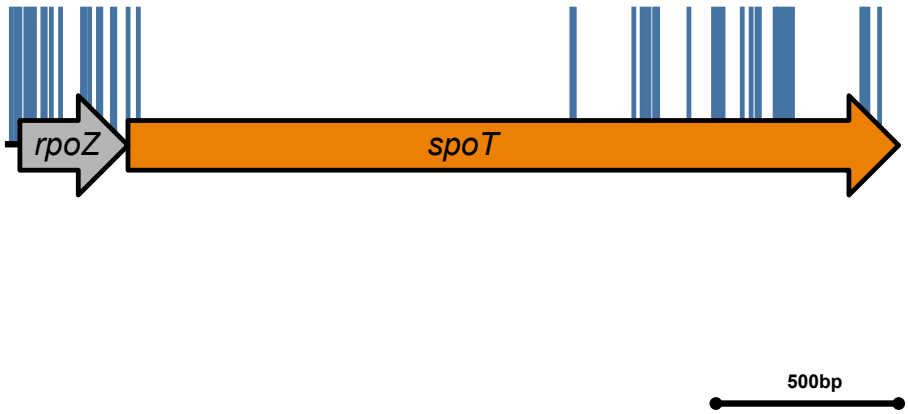


Figure 3.3. Transposon insertion plot demonstrating domain-level essentiality in the (p)ppGpp synthase/hydrolase-encoding *spoT* (orange). Transposon insertions (blue lines; not reflecting readcount differences) are primarily located in the 3' 'ACT' domain region, required for the amino acid concentration-dependent control of SpoT activity (Chipman & Shaanan, 2001).

3.1.2. Functional enrichment analysis

Clusters of orthologous groups (COGs) analysis was conducted to provide insight into the functions of the 428 essential genes (Tatusov *et al.*, 1997). This analysis showed that essential genes appeared to be

enriched in processes involved in the biosynthesis of integral cell components such as the cell membrane, the control of core cell processes like transcription, translation, and cell division, as well as the biosynthesis of metabolic enzymes and cofactors (Figure 3.4).

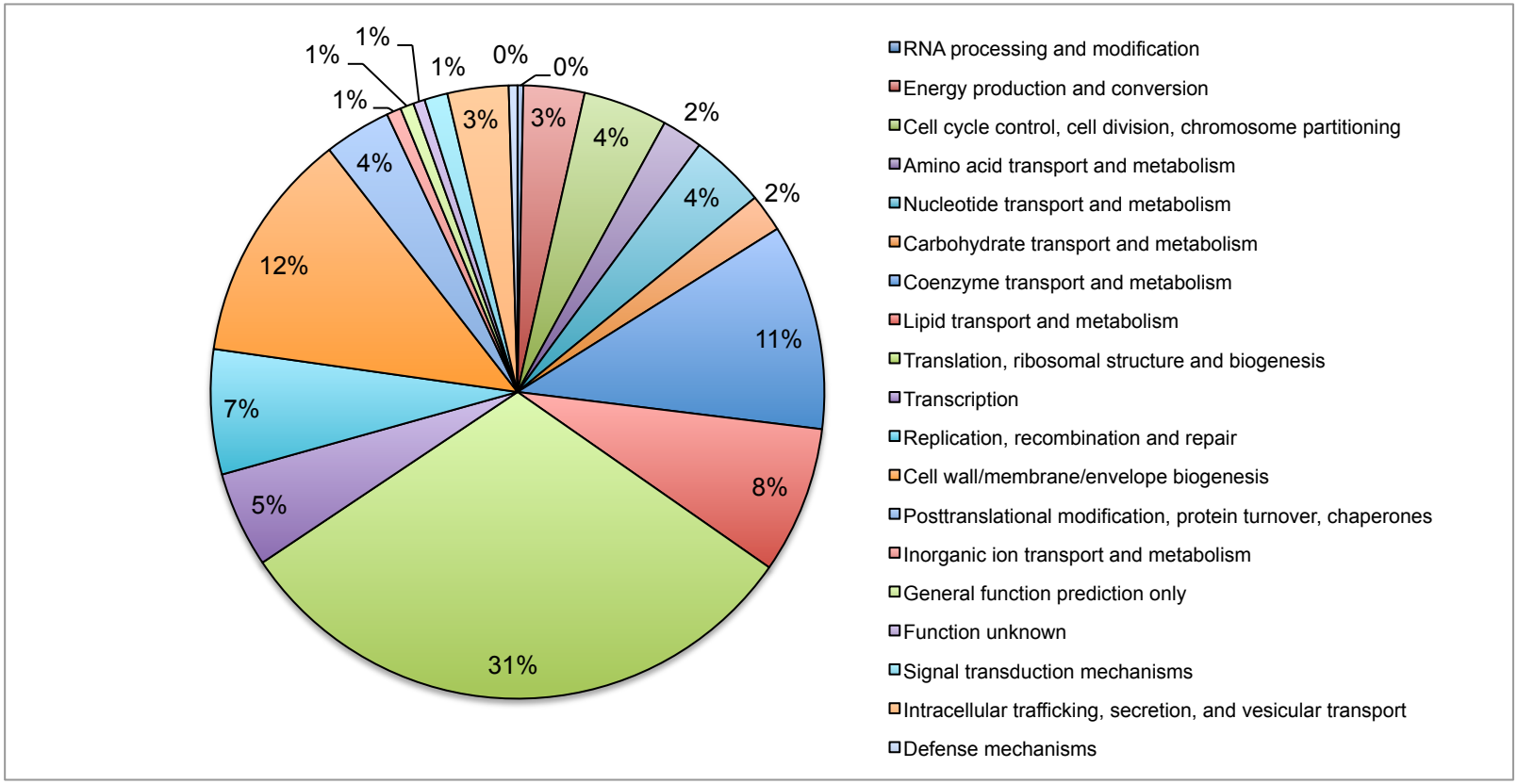


Figure 3.4. Significantly enriched COG biological process terms in essential genes.

3.2. A comparative analysis of the essential genome of E. coli K-12

An analysis of the essential genome of *E. coli* K-12 strain BW25113 using TraDIS following growth in LB was recently published (Goodall *et al.*, 2018). BW25113 and MG1655 are both derivatives of the W1485 lineage of *E. coli* K-12, but these strains differ in a number of ways (Table 2.1 in Materials and Methods provides a full list of genotypic differences). For example, both *araBAD*, encoding enzymes for L-arabinose degradation, and *rhaDAB,* encoding enzymes for L-rhamose degradation, are deleted in BW25113. Furthermore, a section of *lacZ* (encoding β-galactosidase) in BW25113 is replaced with four tandem *rrnB* repeats and a frameshift mutation is found in *hsdR*, encoding a component of the Type I restriction enzyme, EcoKI (Grenier *et al.*, 2014). In their study, Goodall *et al*. compared their list of essential genes with two other studies of gene essentiality in *E. coli* K-12: the PEC database (Yamazaki *et al.*, 2008) and the Keio collection (Baba *et al.*, 2006). This comparative analysis highlighted the significant impact of methodology on determining gene essentiality but also generated a more definitive list of candidate essential genes. Therefore, the essential gene list generated in this study was compared with the analysis undertaken by Goodall *et al*, not only to generate a more comprehensive understanding of genetic requirements for growth in *E. coli*, but also to potentially understand the impact of strain differences on gene essentiality. A Venn diagram illustrating numbers of genes shared between these four studies is presented in Figure 3.5. Candidate essential genes common to all studies of were identified; however, many essential genes that were unique to the two TraDIS

studies were also identified, again highlighting the importance of methodology in gene essentiality studies.

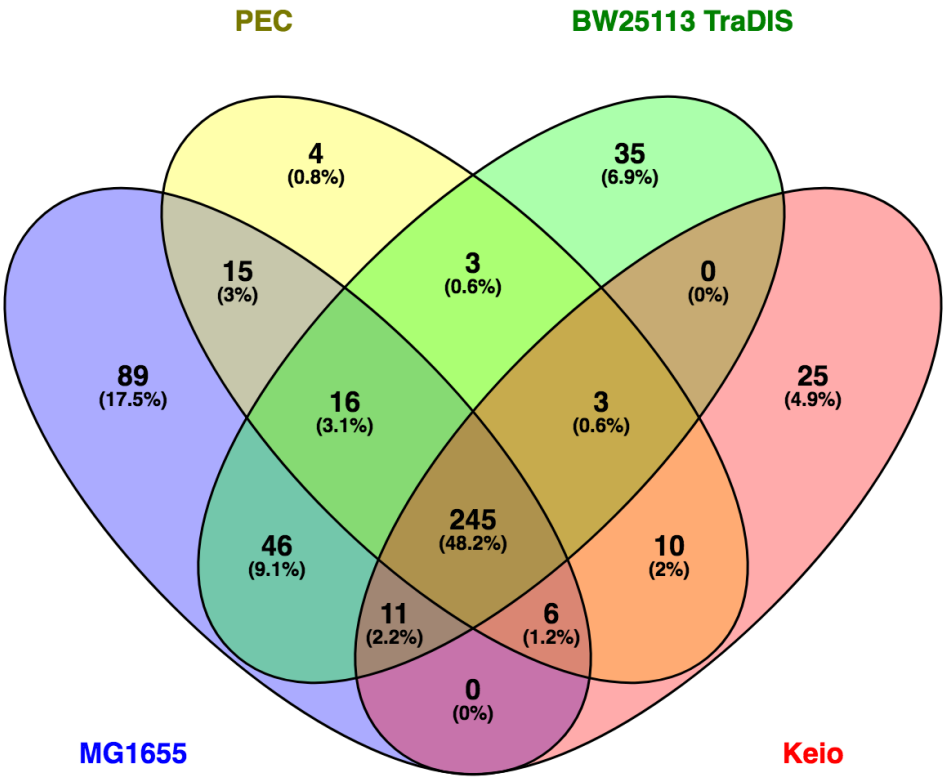
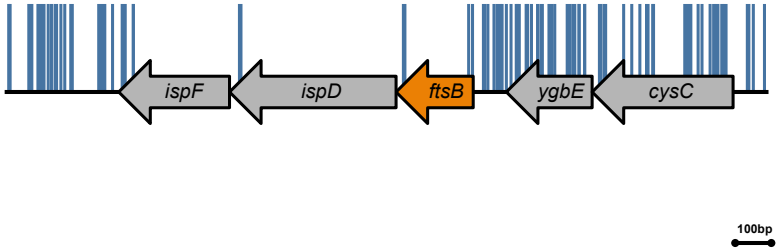


Figure 3.5. Venn diagram comparing candidate essential genes for growth in LB in this study (MG1655; blue), Profiling of the *E. coli* Chromosome database (PEC; Yamazaki *et al.,* 2008; yellow), the Keio library (Baba *et al*; red), and the TraDIS study conducted by Goodall *et al*. (BW25113 TraDIS; green).

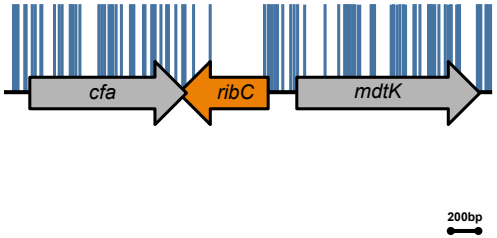
3.2.1. Essential genes identified in all studies

In their study, Goodall *et al*. identified 248 genes that were classified as essential in all studies (Goodall *et al.*, 2018). When the MG1655 TraDIS data from this work was included in this analysis, the number of candidate essential genes was reduced to 245 (Figure 3.5). This suggests that three genes previously identified as essential were not classified as essential in MG1655. These three genes are *ftsB, ribC,* and *glmS*. Both *ftsB* and *ribC,* were annotated as ambiguous in this study. *ftsB* contained two insertions (Figure 3.6A), one at the extreme 3' end of the gene that could still produce a functional gene product, but another at the extreme 5' end of the gene which should interrupt expression of a functional gene product. However, the first 3 amino acids encoded by *ftsB* are not known to influence protein function (The UniProt Consortium, 2017). Therefore, *ftsB* could be reannotated in this study as essential. On the other hand, *ribC* contained two insertions, including one within a predicted lumazine substrate binding region (The UniProt Consortium, 2017). Therefore, *ribC* is likely to contain non-essential regions. However, the remainder of the gene contained no insertions, indicating that the majority of the protein was essential for survival. Therefore, *ribC* could also be nominated as an essential gene.

(A)


ispF
ispD
ftsB
ygbE
cysC
100bp

(B)


cfa
ribC
mdtK
200bp

(C)

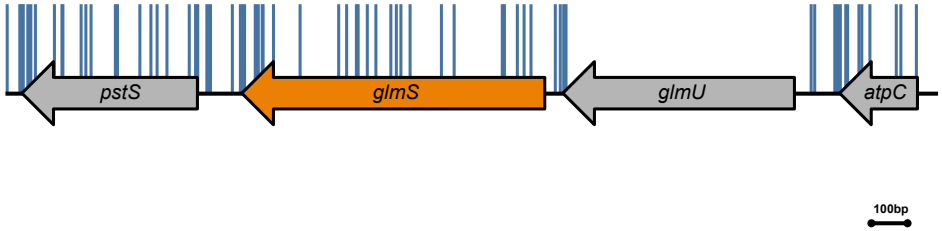

pstS
glmS
glmU
atpC
100bp

Figure 3.6. Transposon insertion plots of genes classified as essential in PEC, Keio, and BW25113 TraDIS but not MG1655 TraDIS. Blue lines indicate transposon insertions in the MG1655 library plotted on the MG1655 genome (Genbank accession No. U000096). Note that the length of the blue insertion lines is kept the same for illustrative purposes; the readcounts for each of the insertions above are not necessarily the same. Genes of interest highlighted in orange, namely (A) *ftsB*, (B) *ribC*, and (C) *glmS*. Other genes are highlighted in grey.

Interestingly, *glmS*, encoding L-glutamine—D-fructose-6-phosphate aminotransferase, was identified as essential in previous studies but non-essential in the MG1655 TraDIS (Figure 3.6C). The *glmS* gene is located directly upstream of the essential gene, *glmU,* encoding a fused N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase. Both enzymes are required for the synthesis of UDP-*N*-acetyl-α-D-glucosamine, an essential precursor of cell wall peptidoglycan (Kotnik *et al.*, 2007). There appeared to be an equal number of insertions in both strands of *glmS*, as well as many insertions in the upstream gene, *atpC* (Figure 3.6C) This result is curious, since it has been previously shown that mutants of *glmS* could only grow in media supplemented with N-acetylglucosamine or glucosamine, or during overexpression of catabolic D-glucosamine-6-phosphate isomerase, *nagB*, involved in the degradation of N-acetylglucosamine (Vogler *et al.*, 1989; Wu & Wu, 1971). However, it must be noted that the insertion index for *glmS* (0.0115) was only marginally above the non-essential cutoff value (0.01). This may highlight the impact of the different ways in which TraDIS libraries were generated between this study and Goodall *et al*. In this study, TraDIS was conducted on transposon mutants taken directly from LB agar plates, whereas Goodall *et al*. analysed their library after passaging (growing) several times in LB (Goodall *et al.*, 2018). Therefore, mutants with slow growth phenotypes, potentially including *glmS*, were under greater selective pressure and were thus removed from the library.

Interestingly, the essentiality of *glmS* mutants in previous studies was determined based on growth in liquid broth (Baba *et al.*, 2006; Goodall *et al.*, 2018; Vogler *et al.*, 1989; Yamazaki *et al.*, 2008). This study reveals that *glmS* mutants are able to form colonies on LB agar, suggesting either the presence of an exogenous source of N-acetylglucosamine/glucosamine for those mutants to use differences in metabolic requirements during growth on agar compared to liquid broth may permit the growth of *glmS* mutants.

3.2.2. Essential genes identified uniquely by both TraDIS studies

This study identified 428 genes as essential in MG1655 (Table A21). Of these, 245 genes were also classed as essential in *E. coli* in other studies (Goodall *et al.*, 2018), and an additional 46 genes were shared with the BW25113 TraDIS study (see Table A3 and Table A4, Appendix). These 46 genes therefore represent candidate essential genes that could only be identified using TraDIS. Using gene ontology (GO) and manual analysis, this group of genes was grouped into several categories, as shown in Figure 3.7.

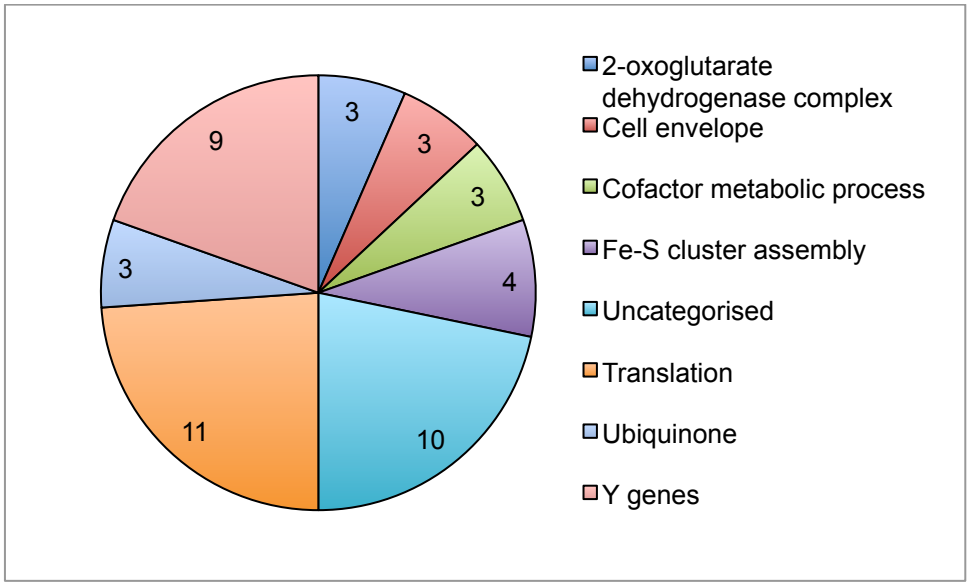


Figure 3.7. Categories of genes unique to and shared by both TraDIS studies.

Insertions in many of these genes are likely to have been identified due to polar effects of the insertion(s) on downstream essential genes (see Table A5, Appendix). In addition, mutants could be falsely identified as essential in TraDIS as the result of a slower growth rate that, in the context of a pooled library, would be out-competed by other mutants in the pool. For example, all three genes encoding the α-ketoglutarate dehydrogenase complex, *lpd, sucA,* and *sucB* are identified as essential in both TraDIS studies. This enzyme complex catalyses the oxidative decarboxylation of α-ketoglutarate to generate succinyl-coenzyme A (CoA) and CO_2 during the TCA cycle (Figure 3.8).

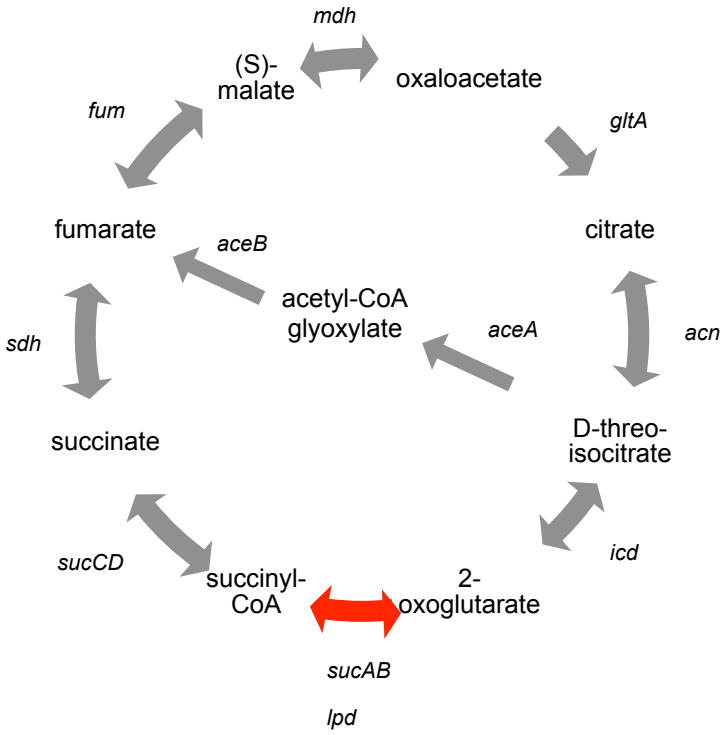


Figure 3.8. The TCA cycle. The interconversion of 2-oxoglutarate (α-ketoglutarate) to succinyl-CoA by the α-ketoglutarate dehydrogenase complex is highlighted in red.

Deletion mutants of *sucA, sucB,* and *lpd* have been shown to be unable to respire several different carbon sources in a Biolog assay, including L-aspartate, L-asparagine, and L-serine (Ito *et al.*, 2005; Sezonov *et al.*, 2007). Therefore, this reduced metabolic flexibility might induce a sufficient selection under the conditions of intense competition such as those experienced during pooled library screening. Other identified genes that might be explained in a similar manner include some that encode for ribosomal subunit proteins (e.g. *rplY, rplA,* and *rps* genes), deletion

mutants of which have been previously shown to reduce the growth rate of *E. coli* K-12 (Korepanov *et al.*, 2007), alongside mutations in members of the *isc* operon (e.g. *fdx, iscU,* and *iscS*) which have all been shown to result in significantly increased doubling times compared to the wild type (Tokumoto & Takahashi, 2001).

3.2.3. Essential genes identified by TraDIS uniquely in MG1655

In total, 89 genes were identified as being uniquely essential in MG1655 (see Table A5 and Figure 3.5). However, many of these genes (n = 37) were small tRNA genes in the region of 75 bp in length. As mentioned above, 75 bp was suggested to be the shortest IFR that could be considered significant (Goodall *et al.*, 2018). Therefore, the small size of these tRNA genes increases the chance that they will be falsely assigned as essential. In addition, all 7 rRNA operons in *E. coli* were represented in this group, comprising 20 genes in total that were identified as uniquely essential in MG1655. This is interesting, as there are 7 copies of the *rrn* operon in *E. coli* and disruption of one operon should not significantly affect growth (Asai *et al.*, 1999; Ellwood & Nomura, 1980). However, it is likely that the lack of insertions in these regions may be due to complex DNA secondary structure (Goodall *et al.*, 2018).

Therefore, as discussed here and in previous sections, false positives can be identified as a result of: (a) genes containing insertions but at a sufficiently low density to provide an essential insertion index score, e.g. *mtn, hscB*; (b) genes being too small to contain significant numbers of insertions e.g. *ibsC, ibsE* (both 59 bp in length); or (c)

insertions exhibiting polarity on nearby essential genes. When these genes are removed from the analysis, seven genes that were exclusively selected for essentiality in MG1655 remained: *ptsH, tufA/tufB, yciM, ydfK, ykgS, ynaE*, and the non-protein-encoding gene, *sdsR* (*ryeB*).

3.2.3.1. ptsH

The *ptsH* gene encodes HPr, the non-sugar-specific component of the sugar phosphotransferase system (PTS^{Sugar}), required for the transfer of a phosphoryl group from Enzyme I to Enzymes II (Figure 3.9).

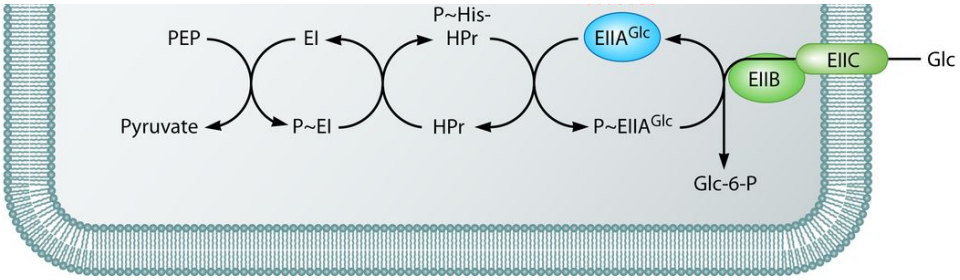


Figure 3.9. The PTS system in *E. coli*. Adapted from (Deutscher *et al.*, 2014).

HPr also has several important regulatory roles, including the allosteric activation of pyruvate kinase (PykF), phosphofructokinase (PfkB), glucosamine-6-phosphate deaminase (NagB), and inhibition of adenylate cyclase, Adk (Rodionova *et al.*, 2017). Moreover, dephosphorylated HPr binds to Rsd, an inhibitor of sigma70, thus preventing Rsd-mediated inhibition of sigma70 during exponential growth (Park *et al.*, 2013). Loss of HPr is not expected to result in a lethal phenotype, with no reports of significant growth defects with *ptsH* mutants

in either BW25113 or MG1655 in LB (Baba *et al.*, 2006; Gerdes *et al.*, 2003; Joyce *et al.*, 2006). TraDIS may have been able to identify a subtle phenotype for *ptsH* under the specific growth conditions in this study, since mutation of *ptsH* could result in global changes in carbon and energy metabolism in LB that would affect growth rate (Rodionova *et al.*, 2017). However, it is not clear why *ptsH* was not selected in the TraDIS study of BW25113 (Goodall *et al.*, 2018). The library constructed by Goodall and colleagues had an insertion density equivalent to 1 insertions per 5.14 bp, approximately five times more dense than our library of 1 insertion per 24 bp (Goodall *et al.*, 2018). Therefore, the MG1655 library constructed for this study is more likely to identify genes as essential due to growth rate differences, since a reduction in growth, and this readcount/insertion count thresholds may be met more quickly.

3.2.3.2. tufA and tufB

Both *tufA* and *tufB* encode the translation elongation factor Tu (EF-Tu), a protein that binds and delivers aminoacylated tRNAs, to the translating ribosome (see Figure 3.10). EF-Tu is also the most abundant protein in *E. coli* (Weijland *et al.*, 1992). It is possible to create viable mutants of *tufA,* however this results in reduced growth and ribosome production (Gausing, 1981). Interestingly, mutants of *tufB* have been shown previously to display growth rate defects in *E. coli* K-12 strain LB1001 (Van de Klundert *et al.*, 1978). Furthermore, in a comparative analysis using Tn-seq, *tufB* was shown to be essential in *Shigella flexneri* 2a 2457T but not BW25113 (Freed *et al.*, 2016). Therefore, transposon

sequencing appears to be able to identify strain-specific differences in the requirement for EF-Tu in *E. coli*. A molecular explanation for these different requirements has not yet been established.

3.2.3.3. yciM

The *yciM* gene (also called *lapB*) encodes a lipopolysaccharide assembly protein which couples LPS biosynthesis and transport (Figure 3.11; (Klein *et al.*, 2014). LapB is also employed during the heat shock response to maintain envelope integrity (Nicolaes *et al.*, 2014). The *lapB* gene is located immediately downstream from *lapA* and both genes are predicted to be transcribed as an operon (Nonaka *et al.*, 2006). Interestingly, *lapA* was identified as essential in both MG1655 and BW25113 TraDIS studies (this study and Goodall *et al.*, 2018).

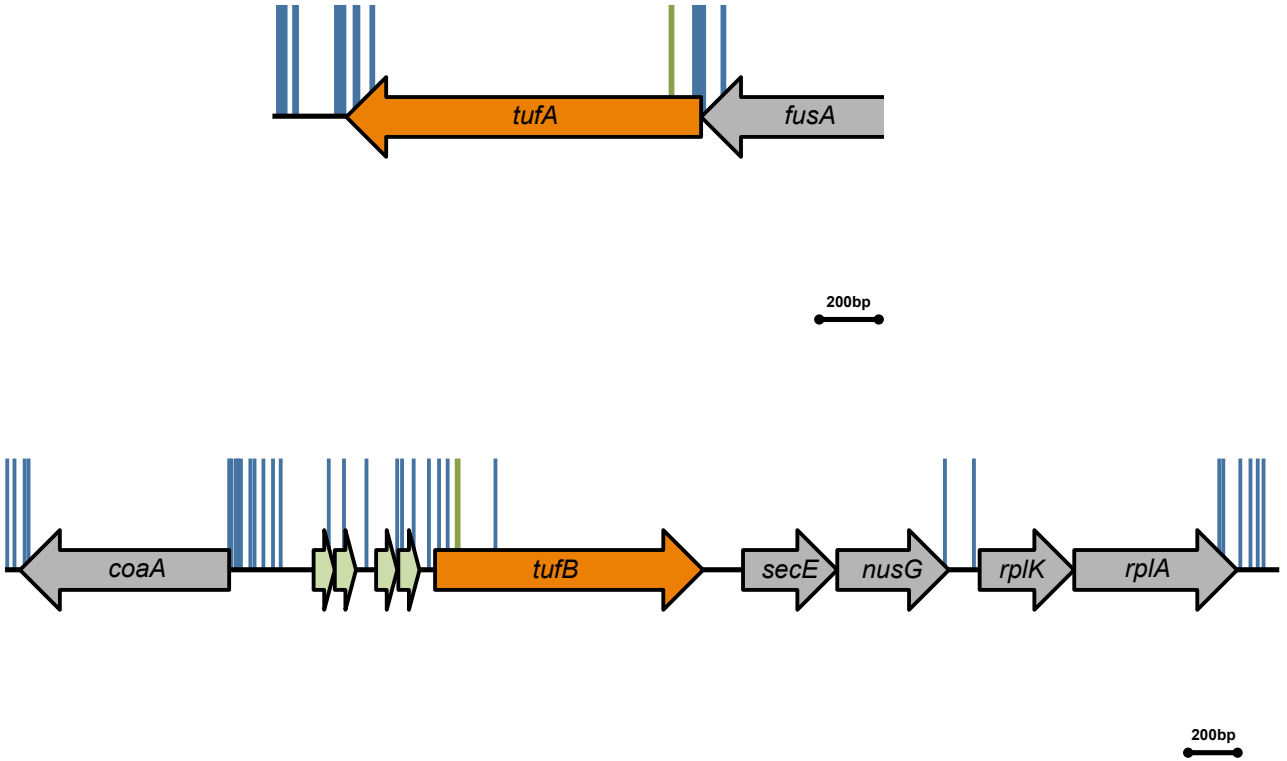


Figure 3.10. Transposon insertions in *tufA* (top) and *tufB* (bottom). Blue lines indicate transposon insertions (length of line does not reflect readcounts). Green lines indicate transposon insertions in GTPase domains.

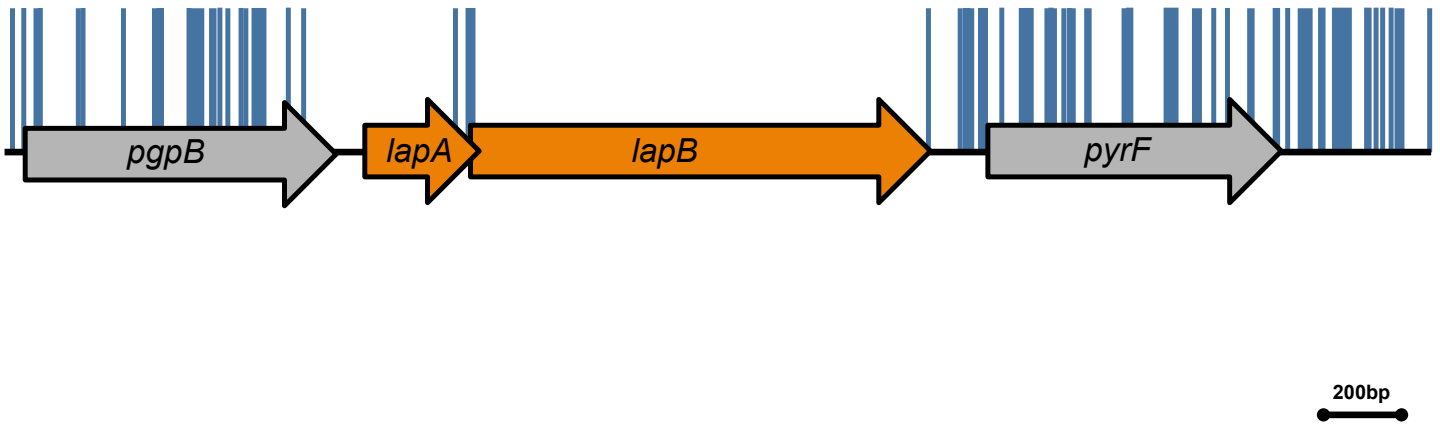


Figure 3.11. Transposon insertions in *lapAB* and surrounding genes. Blue lines indicate transposon insertions (length of line does not reflect readcounts).

However, *lapB* is only identified as essential in MG1655 and other studies are conflicting in their findings. For example, it has been shown that *lapB* mutants of BW25113 could grow in standard LB, although one study (Baba *et al.*, 2006) deemed the growth of their mutant 'indeterminate' (Baba *et al.*, 2006; Klein *et al.*, 2014). On the other hand, another study showed that *lapB* was required for growth in MG1655 (Mahalakshmi *et al.*, 2014). This study demonstrated that the Δ*lapB* mutant strain present in the Keio library contained a suppressor secondary mutation which mitigated the essentiality of *lapB*. Indeed, transfer of the Δ*lapB* mutation to the Keio parental strain (BW25113) or to MG1655 by P1 transduction resulted in no transductants in LB, showing that *lapB* was required for growth in both K-12 strains (Mahalakshmi *et al.*, 2014). It should be noted that the essentiality of *lapB* was deemed 'unclear' in BW25113 following TraDIS, implying some *lapB* insertion mutants could grow, albeit poorly (Goodall *et al.*, 2018). Therefore, while some studies do indicate MG1655 requires *lapB,* including this TraDIS analysis, it is not entirely clear whether this is the case in BW25113.

3.2.3.4. Prophage-associated genes

ynaE, ydfK, and *ykgS* were identified as essential in the MG1655 TraDIS only, and the genes are each found within the predicted genomes of the Rac, Qin, and CP4-6 prophages, respectively (Figure 3.12).

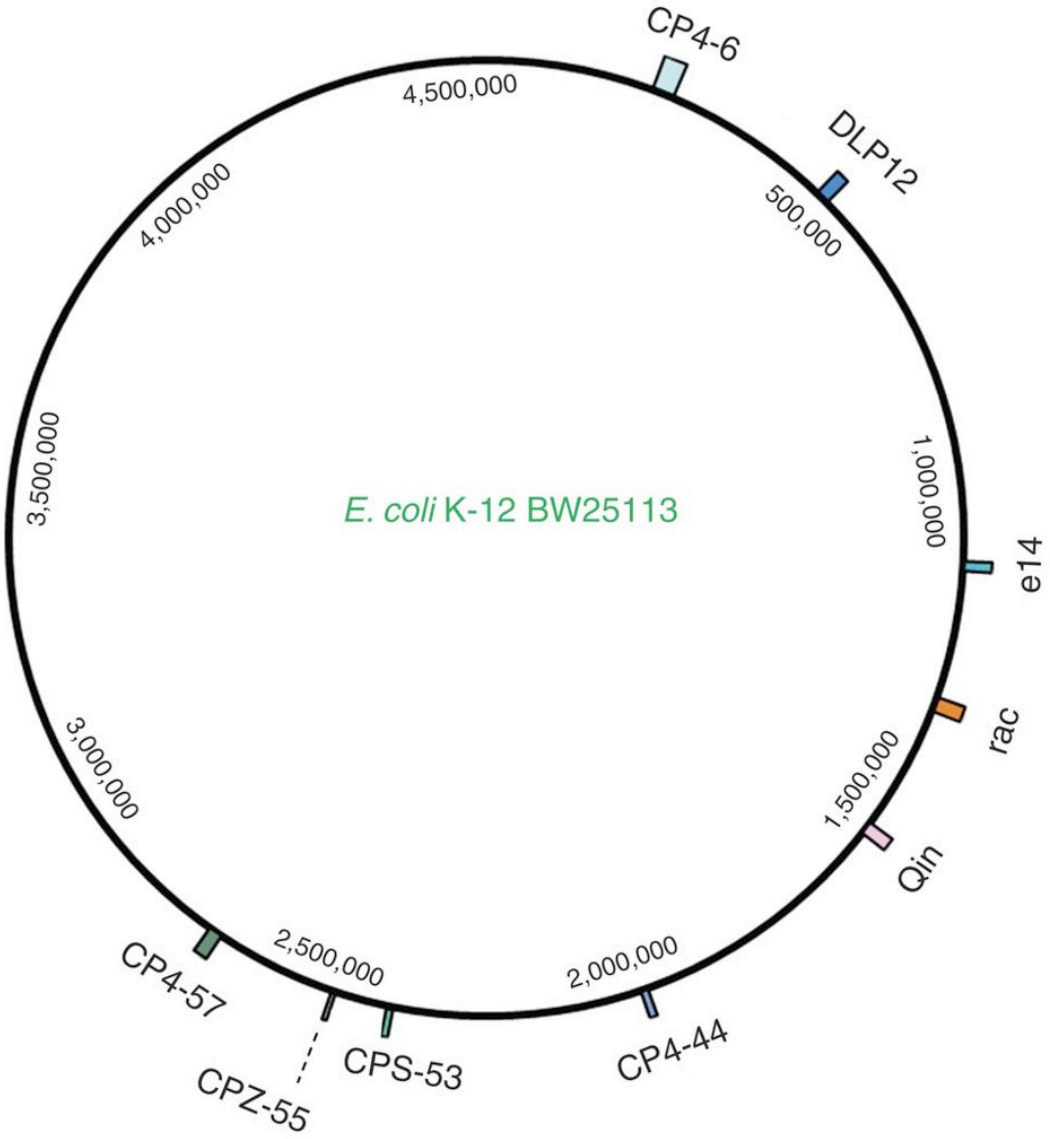


Figure 3.12. Map of the *E. coli* K-12 BW25113 genome illustrating its cryptic prophages. Adapted from Wang *et al*. (Wang *et al.*, 2010).

Both *ynaE* and *ydfK* appear to encode cold shock proteins that are upregulated in response to cold shock (Polissi *et al.*, 2003; Raghavan *et al.*, 2011), while *ykgS* is a small protein that is expressed at a greater level during stationary phase compared to exponential phase (VanOrsdel *et al.*, 2018). The precise role for these genes during growth is not known, however it is thought that cryptic prophages in *E. coli* may play an

important role during adaptation to some stresses (Wang *et al.*, 2010). It is difficult to explain why there was no selection for these genes in the BW25113 TraDIS, but it may be related to differences in library screening conditions, transposon density, or other unknown factors.

3.2.3.5. SdsR

One advantage of TraDIS is that it can also identify and measure the essentiality of non-protein-encoding regions of DNA. This is highlighted by the identification of the small regulatory RNA, SdsR (formerly RyeB), as being essential in MG1655. SdsR expression is activated in stationary phase or in cells exposed to heat shock, and is subject to control by σ^S (Fröhlich *et al.*, 2012). SdsR base-pairs with mRNA of the mismatch repair protein, MutS, repressing its synthesis post-transcriptionally (Gutierrez *et al.*, 2013). Furthermore, it has been shown to downregulate the synthesis of the major outer membrane porin, OmpD, in *Salmonella*, via Hfq-dependent base pairing with *ompD* mRNA (Fröhlich *et al.*, 2012). However, *mutS* is not identified as essential in any of the studies mentioned here, and *E. coli* K-12 strains do not encode an *ompD* homologue, therefore it is possible that SdsR interacts with other mRNAs in *E. coli* K-12.

3.2.4. Essential genes uniquely identified by TraDIS uniquely in BW25113

The comparative analysis carried out in this study identified 34 genes that were unique to the BW25113 TraDIS study published by Goodall *et al.* (Goodall *et al.*, 2018). Interestingly, many genes allocated as 'ambiguous'

in MG1655 TraDIS were considered essential in Goodall *et al*'s analysis (see Table A2 and Table A6, Appendix), suggesting that under certain conditions these mutants could survive. This may again highlight the impact of a number of different factors mentioned this chapter, including varying starting populations of mutants, differences in library screening, or differences in the composition of LB medium between studies. Either way, the absolute requirement for these genes (and indeed those unique to MG1655) for growth in LB is circumspect, and it is likely that most were selected for the very same reasons as genes unique to MG1655 TraDIS.

Conclusions

The composition of a pooled transposon mutant library in *E. coli* K-12 strain MG1655 was characterised using TraDIS. This library was found to contain 193,495 unique insertion sites in the MG1655 genome, providing a resolution of approximately 1 insertion every 24 bp. According to the parameters of insertion index and log likelihood ratio, 428 genes were found to be essential i.e. containing a sufficiently low insertion index to be considered as essential for growth on LB agar plates. Comparison of these 428 genes with other lists of essential genes in *E. coli* K-12 identified 247 genes that were universally essential for growth in LB. 46 genes were identified as essential uniquely in TraDIS studies of MG1655 and BW25113, while 89 genes were identified as essential by the MG1655 TraDIS study alone. Most genes in the latter two categories (i.e. TraDIS studies) were probably selected by due to reduced mutant growth

rates, however some genes, including *tufA*, *tufB*, and *lapB* were shown to be uniquely required for growth by MG1655 in a way that indicates potential strain-specific requirements for these genes. However, the comparative analysis also identified many 'false positives', annotating genes as essential not due to a core metabolic or physiological requirement for these genes, but due to reasons of e.g. growth rate differences of mutants, differences in library construction (e.g. selection on agar vs liquid broth), differences in library density, or gene size being too small to accurately assign essentiality (e.g. when applying an IFR filter). This highlights the importance of standardising production and analysis of TraDIS libraries to facilitate future comparative studies.

Chapter 4 TraDIS analysis of genetic requirements of *E. coli* growth in the presence of bile

Introduction

One of the strongest anatomical barriers to entry of colonising microorganisms to the GI tract is bile (Sarker & Gyr, 1992). Therefore, to understand *E. coli*'s responses to bile is crucial in describing the challenges the bacterium faces when colonising the GI tract. Moreover, it has been suggested that the infectivity of pathogenic bacteria could be directly associated with their ability to resist bile, but that this would require an understanding of the bacterium's ability to grow within environments that contain bile (Merritt & Donaldson, 2009).

Growth in the presence of bile involves withstanding its antimicrobial effects, the most potent of which is its ability to disrupt the cell membrane (Begley *et al.*, 2005). This can be seen microscopically, where cells exposed to bile are shrunken and empty, or biochemically, where cell leakage can be detected in cells exposed to bile via enzyme assays (Fujisawa & Mori, 1996; Leverrier *et al.*, 2003; Noh & Gilliland, 1993; de Valdez *et al.*, 1997). Also, many membrane-associated functions and structures are mobilised by bacteria upon exposure to bile such as OMPs, efflux pumps, and cell membrane biosynthesis genes (Lee *et al.*, 2000; Nikaido, 2003; Rincé *et al.*, 2003; Ruiz *et al.*, 2007; Thanassi *et al.*, 1997). Moreover, LPS, the enterobacterial common antigen (ECA), PhoPQ, Tol-Pal, and other membrane-linked functions have been shown to be required for bile resistance in members of the *Enterobacteriaceae* (Merritt & Donaldson, 2009; Nikaido, 2003; Ramos-Morales *et al.*, 2003; Ray *et al.*, 2000; van Velkinburgh & Gunn, 1999).

However, many factors contribute to the antimicrobial property of bile, including low pH, the presence of immunoglobulin A and antimicrobial peptides, high osmolarity, and calcium and iron chelators. Furthermore, bile acids (also referred to as bile salts) can induce oxidative damage that is harmful to DNA and proteins (Begley *et al.*, 2005; Merritt & Donaldson, 2009). Owing to the fact that bile elicits so many antimicrobial effects at once, it is likely that several cellular systems are employed simultaneously in order to resist its effects, pointing towards the importance of studying bile resistance using whole-cell or systems-level analyses. Surprisingly, despite the advent of high-throughput genomic technologies, few studies of the whole-cell or whole-genome response of *E. coli* to bile exist. Using microarrays, it was shown that, in *E. coli* O157:H7, bile induces expression of the *acrAB* efflux system, the *basRS* two-component system, a lipid A modification system (*arn, ugd*), and increased resistance to polymyxin, suggesting that bile salt stress induces *E. coli* O157:H7 to produce protective mechanisms for the outer membrane (Kus *et al.*, 2011). In a similar study in O157:H7 growing in the presence of bile salts, transcriptomic analysis showed significant alterations in the expression of the locus of enterocyte effacement (LEE) pathogenicity island, increased expression of the flagella hook-basal body structure, decreased expression of 'late' flagellar genes, such as those for the filament and stator motor, decreased expression of chemotaxis genes, and increased expression of genes required for iron scavenging and metabolism (Hamner *et al.*, 2013). This

indicates that pathogenic *E. coli* may use bile as a signal to modulate virulence.

The first published TraDIS study identified the genomic requirements of *Salmonella enterica* Typhi during growth in LB medium supplemented with ox bile (Langridge *et al.*, 2009). This study identified mutants in 168 genes that displayed significantly reduced fitness during growth in the presence of 10% ox bile. These genes grouped into energy metabolism, membrane/surface structures, central/intermediary metabolism, and the degradation of macromolecules (Langridge *et al.*, 2009). Some genes within these categories were previously implicated in bile resistance, such as *waa* genes (encoding the LPS core), the *acrAB-tolC* bile salt efflux pump, the *phoPQ* two-component system, and *dam* DNA methyltransferase (Begley *et al.*, 2005; Langridge *et al.*, 2009). However, many genes had not previously been implicated in bile resistance, including over 30 putative/hypothetical genes (Langridge *et al.*, 2009).

The aim of this study was to use TraDIS to comprehensively investigate the genetic requirements for the growth of *E. coli* K-12 in the presence of ox bile. Results obtained by TraDIS would then facilitate a comparative analysis with *S.* Typhi in order to identify conserved bile resistance strategies in the Enterobacteriaceae. This study reveals broad insights into the genetic requirements of *E. coli* during growth in bile, including genes with and without prior links to bile tolerance.

4.1. TraDIS

The MG1655 transposon mutant library described in Chapter 3 was inoculated into LB containing either 2% or 10% ox bile as described in Materials and Methods. Cultures were set up in duplicate, generating four libraries that were put forward for TraDIS sequencing. TraDIS sequencing was carried with the help of Fiona Crispie and Laura Finnegan of Teagasc Moorepark. All subsequent analysis was carried out during this thesis. The INPUT library was taken from the mutant library grown in LB and the OUTPUT library was taken after the INPUT library was grown in LB medium with 2% bile, followed by LB medium with 10% bile.

File	Total Reads	% Mapped	Unique Insertion Sites (UIS)	Sequence Length/UIS
INPUT 1	3551492	95.25184464	156452	29.66821773
INPUT 2	3611440	98.07458265	159343	29.12993982
OUTPUT 1	3372647	97.69497317	181451	25.58074632
OUTPUT 2	3377357	97.27516664	190842	24.32196267

Table 4.1. Mapping statistics for TraDIS library replicates.

4.1.1. Sequencing results

The sequencing reads were processed and mapped to the *E. coli* str. K-12 MG1655 reference genome (U00096) as described in Materials and Methods and mapping statistics are shown in Table 4.1. Approximately 3.5 million reads were generated for each library, with at least 95% of reads mapping to the MG1655 genome. Scatterplots of readcounts and transposon insertion counts for each gene between replicate libraries were generated and with the respective R^2 values calculated. These indicated a high degree of correlation between replicates (see Figure 4.1).

Fitness requirements for each gene were measured based on relative changes in mutant frequency between INPUT and OUTPUT libraries and were expressed as Log_2 fold-change (logFC) values for each gene. LogFC values were subsequently used as the primary measure of fitness (i.e. requirement for growth in bile). In total, mutants in 214 genes displayed significant logFCs (see Table A7, Appendix); mutants in 137 genes had a negative logFC value after growth in bile (suggesting that these genes were required for fitness under these conditions), while mutants in 77 genes had a positive logFC value after growth in bile (suggesting that mutations in these genes provided a fitness advantage during growth in bile; see Table A7).

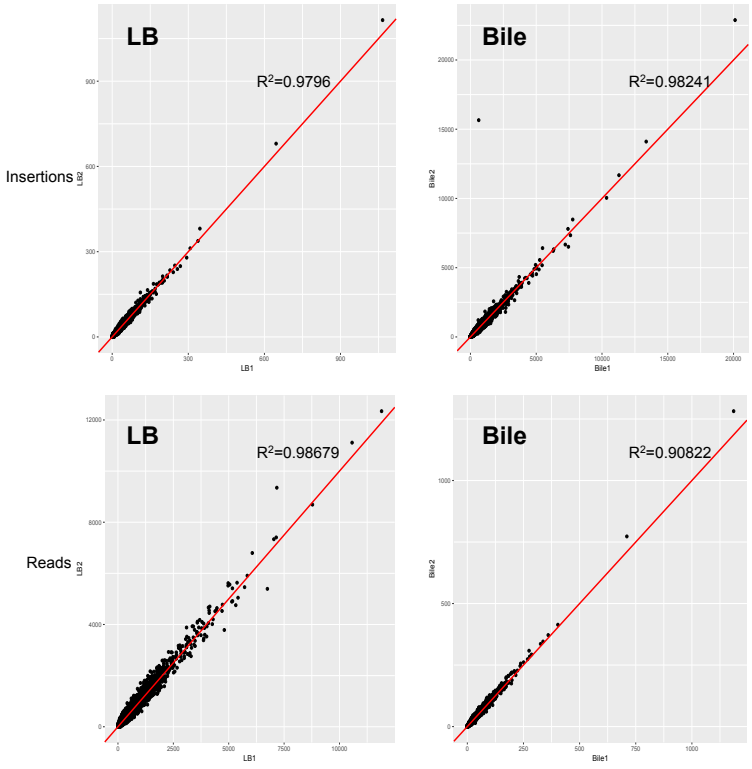


Figure 4.1. Scatterplots and R^2 values of INPUT (left column; 'LB1' and 'LB2') and OUTPUT (right column; 'Bile1' and 'Bile2') replicate libraries. Insertion counts (top row) and readcounts (bottom row) are shown.

To identify general functions required for fitness in bile, all genes with significant logFC values were initially assigned a gene ontology (GO) term and then grouped by functional enrichment analysis. However, this resulted in many genes being classified into uninformative higher-order GO terms such as 'metabolic process'. Therefore, genes were analysed individually for annotation/predicted function by referral to online databases including Ecocyc (https://ecocyc.org/), Ecogene (http://ecogene.org/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed), and others (Keseler *et al.*, 2017; Mi *et al.*, 2017; Zhou & Rudd, 2012). In this way, the following potential functional groups were manually generated: cell envelope constituent/homeostasis; lipopolysaccharide (LPS); enterobacterial common antigen (ECA), outer membrane proteins (OMPs); lipid transport; cell division; peptidoglycan (PG) synthesis and turnover; osmoregulation; NADH dehydrogenase; response to bile; oxidative stress; and DNA replication and repair (Table 4.2; also see Table A8, Appendix).

Functional group	Example gene/operon	Input insertion count	Ouput insertion count
Cell envelope constituents	*ompA*	19	3
LPS	*waaQGPSBOJYZU*	379	116
ECA	*rfe-wzzE-wecBC-rffGHC-wecE-wzxE-rffT-wzyE-rffM*	242	71
Cell division	*damX*	23	1
PG synthesis and turnover	*mltA*	29	16
Osmoregulation	*opgGH*	78	9
NADH dehydrogenase I	*nuoABCEFGHIJKLMN*	91	1
Bile tolerance	*acrAB*	188	1

Table 4.2. Functional categories of genes with significant logFCs in TraDIS. Example genes/operons listed alongside total numbers of insertions in input and output libraries.

4.1.2. Genes with negative logFC values

The majority of the 214 genes identified by TraDIS displayed negative logFC values (n = 137 or 64%), of which many have previously been identified as important for bile adaptation and tolerance in *E. coli* and *Salmonella* (Hernández *et al.*, 2012). These genes included the well characterised bile resistance mechanisms mentioned previously, such as *acrAB-tolC* (bile salt efflux)*, phoPQ* (transcriptional response to bile)*, waa* (LPS core biosynthesis)*, wec/rfe* (ECA biosynthesis)*, tol-pal* (cell envelope integrity), and *dam* (DNA methylation) (Merritt & Donaldson, 2009; Nikaido, 2003; Ramos-Morales *et al.*, 2003; Ray *et al.*, 2000; van Velkinburgh & Gunn, 1999). Furthermore, many genes with mutants showing reduced fitness were cell envelope-associated, eluding to the increased selective pressure bile can exert upon the cell membrane (Begley *et al.*, 2005). Therefore, the list of genes with negative logFCs verified that TraDIS accurately described a condition under which bile was a selective pressure, and, as such, results obtained from TraDIS should accurately reflect genetic requirements for growth of *E. coli* in bile.

The roles of some genes with negative fitness scores during growth in bile were unexpected. For example, most *nuo* genes, encoding NADH dehydrogenase I, showed logFC values of < -8, indicating a substantial negative selection due to bile (see Table 4.3).

Gene	Function	logFC
nuoL	NADH:ubiquinone oxidoreductase, membrane subunit L	-9.099262266
nuoJ	NADH:ubiquinone oxidoreductase, membrane subunit J	-9.127948557
nuoI	NADH:ubiquinone oxidoreductase, chain I	-8.31115119
nuoH	NADH:ubiquinone oxidoreductase, membrane subunit H	-9.077033141
nuoG	NADH:ubiquinone oxidoreductase, chain G	-10.51441405
nuoF	NADH:ubiquinone oxidoreductase, chain F	-9.991423561
nuoC	NADH:ubiquinone oxidoreductase, fused CD subunit	-9.710277668
nuoB	NADH:ubiquinone oxidoreductase, chain B	-9.408109621
nuoA	NADH:ubiquinone oxidoreductase, membrane subunit A	-4.275454887
nuoN	NADH:ubiquinone oxidoreductase, membrane subunit N	-9.064111238
nuoM	NADH:ubiquinone oxidoreductase, membrane subunit M	-9.358471301

Table 4.3. LogFC values of members of the *nuo* operon, encoding NADH dehydrogenase I.

The role NADH dehydrogenase during growth in bile is not clear, although it has been shown to be required for bile resistance in *Listeria monocytogenes,* possibly by influencing cellular redox state (Wright *et al.*, 2016). NADH dehydrogenase is also embedded in the inner membrane, therefore it is also possible that disruption of this enzyme could disturb inner membrane integrity. Moreover, it has been shown previously that deletion of *nuo* operon genes prevented the stationary phase induction of the Cpx response, required for the coordination of the envelope stress response in *E. coli*, and that NADH dehydrogenase is generally required for maintaining basal levels of the activity of the Cpx two-component system (Guest *et al.*, 2017).

Furthermore, 17 unannotated y genes were included in this list, indicating that some processes required for fitness in bile remain to be characterised. Significantly, many of the y genes have putative functional annotations related to other processes selected in TraDIS, potentially validating these annotations. For example, the expression of *yrbL*, encoding a protein kinase-like domain-containing protein, has been shown to be upregulated by PhoP in response to Mg^{2+} concentration, and also in a mutant with a constitutively active EvgSA two-component system, which is known to interact with the PhoPQ system (Eguchi *et al.*, 2004; Minagawa *et al.*, 2003). Cell division genes form a significant functional group within the TraDIS dataset. YraP is implicated in the NlpD-mediated activation of AmiC (N-acetylmuramoyl-L-alanine amidase C), an important process during cell separation, and genes encoding these three proteins (i.e. *yraP, nlpD,* and *amiC*) are under negative selection in TraDIS (Tsang *et al.*, 2017). Furthermore, *yraP* is a member of the RpoE (the cell envelope stress sigma factor, σ^E) regulon, suggesting it may have a role during envelope stress (Dartigalongue *et al.*, 2001). The fitness requirement for *yraP* in bile may therefore be evidence that YraP plays a role in facilitating cell separation under conditions of cell envelope stress.

4.1.3. Genes with positive logFC values

A relatively large number of genes (approximately 36% of the total list) with positive logFC values were also identified, indicating that insertions in these genes improved relative fitness during growth in bile. For

example, mutants in the transcriptional repressor, *acrR*, show positive logFCs, suggesting that lifting AcrR-mediated repression is advantageous for growth in the presence of bile. In particular, AcrR is a repressor of *acrAB*, suggesting that derepression of this operon is advantageous since it would allow increased expression of this vital bile adaptation mechanism (Su *et al.*, 2007). However, the fitness advantage of mutants of several genes was less immediately clear, including mutants of *dsbA, skp, lptC, lpxM, kdsD,* and *kdsC*.

4.1.3.1. dsbA

DsbA is a periplasmic protein disulfide isomerase that is responsible for generating disulfide bonds in the *E. coli* periplasm (Landeta *et al.*, 2018; Shouldice *et al.*, 2011). The activity of DsbA requires redox activity, and DsbB is required to reoxidise DsbA during disulfide bond formation (Figure 4.2).

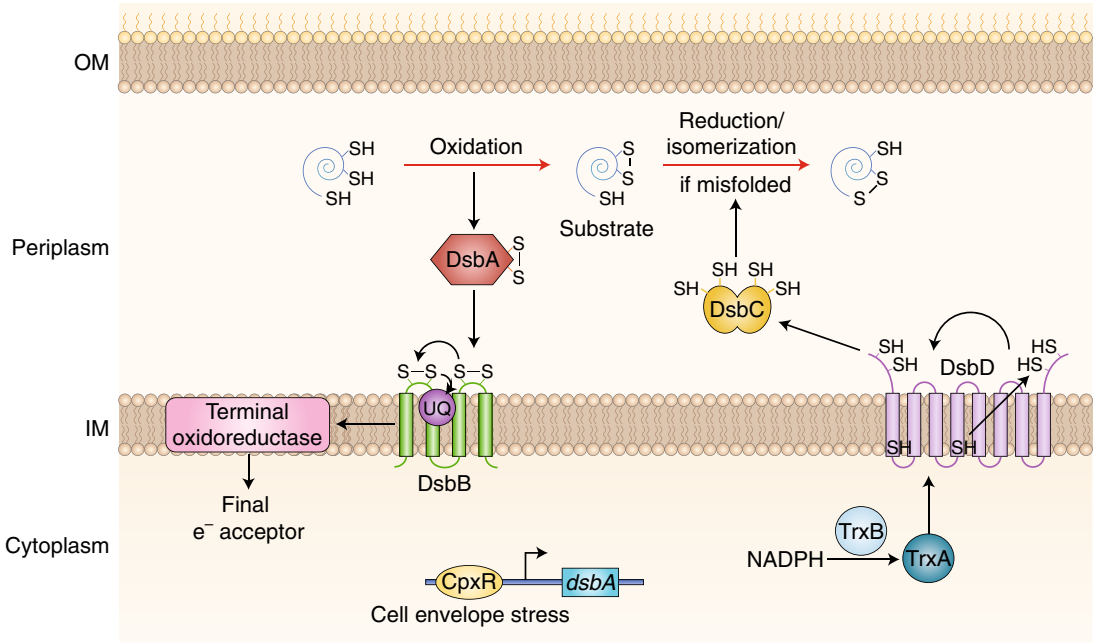


Figure 4.2. The prototypical disulfide bond (DSB)-forming system in *E. coli*. To start a new catalytic cycle, DsbB reoxidizes DsbA by transferring the electrons to ubiquinone. If DsbA introduces incorrect disulfides in the substrate, DsbC reduces the non-native DSBs and allows the formation of native bonds. The electrons for this reduction process come from cytoplasmic thioredoxin to DsbD, and then to DsbC. From Landeta *et al*., 2018.

However, DsbA can introduce inappropriate disulfide bridges that negatively affect protein function (Berkmen *et al.*, 2005). As a countermeasure to this, the DsbC/DsbD system corrects DsbA-induced misfolding so that disulfide bonds are formed in their correct positions (Vertommen *et al.*, 2008). In TraDIS, *dsbA, dsbB,* and *dsbC* all had positive logFC values with observable increases in transposon insertion frequency in the output libraries (Figure 4.3A). Indeed, *dsbA* was represented among the genes with the most positive logFC values (logFC of +3.15), suggesting that insertion mutants in this system resulted in a substantial fitness advantage during growth in the presence of bile. This observation was verified by subsequent competition assays between a BW25113 Δ*dsbA* mutant from the Keio library and the wild type (Figure 4.3B). In this experiment, the Δ*dsbA* mutant displayed a competitive index (CI) of >4, confirming that the Δ*dsbA* mutant has a competitive advantage over the wild-type during growth in the presence of bile (Figure 4.3B). In contrast to what was observed in TraDIS, the Δ*dsbB* mutant from the Keio library did not display a significant competitive advantage over the wild type, and a Δ*dsbC* deletion mutant was less competitive than the wild type, in the presence of bile.

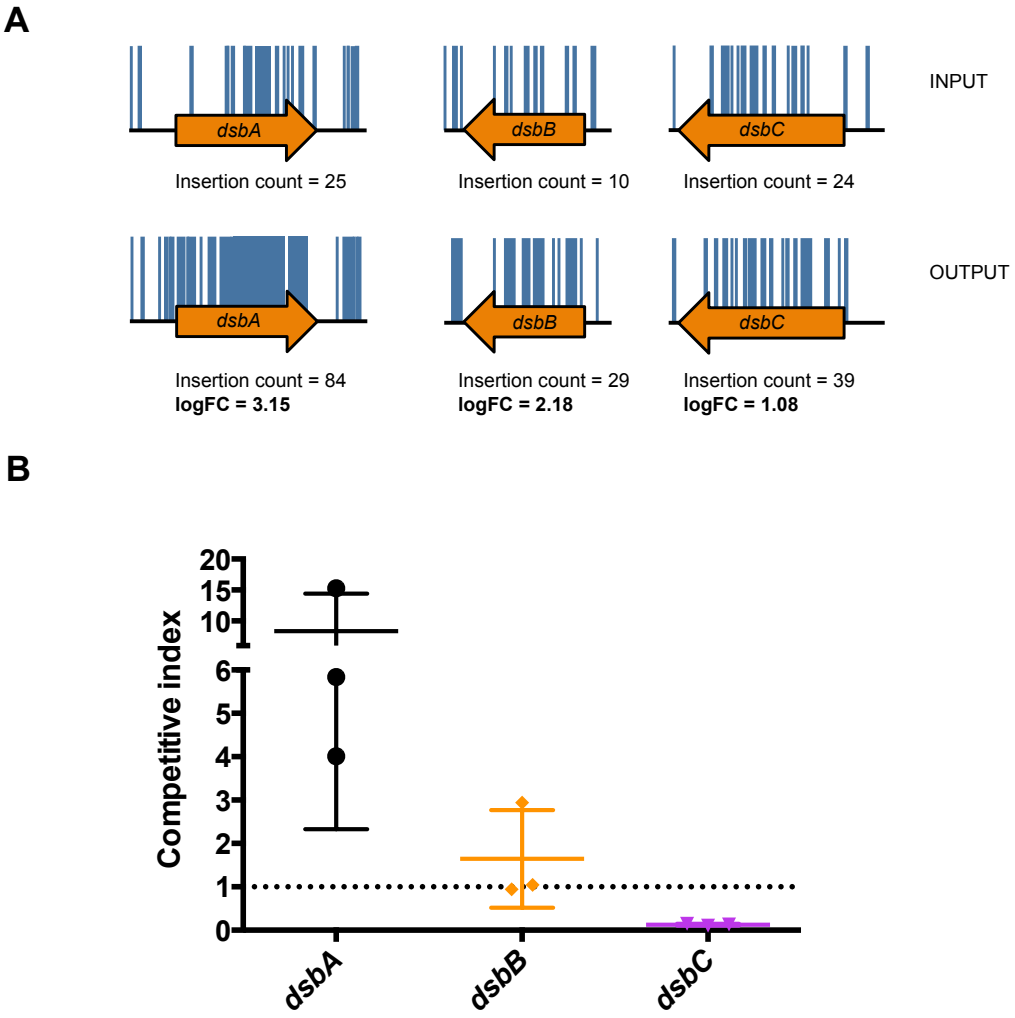


Figure 4.3. (A) Transposon insertion plots of *dsbA, dsbB,* and *dsbC*. Blue lines indicate transposon insertions (length not scaled to readcounts) in the input (top row) and output (bottom row). Insertion counts and LogFC values for each gene indicated. **(B)** Competition assay of *dsbA, dsbB,* and *dsbC* whole-gene deletion mutants in strain BW25113, derived from the Keio library. The dotted line indicates a competitive index of 1. Each dot represents the CI calculated from a biological replicate and experiments were carried out at least three times.

DsbA has a significant impact on the activity of a key mediator of bile resistance, PhoQ. The activity of the PhoQ sensor protein is repressed by MgrB, a protein that requires the generation of disulfide bonds at C28 and C39 via DsbA. Therefore, in the absence of either DsbA or DsbB, MgrB-mediated repression of PhoQ does not occur, resulting in the possible overexpression of PhoQ-regulated genes and increased bile resistance. Furthermore, *dsbA* knockout mutants have been shown to display increased abundances of proteins encoded by genes determined as being required for fitness in this study i.e. *ompA, opgG,* and *yrpG* (Vertommen *et al.*, 2008). Whether altered abundance of these proteins is also evident in *dsbB* and *dsbC* mutants is not known.

4.1.3.2 skp

Skp encodes a periplasmic chaperone, active during the folding and transport of OMPs to the outer membrane (OM). In particular, Skp is required for the folding and insertion of OmpA into the OM, as well as contributing to the assembly of the essential LPS assembly machinery component, LptA (Bulieris *et al.*, 2003; Schwalm *et al.*, 2013). However, Skp may interact with over 30 other proteins, primarily OMPs and other membrane-associated proteins (Jarchow *et al.*, 2008). According to TraDIS, insertions in *skp* were highly advantageous during growth in bile with a positive logFC value of 2.51 (Figure 4.4A). Moreover, when the Δ*skp* mutant from the Keio library was tested in a competition assay against the wild type, it displayed a competitive index as high as 20,

confirming that the loss of *skp* function was highly advantageous during growth in bile (Figure 4.4B).

In the absence of Skp, OmpA is not released from the inner membrane (Schäfer *et al.*, 1999). OmpA is a nonspecific diffusion channel, therefore it could be hypothesized that loss of OmpA leads to increased bile resistance due to less crossover of harmful substances into the cell. However, this is unlikely as insertions in *ompA* mutants displayed negative logFC values (-3.37) in TraDIS, suggesting a fitness disadvantage associated with loss of OmpA. Moreover, it has been shown in *E. coli* O157:H7 that expression of *ompA* is not decreased in response to bile; in fact, expression of a repressor of *ompA* expression, *tdcA*, decreases in the presence of bile (Hamner *et al.*, 2013). A Tn10 transposon insertion mutant of *skp* has been shown to have an induced σ^E regulon and this might be responsible for the increased fitness observed in the presence of bile (Missiakas *et al.*, 1996).

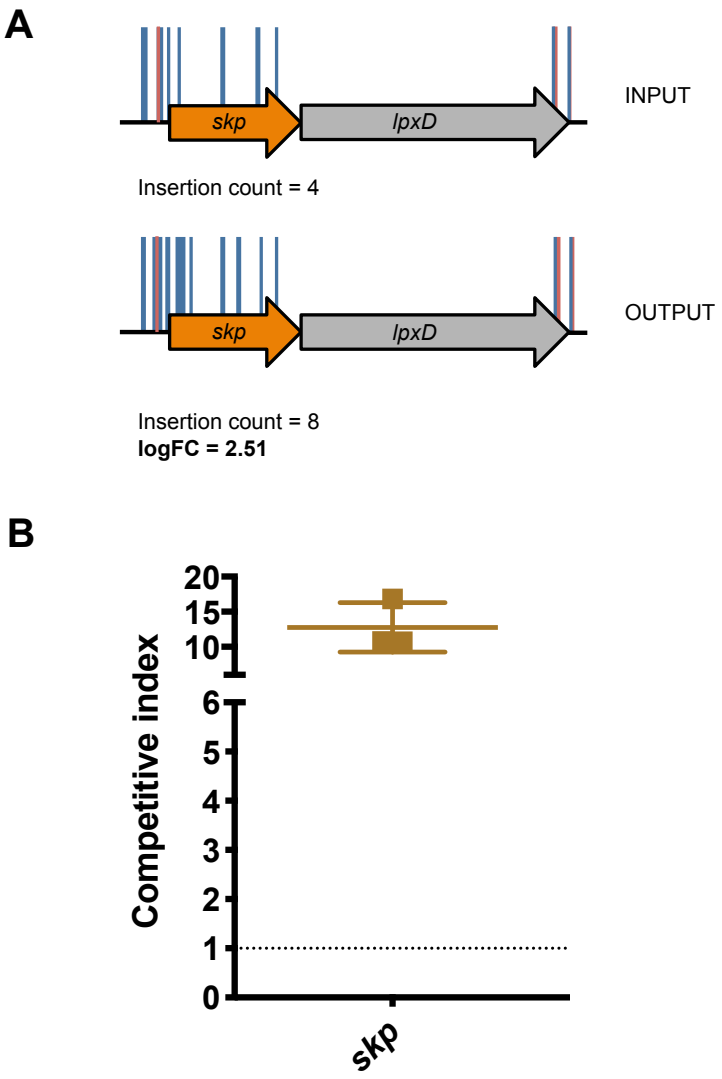


Figure 4.4. (A) Transposon insertion plots of *skp*. Insertions plots of the input (top row) and output (bottom row) are shown. The majority of transposon insertions in *skp* are oriented in one direction only, and so these orientations are highlighted by blue (plus strand) and red lines (minus strand). Newly acquired insertions are highlighted in purple The length of insertion lines is not scaled to readcounts. Insertion counts and logFC value indicated. **(B)** Competition assay of the *skp* Keio whole-gene deletion mutant. The dotted line indicates a competitive index of 1, equivalent to the competitiveness of the wild type.

The *skp* gene is located upstream from the essential gene, *lpxD*, involved in the third step of lipid A biosynthesis (Figure 4.4A). Both genes can be expressed on the same transcriptional unit, and expression is mediated by the transcriptional regulator, CpxR (Dartigalongue *et al.*, 2001). Examination of transposon insertion plots (Figure 4.4A) reveals that, after growth in bile, there is an apparent increase in transposon insertions in *skp*. However, these insertions are all oriented in one direction, indicating that *skp* is dispensable, and that expression from transposon insertions might somehow affect downstream *lpxD* expression in an advantageous way in the presence of bile. Therefore, while *skp* is definitely dispensable during growth in bile, the increased fitness associated with insertion in the *skp* gene may be the result of polar effects on neighbouring genes.

4.1.3.3. lpxM, kdsD, kdsC, lptC

Several genes with the most positive logFC values are involved in aspects of LPS biosynthesis and transport, including *lpxM, lptC, kdsD,* and *kdsC*. What is noticeable about these genes is the sheer number of new detectable insertions in these genes in the output library compared to the input, particularly in *lpxM* and *kdsD* (Figure 4.5A). This suggests that there were mutants in these genes in the input library, but these mutants were not detected during sequencing. However, the relative abundance of these mutations increased following growth in bile.

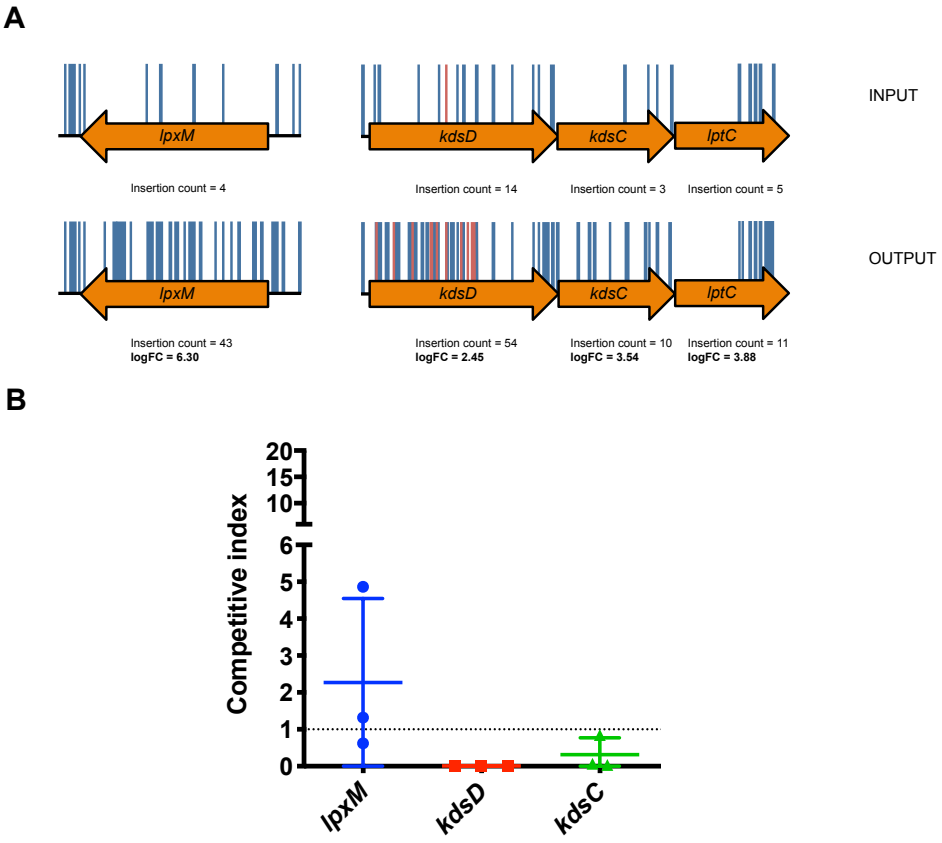


Figure 4.5. (A) Transposon insertion plots of *lpxM, kdsD, kdsC,* and *lptC*. Blue lines indicate transpon insertions (length not scaled to readcounts) in the input (top row) and output (bottom row). For *kdsD, kdsC,* and *lptC*, the orientation of transposon insertions is highlighted by blue (plus strand) and red lines (minus strand). LogFC values from TraDIS indicated. **(B)** Competition assay of *lpxM, kdsD,* and *kdsC* Keio whole-gene deletion mutants. The dotted line indicates a competitive index of 1, equivalent to the competitiveness of the wild type.

lpxM (Figure 4.5A) contained insertions in both orientations and throughout the gene (not shown), but Δ*lpxM* mutants showed a range of competitive indices, from 0.6 to 4.8 (Figure 4.5B). *lpxM* encodes a myristoyl-acyl carrier protein (ACP)-dependent acyltransferase, which transfers myristoyl-ACP to 2-keto-3-deoxy-octulosonate $(KDO)_2$-(lauroyl)-lipid IV_A. It has been reported that *lpxM* mutants display elevated σ^E levels and pentaacylated LPS, as opposed to hexaacylated LPS found in wild-type cells (Tam & Missiakas, 2005). The loss of *lpxM* in *Klebsiella pneumoniae* has been shown to cause increased sensitivity to bile (Clements *et al.*, 2007), however in *E. coli* a Δ*lpxM* knockout mutant was resistant to four times more deoxycholate than the wild type; therefore, it is likely the TraDIS is reflecting this phenotype (Karow & Georgopoulos, 1992).

The genes, *kdsC* and *kdsD*, encode 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase and rabinose-5-phosphate isomerase, respectively. Together, these enzymes participate in the generation of KDO, a component of LPS. *kdsD* and *kdsC* mutants have been shown not to display defects in LPS biosynthesis or growth under standard laboratory conditions due to the presence of isozymes or other compensatory mechanisms (Meredith & Woodard, 2005; Sperandeo *et al.*, 2009). In this study, transposon insertions in *kdsC* and *kdsD* are either mostly or entirely inserted in one orientation, suggesting that these insertions may have advantageous polar effects on nearby genes e.g. the essential *lptC, lptA,* and/or *lptB*.

The *lptC* gene is transcribed with *kdsD* and *kdsC*, and encodes a component of the LPS transport system. Notably, the increase in insertions is localised to the C terminus-encoding region (Figure 4.5A (Sperandeo *et al.*, 2009)). A Δ*lptC* mutant was not available in the Keio collection and therefore this gene is considered to be essential for growth in *E. coli* (Baba *et al.*, 2006; Sperandeo *et al.*, 2009). However, the C terminal region of LptC is implicated in binding to LptA, the periplasmic component of the LPS transport apparatus, and it has previously been shown that mutations within this region result in a reduction of levels of this essential periplasmic component of the Lpt system (see Figure 4.6; (Sperandeo *et al.*, 2011). It is interesting to note that DsbA and Skp also play roles in the assembly of LptD (Ruiz *et al.*, 2010; Sperandeo *et al.*, 2009), another component of the Lpt system, and that mutations in *dsbA* and *skp* are also advantageous in the presence of bile. This indicates a possible effect of bile on the requirements for the Lpt transport system which provides a fitness advantage to these mutants. However, it is also possible that the insertions within *lptC* provide a fitness advantage by increasing the expression of the downstream essential gene, *lptA*, since all transposons in *lptC* were inserted in one orientation only (Figure 4.5A).

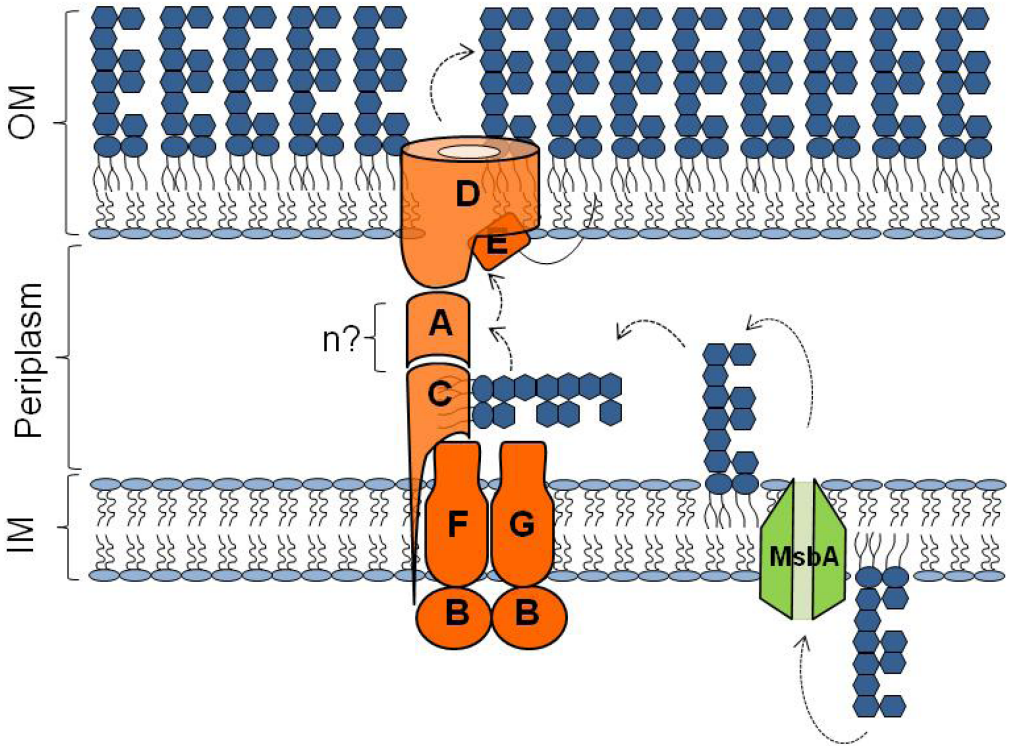


Figure 4.6. LPS transport in *E. coli*. LPS is flipped across the inner membrane by the MsbA protein before being exported to the cell surface by the Lpt machinery (highlighted in orange). MsbA (highlighted in green) flips the lipid A-core molecule across the inner membrane. LptBGF, an ATP binding cassette transporter, together with LptC facilitate the release of LPS from the inner membrane. The periplasmic protein, LptA, acts as a bridge between the inner and outer membranes. The outer membrane complex LptDE assembles LPS at the cell surface. Figure adapted from (Martorana *et al.*, 2014).

4.2. Comparison of end-point growth and logFCs

Previously, results from TraDIS were compared to screens of defined mutant libraries in order to validate the list of essential genes obtained in TraDIS (see Chapter 3 of this thesis). This is because logFC values obtained from a TraDIS experiment do not provide information as to why a mutant was less fit under an experimental condition e.g. is it unviable? Is it viable but has a lower growth rate? Does it interact with other mutants in the pooled population in a manner detrimental to growth? Such information could be inferred by comparing logFC values with the end-point growth (i.e. OD of mutant cultures) of mutants after growth in the presence of bile, since final cell density can provide a simple measure of the ability of a mutant to grow in the absence of other competing mutants. Therefore, a selection of single-gene knockout mutants from the Keio library was grown in the presence of bile and compared to results of the TraDIS screen.

4.2.1. Growth assay results

214 mutants of the Keio library whose corresponding genes displayed significant logFC values in TraDIS, were assayed in triplicate for end-point growth in LB supplemented with 10% bile, following initial growth in LB with 2% bile. This passaging in bile was done to ensure comparability of results with the TraDIS assay, which was screened under similar conditions (Langridge *et al.*, 2009). A difference in growth was deemed statistically significant if final OD_{595} absorbance readings were two standard deviations above or below the mean absorbance measurement

of all mutant cultures together (see section 2.6.1, Materials and Methods). Following growth in bile, 27 mutants showed significantly decreased growth and 36 mutants showed significantly increased growth; the remaining 151 mutants showed no significant difference in growth (Figure 4.7). Within the Keio mutants showing decreased growth in bile, 23 genes had a negative logFC and 4 genes has a positive logFC after TraDIS analysis. Similarly, within the Keio mutants showing increased growth, 24 genes were associated with a negative logFC and 12 genes were associated with a positive logFC following TraDIS analysis. Finally, within the Keio mutants that showed no differences in growth, 90 genes were associated with a negative logFC and 61 genes were associated with a positive logFC following TraDIS analysis. Therefore, following this preliminary analysis, there does not appear to be a very good correlation between TraDIS and end-point growth assays. Indeed, only 35 genes (16.4%) showed agreement between TraDIS and Keio with respect to growth in bile. Genes were then categorised based on the correlation between growth value and logFC (see Table A9, Appendix).

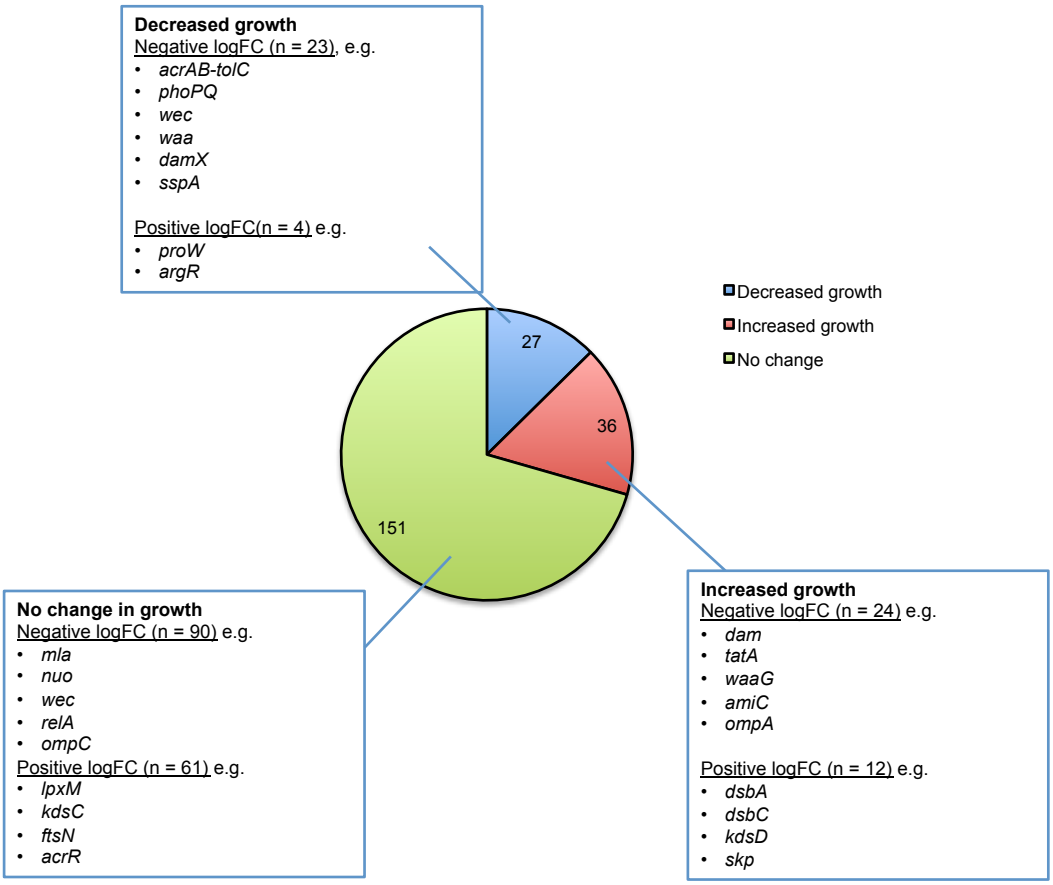


Figure 4.7. Keio library mutants whose genes displayed significant logFC values in TraDIS, grouped by whether their end-point growth was significantly decreased (blue), increased (red), or not changed (green). A comparison to logFC values is shown for each category in the blue boxes.

4.2.2. Negative logFC and reduced growth

This category of genes (n = 23; see Table 4.4) represents those whose mutants are most sensitive to the deleterious effects of bile, exhibiting both reduced growth and fitness. Many of these functions include those which have been mentioned in this chapter, with well defined roles in bile tolerance and resistance. These genes included *acrAB-tolC*, *phoPQ, wec*

(rff), waa, damX, and *rpoS* (Begley *et al.*, 2005). Other genes in this list, including *mltG, nlp, fimE, opgH,* and *tatC*, do not have well characterised roles in bile tolerance, yet their presence in this list indicates that they are probably required for optimal growth and fitness in bile.

Gene	logFC	% reduction in growth	Gene	logFC	% reduction in growth
acrA	-13.457	32.6	*rpoS*	-6.148	13.9
phoQ	-11.490	11.8	*wecB*	-6.065	35.3
tolC	-11.047	17.4	*fimE*	-4.742	14.6
phoP	-10.992	13.8	*wecC*	-4.075	30.8
waaQ	-10.536	26.9	*opgH*	-3.467	24.7
acrB	-10.456	23.6	*cvpA*	-3.088	17.6
wecD	-8.995	23.8	*yceG (mltG)*	-2.934	18.4
nlpD	-8.597	15.6	*nlpI*	-2.692	24.4
lapA	-8.385	21.7	*yajG*	-1.695	14.7
tatC	-8.317	20.8	*lpoA*	-1.519	17.7
sspA	-7.338	33.3	*rlmL*	-0.555	13.8
damX	-6.295	14.2			

Table 4.4. Keio library mutants listed by increasing logFC value. Percentage reduction in growth indicated. Reduction in growth is calculated based on the percentage difference between the average mutant end-point OD_{595} value and the average end-point OD_{595} value of all other mutants assayed on the same well plate (see Materials and Methods for more detail).

4.2.3. Positive logFC and increased growth

This list of genes (n = 12; see Table 4.5) included those whose mutants displayed an increased cell density and increased competitiveness in the presence of bile. This list also included some of the genes discussed previously, such as *dsbA,* and *skp*, further confirming that mutants in these genes are advantageous for growth in the presence of bile. Moreover, the Δ*dsbB* mutant, which previously was shown not to display increased CI in a competition assay, showed increased growth (Figure 4.3B). Similarly, the Δ*kdsD* mutant, which showed a substantial reduction in competitiveness compared to the wild type (Figure 4.5B) also showed increased growth, as discussed above. This would suggest that Δ*dsbB* and Δ*kdsD* mutants are capable of growing to high cell densities but probably at a slower growth rate, but also that competitive pressures were different in TraDIS and competition assays.

Gene	**logFC**	**% increase in growth**	**Gene**	**logFC**	**% increase in growth**
dsbA	3.154	49.4	*kgtP*	1.180	44.9
skp	2.510	38.0	*ebgR*	1.151	13.4
kdsD	2.458	15.2	*rsmG*	0.979	16.4
dsbB	2.188	23.2	*uxuB*	0.901	47.0
clsA	1.697	16.0	*bacA*	0.884	39.5
secG	1.531	22.0	*eptB*	0.852	18.4

Table 4.5. Keio library mutants (listed by gene) that showed increased end-point growth alongside their equivalent logFC values.

4.2.4. Mutants with no significant growth differences

The largest group of genes in this list encompassed those mutants with decreased logFC values but no significant differences in end-point growth of their Keio library equivalent mutants (n = 90). In other words, these mutants could grow to sufficiently high cell densities, but were negatively affected by the presence of competing populations of mutants. This phenotype is most likely explained by these mutants having reduced growth rates in the presence of bile, which would not necessarily limit their ability to reach a relatively high final OD_{595}, but would put them at a fitness disadvantage under competition. Notable genes or operons within this list include *mla, nuo, waa, wec, relA,* and *ompC* (see Table A9).

In addition, within this group are mutants that showed a fitness advantage (i.e. positive logFC values in TraDIS) but no significant changes in growth (n = 61). It is possible that these mutants displayed a fitness advantage because they were more resilient to the deleterious effects of bile, as opposed to being able to grow to greater numbers within the mutant pool. Again, the apparent discrepancies between TraDIS and end-point growth assays may reflect differences in assay conditions or else subtle changes between phenotypes of whole-gene deletion and transposon insertion mutants. Indeed, the non-congruence of outputs from TraDIS (i.e. logFC), end-point growth, and CI values indicates the need to employ a variety of different approaches in order to fully describe the role of a gene under a specified selective growth condition.

4.3. Comparison to S. enterica TraDIS

As mentioned in the introduction, TraDIS has been used previously to identify the genetic requirements for growth for *S. enterica* Typhi growth in the presence of bile (Langridge *et al.*, 2009). Therefore, in order to understand broader genetic requirements within the Enterobacteriaceae, a comparative analysis between *E. coli* and *S. enterica* TraDIS studies was conducted.

In *S.* Typhi*,* 168 genes displayed significant reductions in insertion count following growth in bile, and these were previously categorised into four main classes: energy metabolism, membrane/surface structures, degradation of macromolecules, and central/intermediary metabolism (Langridge *et al.*, 2009). It was not indicated, however, which genes displayed increased insertions following growth in bile, therefore the list of 168 genes was compared with genes displaying negative logFC values in *E. coli* (n = 137) Of these, 42 genes were shared with *E. coli* (Figure 4.8).

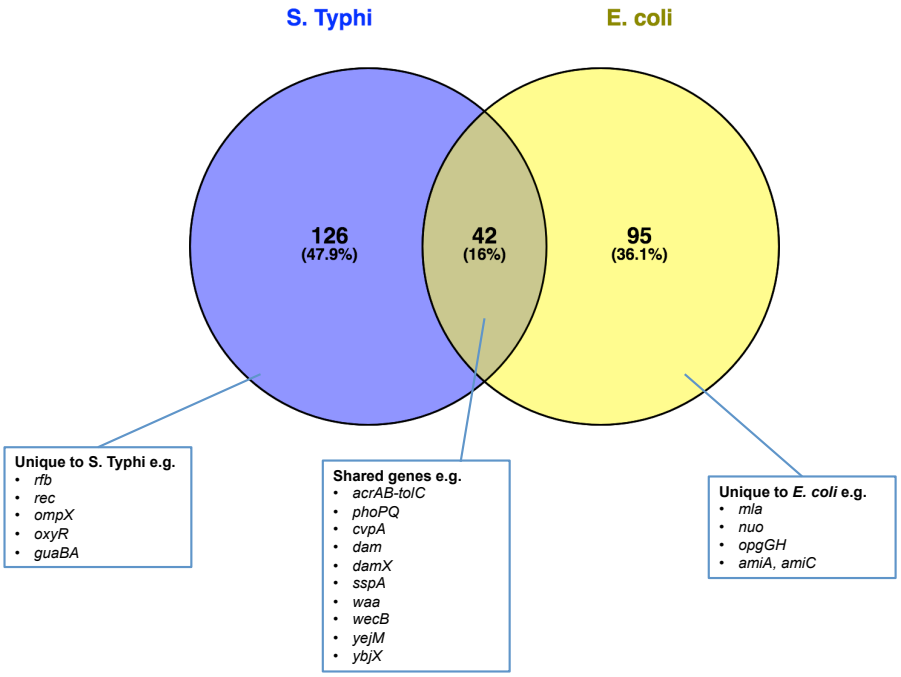


Figure 4.8. Venn diagram comparing genes required for growth in bile in *S. Typhi* and *E. coli* MG1655 (Langridge *et al*., 2009 and this study). Example genes in each category are highlighted. Full gene lists can be found in Table 4.6 (shared genes), and Tables A10 and A11, Appendix (unique genes).

4.3.1. Genes shared between S. Typhi and E. coli

Most genes that were shared in both studies have been mentioned previously in this chapter and include the major bile resistance determinants or genes shown to play a role in optimal growth in bile, i.e. *acrAB, tolC, phoPQ, waa* genes, *wec* genes, *damX*, *dam,* and *rob* (encoding a response regulator which coordinates the response to organic solvents such as bile acids; see Table 4.6).

Gene	Function
acrA	multidrug efflux system
acrB	multidrug efflux system protein
ampG	muropeptide transporter
cpxR	response regulator in two-component regulatory system with CpxA
cvpA	colicin V production protein
dacA	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 5)
dam	DNA adenine methyltransferase
damX	cell division protein that binds to the septal ring
dedA	DedA family inner membrane protein
dsdA	D-serine dehydratase
ftsP	septal ring component that protects the divisome from stress; multicopy suppressor of ftsI(Ts)
galR	galactose-inducible d-galactose regulon transcriptional repressor; autorepressor
mlaA	ABC transporter maintaining OM lipid asymmetry, OM lipoprotein component
mltA	membrane-bound lytic murein transglycosylase A

ompA	outer membrane protein A (3a;II*;G;d)
ompC	outer membrane porin protein C
phoP	response regulator in two-component regulatory system with PhoQ
phoQ	sensory histidine kinase in two-component regulatory system with PhoP
prc	carboxy-terminal protease for penicillin-binding protein 3
ptsP	PEP-protein phosphotransferase enzyme I; GAF domain containing protein
relA	(p)ppGpp synthetase I/GTP pyrophosphokinase
rfbD	dTDP-L-rhamnose synthase, NAD(P)-dependent dTDP-4-dehydrorhamnose reductase subunit
rob	right oriC-binding transcriptional activator, AraC family
sanA	DUF218 superfamily vancomycin high temperature exclusion protein
sspA	stringent starvation protein A, phage P1 late gene activator, RNAP-associated acid-resistance protein, inactive glutathione S-transferase homolog
tig	peptidyl-prolyl cis/trans isomerase (trigger factor)
tolC	transport channel
uvrD	DNA-dependent ATPase I and helicase II
waaB	lipopolysaccharide 1,6-galactosyltransferase; UDP-D-galactose:(glucosyl)lipopolysaccharide-1, 6-D-galactosyltransferase
waaG	UDP-glucose:(heptosyl)lipopolysaccharide alpha-1,3-glucosyltransferase; lipopolysaccharide core biosynthesis protein; lipopolysaccharide glucosyltransferase I
waaL	O-antigen ligase
waaO	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase

waaP	kinase that phosphorylates core heptose of lipopolysaccharide
waaQ	lipopolysaccharide core biosynthesis protein
waaR	lipopolysaccharide 1,2-glucosyltransferase; UDP-glucose:(glucosyl)LPS alpha-1,2-glucosyltransferase
waaU	lipopolysaccharide core biosynthesis
waaY	lipopolysaccharide core biosynthesis protein
wecA	UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase
wecB	UDP-N-acetyl glucosamine-2-epimerase
wecC	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase
ybjX	DUF535 family protein
yejM	essential inner membrane DUF3413 domain-containing protein; lipid A production and membrane permeability factor

Table 4.6. Genes under negative selection in both *S.* Typhi and *E. coli* TraDIS studies.

Many other genes encode proteins with roles related to the functioning of the cell envelope, i.e. *ampG, cpxR, dacA, dedA, mltA, ompA, ompC, prc, sanA,* as well as the putative cell envelope-associated genes *ybjX* and *yejM* (Table 4.6). The presence of *yejM* (encoding a putative cardiolipin transport protein) is interesting, since deletion of the *yejM* gene has been reported to be lethal (De Lay & Cronan, 2008). Cardiolipins contribute to several activities within the bacterial membrane, including the PhoPQ-mediated remodeling of the cell membrane during infection and cytoskeletal arrangement during cell division in *E. coli* (Dalebroux *et al.*, 2015; Renner & Weibel, 2012).

The remaining genes shared by both studies are predicted to encode proteins involved in a variety of different functions, including *cvpA* (colicin production), *dsdA* (D-serine ammonia-lyase, involved in serine degradation), *tig* (trigger factor, required for folding of cytosolic proteins), and *relA* and *sspA* (stringent response; see Figure 4.9).

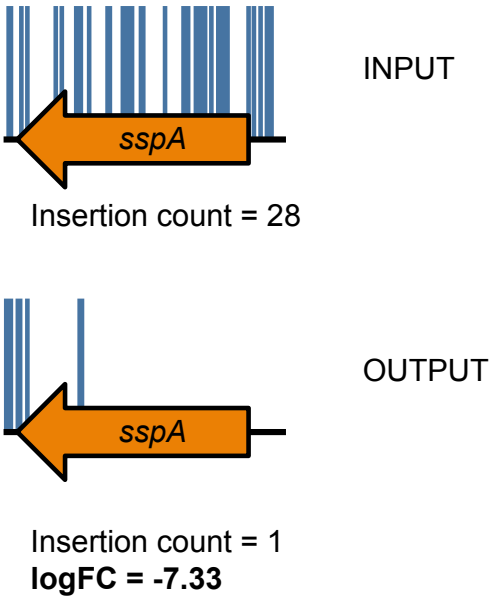


Figure 4.9. Transposon insertion plots of *sspA* from TraDIS conducted in this study. Blue lines indicate transposon insertions (length not scaled to readcounts) in the input (top row) and output (bottom row). Insertion counts and logFC value indicated.

The selection for both *relA* and *sspA* strongly implies a role for the stringent response in bile tolerance. The *relA* gene encodes GDP/GTP pyrophosphokinase, an enzyme responsible for producing the global regulatory molecules ppGpp and pppGpp (referred to collectively as (p)ppGpp). The primary role of RelA is the production of (p)ppGpp in response to amino acid starvation i.e. the stringent response. RelA associates with the ribosome and promotes the synthesis of (p)ppGpp when an uncharged tRNA binds to the acceptor site of the ribosome. The accumulation of ppGpp then results in the cessation of rRNA and tRNA synthesis (Brown *et al.*, 2016). One study has shown a potential role for RelA in adaptation to bile salts in *Enterococcus faecalis*, in modulating the virulence properties of the bacterium such that it can more successfully proliferate in an oxidative environment (including in the presence of bile salts; (Yan *et al.*, 2009)). SspA also associates with RNA polymerase and plays an important role in stress responses during stationary phase via inhibition of the accumulation of H-NS, in particular being essential for growth under acid stress (Hansen *et al.*, 2005). The potential role of SspA in bile tolerance is less clear. SspA can account for over 50% of total protein synthesis during amino acid starvation in *E. coli* (O′Farrell, 1978) and it has been shown that the abundance of at least 11 proteins are altered in Δ*sspA* mutants, (Williams *et al.*, 1994). However, most of these 11 proteins were not identified, and of those that were, none were identified in either the TraDIS or Keio mutant screens (Williams *et al.*, 1994). This study did show, however, that Δ*sspA* mutants displayed reduced viability during long-term stationary phase, indicating a

role for the protein in withstanding stress. Therefore, the stringent response may have an important role to play in bile tolerance in *S.* Typhi and *E. coli* by coordinating stress responses, and particularly oxidative stress.

4.3.2. Species-specific differences

The comparative analysis indicated that a wide variety of functions play a role in bile tolerance in both species but is remarkable that there are far more genes unique to each species compared to the number of genes shared by both (see Figure 4.9, Table 4.6, Table A10, Table A11). However, when genes were arranged into functional groups, it appeared that *S.* Typhi and *E. coli* shared many functional groups without necessarily sharing the exact same genes (Table 4.7).

Functional group	**Genes unique to *S.* Typhi**	**Genes unique to *E. coli***
Cell envelope constituents	*clsA, cpxA, fabR, ompX, rcsC, rcsD, rffM, rseB, tamA, tamB, yrfF*	*asmA, bepA, clsA, fabR, mlaB, mlaC, mlaD, mlaE, mlaF, skp, yajG, yedD*
LPS	*pagP, rffM, pgi, rfbA, rfbB, rfbC, rfbE, rfbH, rfbI, rfbK, rfbP, rfbU, wbbH, wzzB*	*lpxT, rfaG, waaZ*
PG synthesis/Cell division	*ldtB, mltC, mltD, mrcA, mrcB, slt, zapB*	*envC, minC, nlpD, nlpI, tolA, ygeR, yraP, emtA, mipA, amiA, amiB, amiC, elyC*
DNA/RNA	*hfq, seqA, yejK, hcpA, sbcB, xseA, pcnB, topB, yejH, pnp, rph, rapA, rmuC, recD, recG, recJ*	*rnhA, dcm, uup*
Chaperones/protein folding	*djlA, dsbD, fkpA, hslU, ridA*	

Table 4.7. A comparison between some gene functional groups in TraDIS studies of *S.* Typhi and *E. coli*. Note that only genes under negative selection are included.

S. Typhi did appear to select for many more genes involved in nucleotide metabolism and the modification and maintenance of DNA, suggesting that bile salt-induced DNA damage may have been under greater selection in that TraDIS study (Langridge *et al.*, 2009). Furthermore, while mutations in many protein chaperones appeared to offer a fitness advantage in *E. coli*, many genes encoding protein folding chaperones were under negative selection in *S.* Typhi (see Table 4.7). Together, these suggest either differences in the redox environment in which *E. coli* and *S.* Typhi TraDIS libraries were grown or may indicate that both species react differently to the challenge imposed by bile. This can be seen in the negative selection for *rcsD, fkpA, ldtB, tgt, oxyR,* and *clsA* in *S.* Typhi but these genes showing positive logFCs in *E. coli*. There was also a notable difference in the selection for LPS-associated genes between the two species (see Table 4.7), which could have had a significant impact on the structure of the cell envelope. In particular, genes involved in the production of O antigen subunits (i.e. *wzzB,* 12 *rfb* genes and *pgi*) were selected for uniquely in *S.* Typhi (Langridge *et al.*, 2009), whereas MG1655 contains rough LPS i.e. its LPS molecules do not contain an O-specific polysaccharide, and so these genes were not under selection. This is significant, since the presence of the O antigen is known to influence bile resistance (Crawford *et al.*, 2012). Therefore, LPS structure may also have influenced the differences in genetic requirements between the two species. Nevertheless, it would appear that *E. coli* and *S.* Typhi share many of the same functional requirements for bile tolerance.

Conclusions

This study described the use of TraDIS to determine the genetic requirements for growth of *E. coli* in the presence of bile. In total, 214 genes were shown to exert a significant impact upon fitness. The majority of these genes displayed negative logFCs and included many whose roles in bile tolerance had been established previously (Begley *et al.*, 2005). However, TraDIS also identified several genes and functions that had no obvious prior link to bile tolerance, e.g. NADH dehydrogenase and a selection of y genes. Therefore, this study has expanded upon the repertoire of candidate genes involved in bile tolerance in *E. coli*. TraDIS also identified a substantial number of genes displaying positive logFCs, e.g. *dsbA* and *skp,* indicating that a trade-off exists between *E. coli* increasing its growth in the presence of bile and maintaining the function of certain cellular components.

This chapter also described a comparison between this TraDIS study with a similar study conducted in *S.* Typhi (Langridge *et al.*, 2009) in order to characterise bile tolerance in the Enterobacteriaceae. This showed that both species share a core set of genes with well established links to bile tolerance, but also that these species do not share many genes in common. However, comparing gene functional groups demonstrated that the requirements of both species to growth in bile share similar functionalities, despite differences in the requirements for specific genes. These differences may have arisen as a result of a range of different experimental conditions, including species differences, but the

absence of an O antigen in *E. coli* MG1655 may also have had a substantial effect. Future comparative studies should employ various Enterobacterial strains with both smooth and rough LPS in order to understand the effect of the O antigen (and species) on bile tolerance in greater detail.

Chapter 5 TraDIS analysis of genetic requirements of *E. coli* growth under anaerobic conditions in the presence of nitrate

Introduction

Anaerobic nitrate respiration is an important fitness determinant during *E. coli* colonisation of the GI tract, enhancing its metabolic capacity and boosting its growth within the oxygen-poor environments of the intestine. This is supported by the fact that *E. coli* unable to respire nitrate exhibit extreme colonisation defects, and that, under the fluctuating redox conditions of gut inflammation during which nitrate is produced, *E. coli* can boost its growth relative to other members of the GI microbiota (Faber & Bäumler, 2014; Jones *et al.*, 2011; Winter *et al.*, 2013). Therefore, understanding the mechanisms of nitrate respiration in *E. coli* is important when characterising the processes at play in gut health and disease.

In the absence of oxygen and in combination with FNR, IHF, and/or other regulators, genes involved in anaerobic nitrate respiration are expressed (Cole & Richardson, 2013). Nitrate respiration fundamentally is the reduction of nitrate to ammonia via an electron transport chain. This chain is made up of a dehydrogenase, which accepts electrons from electron donors such as NADH or formate, a quinone to act as a redox carrier, and a terminal nitrate reductase, which reduces nitrate to nitrite (Figure 5.1B). Together, this generates a proton motive force, allowing for the production of ATP via the F_1F_0 ATPase. There is metabolic flexibility for this process in *E. coli* through the presence of three nitrate reductases, membrane-bound nitrate reductase A (NarGHI) and nitrate reductase Z (NarZYV), and the periplasmic nitrate

reductase, Nap. Each operates based on the concentration of nitrate present in the environment, with nitrate reductase A operating at higher concentrations than nitrate reductase Z or Nap (Cole & Richardson, 2013). Based on the variety of dehydrogenases and nitrate reductases available in *E. coli,* 5 potential variants of the nitrate electron transport chain exist, although in reality only formate dehydrogenase and NADH dehydrogenase I are thought to be able to operate in concert with *E. coli*'s nitrate reductases (Cole & Richardson, 2013).

The molecular mechanisms of the regulation of nitrate respiration and the reduction of nitrate to nitrate and ammonia have been well characterised (Cole & Richardson, 2013; Unden & Bongaerts, 1997). *E. coli* expresses two nitrate-sensing histidine kinases, NarX and NarQ, which are active in the presence of high (micromolar - millimolar) and low (as low as nanomolar) concentrations of nitrate in the environment, respectively (Cole & Richardson, 2013). When activated, they phosphorylate their cognate response regulators, NarL and NarP, respectively, although there is extensive cross-regulation between the two component systems (Figure 5.1A) (Unden & Bongaerts, 1997).

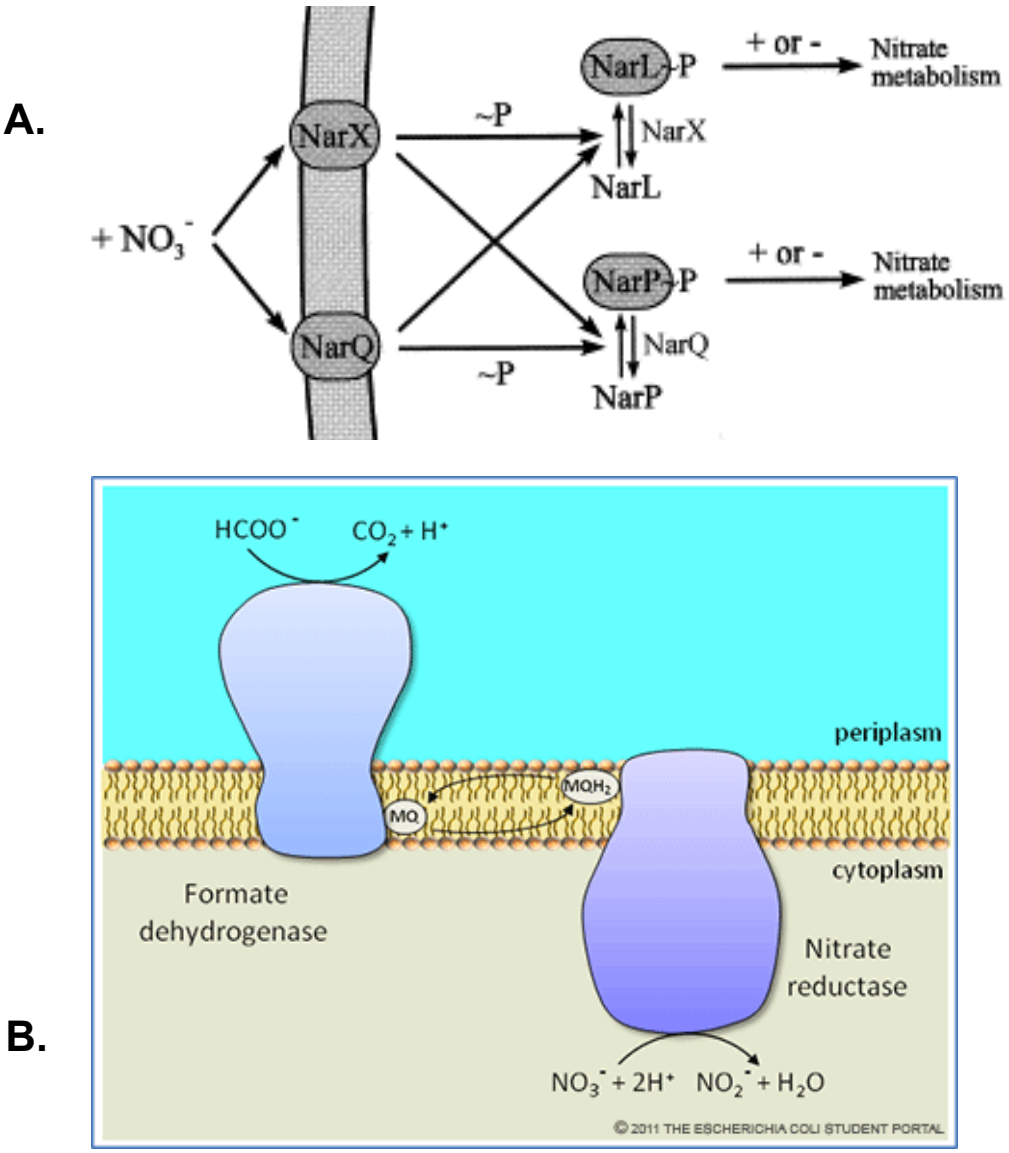


Figure 5.1. (A) Nitrate-sensing two component systems in *E. coli*. Adapted from Unden & Bongaerts, 1997. **(B)** Formate dehydrogenase-to-nitrate reductase electron transfer. From http://ecolistudentportal.org/article_anaerobic_respiration; accessed 23 Jan 2019.

Studies that characterise the global response of the cell to nitrate have focused on analysing the regulons of the major nitrate response regulators i.e. NarP or NarL. For example, transcriptional analysis has showed that NarL is required to induce the expression of over 40 genes in response to nitrate (Constantinidou *et al.*, 2006). While most of these were directly involved in the nitrate respiratory apparatus or the detoxification of byproducts such as NO or nitrate, some included previously uncharacterised functions, such as the unannotated genes *yeaR* and *yoaG* (Constantinidou *et al.*, 2006; Goh *et al.*, 2005; Stewart, 1993). Furthermore, metabolomic analysis has shown that nitrate significantly shifts metabolic flux through the TCA cycle and the pentose phosphate pathway (Toya *et al.*, 2012). Studies on the global effects of nitrate respiration on *E. coli* are, however, lacking.

This study therefore aims to conduct an analysis of the genetic requirements for the growth of *E.coli* under anaerobic conditions in the presence of nitrate using TraDIS. The first section of the analysis is an in-depth look at the effect of nitrate on central metabolism. This is followed by an overview of the requirements for growth under these conditions outside of central metabolism. Results presented in this chapter demonstrate the power of TraDIS as a global analytical method, identifying potential novel functions for both well characterised, and previously uncharacterised, genes.

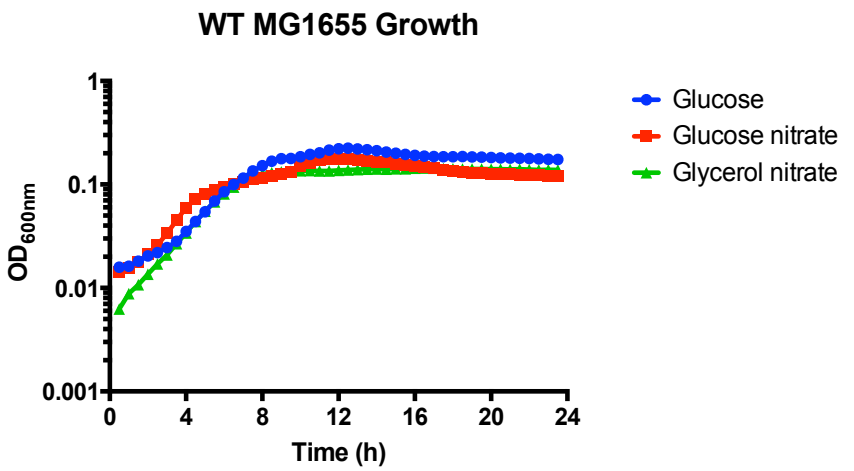
Results & Discussion

5.1. Mutant library screen

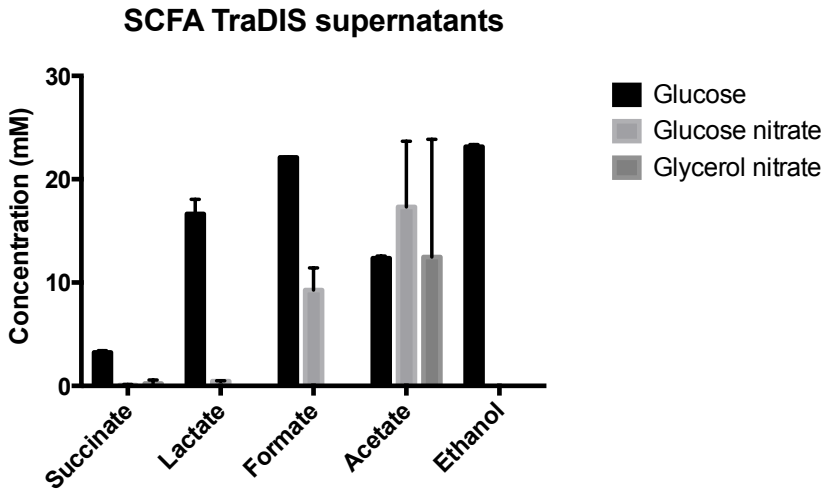
The TraDIS transposon library described in Chapter 3 was initially cultured in M9 minimal medium under aerobic conditions with 0.4% (w/v) glucose as the carbon source. This input library was then passaged three times anaerobically in M9 minimal medium supplemented with glucose only ('glucose' → expected fermentative conditions), glucose + 40 mM sodium nitrate ('glucose nitrate' → expected nitrate respiration conditions) and glycerol + 40 mM sodium nitrate ('glycerol nitrate' → expected nitrate respiration conditions). It should be noted that *E. coli* cannot ferment glycerol and anaerobic growth using glycerol requires the presence of a terminal electron acceptor. Results from TraDIS are dictated by the selective pressures present within the medium in which the mutant library is passaged. Therefore, in order to characterise the growth dynamics and selective pressures present within each growth condition, OD_{600} values, pH readings, total number of generations, and metabolite profiles of cell-free supernatants were measured for both wild-type and pooled mutant library cultures (shown in Table 5.1 and Figure 5.2). Overall, cultures were grown for approx. 20 - 21 generations in glucose, approx. 19 generations in glucose nitrate, and approx. 16 - 19 generations in glycerol nitrate. Growth curves of MG1655 cells cultured under the three conditions showed an initial lag phase in glucose and glucose nitrate, but not in glycerol nitrate (Figure 5.2A). Thereafter, glucose nitrate and

glycerol nitrate appear to display similar growth rates during log phase, however in glucose and glucose nitrate, a form of diauxie emerges, with a flattening of the curve between 4 h and 8 h for glucose, or 8 h - 12 h for glucose nitrate. This may indicate some switch in transition in growth mode such as during the glucose-to-acetate transition (Enjalbert *et al.*, 2015). Interestingly, fermentative growth using glucose resulted in the highest OD_{600} value after 23 h (Figure 5.2A). However, pairwise comparisons of total growth curves indicated no significant differences in total growth between the three conditions. Therefore, while growth dynamics were different between the three different conditions, they did not ultimately show significant differences in overall bacterial growth.

A.


WT MG1655 Growth
OD600nm
1
0.1
0.01
0.001
0
4
8
12
16
20
24
Time (h)
Glucose
Glucose nitrate
Glycerol nitrate

B.


SCFA TraDIS supernatants
Concentration (mM)
30
20
10
0
Succinate
Lactate
Formate
Acetate
Ethanol
Glucose
Glucose nitrate
Glycerol nitrate

C.

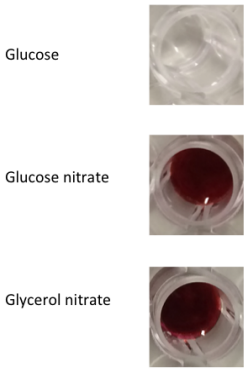

Glucose
Glucose nitrate
Glycerol nitrate

Figure 5.2. Analysis of TraDIS culture conditions. **(A)** Growth curves of WT MG1655 grown in triplicate under anaerobic conditions at 37°C in M9 minimal medium supplemented with 0.4% (w/v) of carbon source and 40 mM nitrate, where indicated. **(B)** Metabolite profiles of cell-free supernatants of TraDIS libraries grown anaerobically on glucose (black bars), glucose nitrate (light grey bars), and glycerol nitrate (dark grey bars). Metabolite levels were measured by HPLC and measurements represent duplicate output library cultures. **(C)** Greiss method for detection of nitrate reductase activity from wild-type MG1655 cultures grown under the different growth conditions, as indicated. A red colour indicates the qualitative presence of nitrites and therefore nitrate reductase activity.

Sample	Passage	OD_{600} Rep1	OD_{600} Rep2	No. generations Rep1	No. generations Rep2	pH Rep1	pH Rep2
Input (M9 glucose aerobic)		2.51	2.18	8	8		
Glucose	1	1.09	0.76	7	6		
	2	1.1	0.95	7	7		
	3 (output)	1.09	0.972	7	7	4.72	4.74
Glucose nitrate	1	0.88	0.87	6	6		
	2	1.23	0.84	7	6		
	3 (output)	0.86	0.999	6	7	6.77	6.72
Glycerol nitrate	1	0.68	0.46	6	5		
	2	1.01	0.63	7	6		
	3 (output)	0.62	0.62	6	6	7.00	7.00

Table 5.1. Optical densities (OD_{600nm}), number of generations, and cell-free supernatant pH values of TraDIS mutant libraries grown in duplicate under the conditions indicated. Number of generations were calculated as the number of times a culture doubled from the starting OD_{600} of 0.01 to the indicated end point OD_{600}. Each passage of growth represents 24h at 37°C.

The presence of specific terminal end-products of metabolism in the cell-free supernatants at the end of the third passage of each growth condition was determined by HPLC (Figure 5.2B). This data showed substantial amounts of ethanol, lactate, formate and succinate produced in the glucose culture, indicative of mixed-acid fermentation. Ethanol, succinate and lactate were not detected in cultures grown with glucose nitrate and glycerol nitrate suggesting that significant levels of fermentation did not occur under these conditions. As expected, pH measurements also indicated increased acidity after growth with glucose compared to glucose nitrate, presumably due to the greater production of acidic byproducts such as acetate (Figure 5.2B, Table 5.1). Finally, the Griess method was used to detect the presence of nitrites in cell-free culture supernatants (Figure 5.2C). Nitrite production is indicative of the reduction of nitrate during anaerobic respiration (Griess & Bemerkungen, 1879). Therefore, as expected, no nitrite production was detected from *E. coli* growing under anaerobic conditions in the presence of glucose. In contrast, nitrites were detected in cell-free supernatants of *E. coli* grown with either glucose or glycerol in the presence of nitrate. Therefore, it appears that the primary metabolic mode of cells grown anaerobically on glucose is mixed-acid fermentation and the addition of nitrate results in a metabolic switch from mixed-acid fermentation to anaerobic respiration.

5.2. TraDIS

5.2.1 Sequencing results

TraDIS sequencing was carried out by the Wellcome Trust Sanger Institute. All subsequent analysis was carried out during this thesis. The reads from the MiSeq run were processed and mapped to the *E. coli* str. K-12 MG1655 reference genome (U00096). An average of 1.4 million reads was generated for each library, with at least 92% of reads mapping to the genome. Mapping statistics are shown in Table 5.2.

5.2.2. Essential genes and logFCs

TraDIS can readily provide lists of essential genes based on insertion indices. However, TraDIS can also be used to generate logFC values for genes when comparing readcounts between an input and output library. In this chapter, both insertion index and logFC were considered when analysing the results of TraDIS. This was done for the following reasons: 1. Selecting genes based on essential gene criteria (i.e. below the lowest calculated insertion index threshold) is straightforward and allows easy comparisons with other studies. e.g. 50 essential genes from growth on glucose, 50 from growth in glucose nitrate, and 46 from growth in glycerol nitrate were shared with a previously published list of genes essential for growth in M9 medium (Joyce *et al.*, 2006).

Sample	Total Reads	% Mapped	Unique Insertion Sites (UIS)	Sequence Length (bp)/UIS
INPUT 1	1380341	94.0584553	237355	19.54740789
INPUT 2	1358614	92.75704596	197456	23.49726015
Glucose 1	1383167	94.09060022	209089	22.1899526
Glucose 2	1394410	94.74634572	211883	21.89734429
Glucose_nitrate 1	1238074	92.69745641	110818	41.86752152
Glucose_nitrate 2	1515866	92.60251621	118058	39.29996273
Glycerol_nitrate 1	1594990	93.38922826	111960	41.44046981
Glycerol_nitrate 2	1550845	93.32456563	211923	21.89321121

Table 5.2. Sequencing statistics of each TraDIS library in duplicate. Reads mapped to Genbank accession number U00096.

2. Comparing logFC values can identify genes with significant roles in fitness under certain conditions despite the gene being non-essential e.g. *sucC* is essential in glucose but not essential in glucose nitrate but does display a logFC value of -1.95 in glucose nitrate. 3. Using logFCs alone to compare genetic requirements for growth can, in some instances, falsely categorise genes as being required more under one condition than another because the total sequencing reads differed between growth conditions and between output libraries (Table 5.2). For example, *pgi* has a logFC value of -6.8 in glucose but -9.4 in glucose nitrate, despite having 0 insertions in both output libraries. As such, genes will be referred to below as essential or as 'important for growth/fitness' in the case of being non-essential but with a significant logFC value. Readcounts and/or transposon insertion plots will be provided for clarification, where required.

Essential gene lists were generated for the INPUT and output libraries as described previously (see Chapter 3 and Materials and Methods). The numbers of essential genes were as follows: INPUT - 573, glucose - 698, glucose nitrate - 834, and glycerol nitrate - 740. Many genes in the output lists were those selected for previously in either the INPUT or base library (see Chapter 3). Therefore, filtering the dataset to remove these genes resulted in the following numbers of essential genes specific for the particular growth condition: glucose - 200, glucose nitrate - 328, and glycerol nitrate - 182 (see Tables A12 and A13, Appendix). LogFC values were calculated for each gene by comparing insertion counts and readcounts from the INPUT and output libraries, and, after

applying the appropriate statistical cutoffs, 163 genes showed significant logFCs after anaerobic growth using glucose, 529 genes with significant logFCs after anaerobic growth with glucose in the presence of nitrate, and 186 genes with significant logFCs after anaerobic growth on glycerol in the presence of nitrate.

All genes with logFC values could reliably be designated as playing a role in fitness specifically under a growth condition, whereas this could not be said for all genes nominated as essential. Therefore, subsequent analyses of the genetic requirements for fitness under the three different growth conditions were based upon the logFC lists (glucose = 163 genes; glucose nitrate = 529 genes; glycerol nitrate = 186 genes; see Table A13), with referral to essentiality where necessary.

In both essential and logFC lists, glucose nitrate displayed higher numbers of genes, suggesting that there may be a greater stress on cells during growth in glucose nitrate, i.e. more genes were required for full fitness.

5.2.3. Functional enrichment analysis

Functional enrichment analysis was conducted on the genes with significant logFC values (see Figure 5.3 and Tables A14-A17, Appendix).

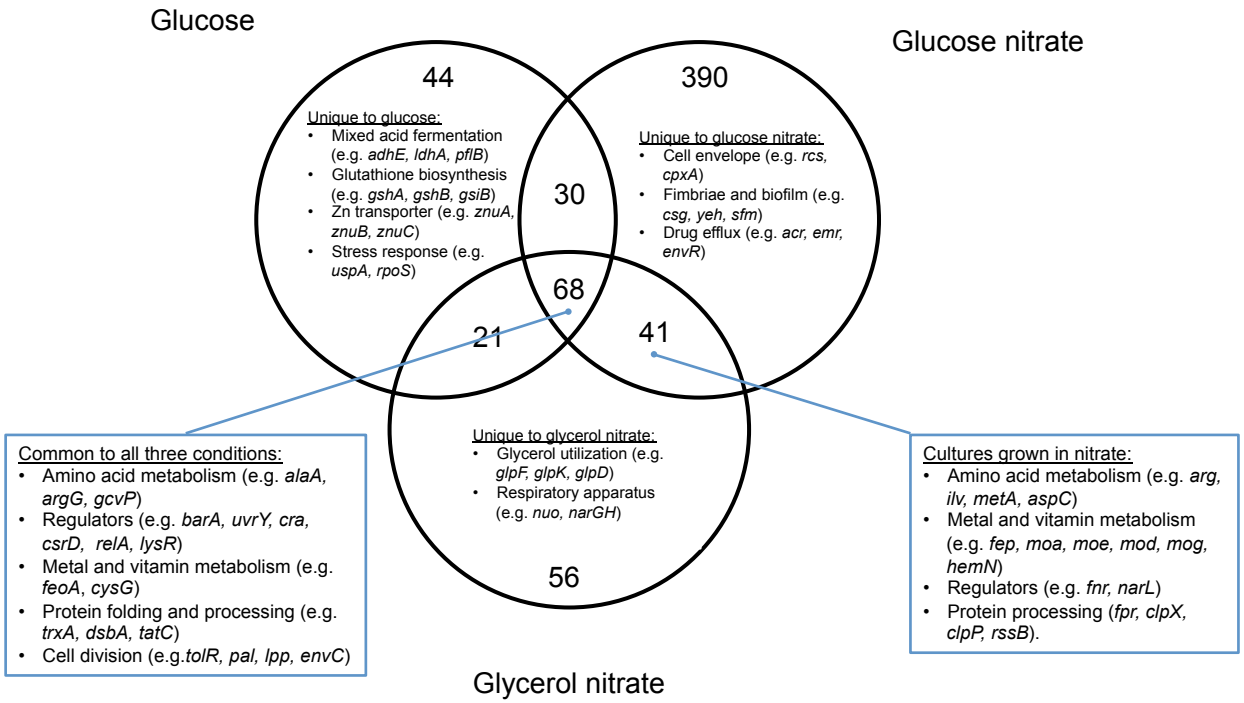


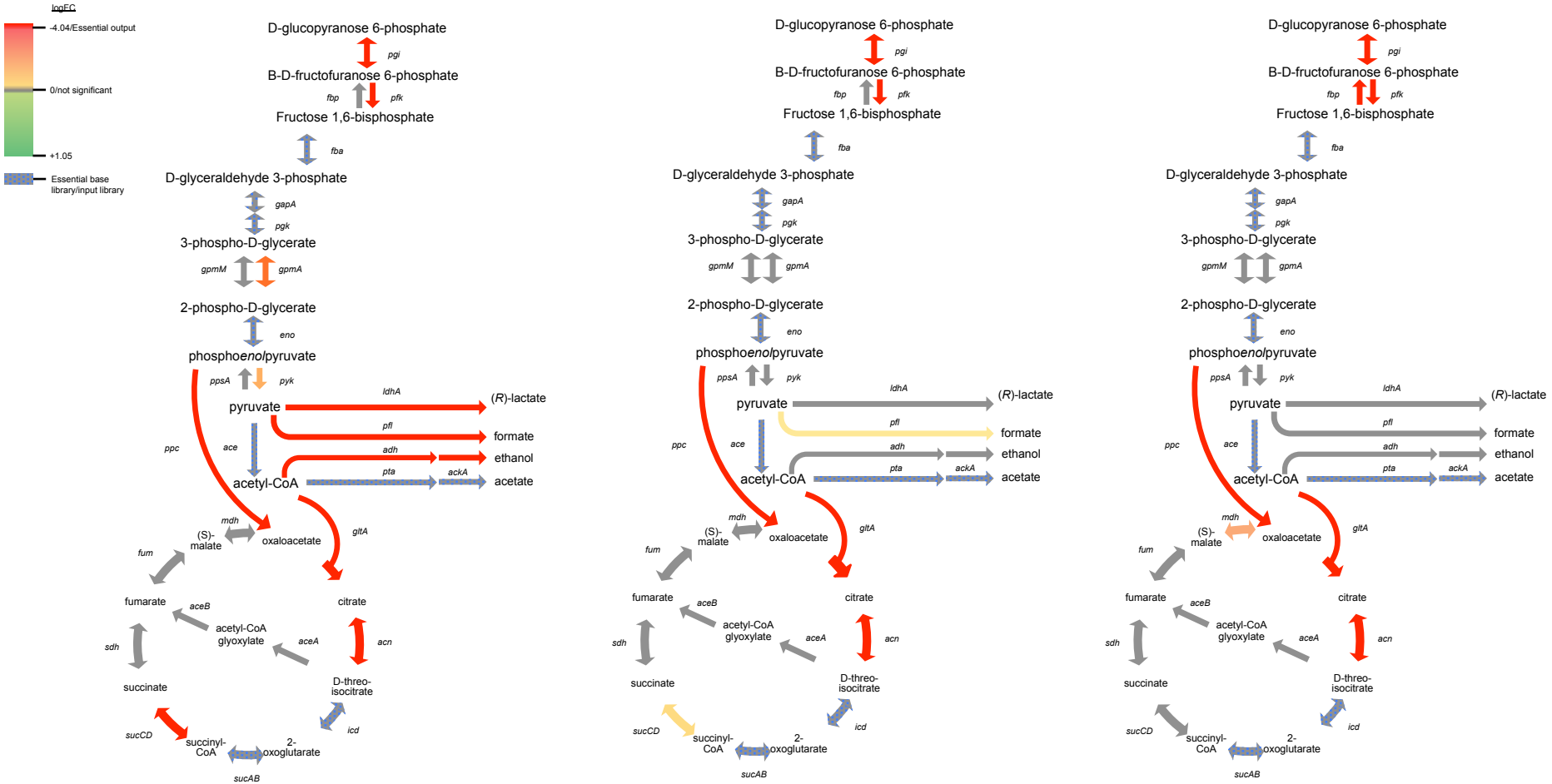
Figure 5.3. Functional groups shared between, and unique to, the three output libraries.

5.3. Gene category analysis - metabolism

5.3.1 Central metabolism

Optimal metabolism is important for efficient growth and defects in metabolism would be expected to result in a fitness disadvantage. Therefore, in order to investigate the effects of the different growth conditions on metabolism, maps of central metabolism were constructed for each growth condition using the logFC as a reporter of fitness. Figure 5.4 highlights these pathways and the degree of selection for genes therein. Not surprisingly, all three conditions shared a requirement for functional glycolysis and the oxidative branch of the TCA cycle, however a number of notable differences in the profiles of genes under the different growth conditions was observed.

Figure 5.4 (overleaf). Central metabolism and mixed acid fermentation in glucose (left), glucose nitrate (centre), and glycerol nitrate (right). Pathway reactions are represented by coloured arrows. The degree of selection for genes in each pathway are highlighted by the colour scale as follows: bright red - genes essential in the output libraries; dark orange to green - genes displaying significant negative to positive logFC values, respectively; grey - non-essential genes displaying no significant changes in fitness; grey arrows with blue spots - genes essential in the base library or input library.


logFC
-4.04/Essential output
0/not significant
+1.05
Essential base library/input library
D-glucopyranose 6-phosphate
pgi
B-D-fructofuranose 6-phosphate
fbp
pfk
Fructose 1,6-bisphosphate
fba
D-glyceraldehyde 3-phosphate
gapA
pgk
3-phospho-D-glycerate
gpmM
gpmA
2-phospho-D-glycerate
eno
phosphoenolpyruvate
ppsA
pyk
pyruvate
ldhA
(R)-lactate
pfl
formate
adh
ethanol
ppc
ace
pta
ackA
acetate
acetyl-CoA
mdh
(S)-malate
oxaloacetate
gltA
fum
fumarate
aceB
citrate
acetyl-CoA glyoxylate
sdh
aceA
acn
succinate
D-threo-isocitrate
sucCD
icd
succinyl-CoA
2-oxoglutarate
sucAB

5.3.1.1. Glycolysis and pyruvate generation

Glycolysis is essential for the generation of both energy and essential precursor metabolites and this was clearly demonstrated clearly in that most genes in glycolysis were deemed to be essential in either the base or INPUT libraries. The very few differences that were observed did illustrate clear differences in metabolism between the 3 conditions. For example, mutants of *fbp* (encoding fructose-1,6-biphosphatase) showed a fitness defect exclusively during growth in glycerol nitrate as Fbp is required for gluconeogenesis.

Another difference was the observed requirement for the *gpmA* gene, encoding phosphoglycerate mutase, during fermentative growth on glucose in the absence of nitrate. Phosphoglycerate mutase is responsible for the interconversion of 2-phospho-D-glycerate and 3-phospho-D-glycerate, and *E. coli* has two non-homologous genes encoding enzymes to carry out this reaction, i.e. cofactor-dependent phosphoglycerate mutase (*gpmA*), and cofactor-independent phosphoglycerate mutase (*gpmM*) where the cofactor is 2,3-bisphosphoglycerate. Previously, it was shown that during growth in super-optimal broth (SOB; a rich medium containing glucose), GpmA levels peak during mid-to-late log phase, as opposed to early exponential phase for GpmM. This suggests that during fermentative growth on glucose there may be an increased requirement for glycolytic activity to be maintained into late log phase (Fraser *et al.*, 1999).

Pyruvate represents an important metabolic node, since it feeds into several other biosynthetic and metabolic pathways. During growth on

glucose, pyruvate can be generated by glycolysis but is primarily generated via the activity of the phosphotransferase (PTS) system, which converts PEP to pyruvate during the uptake of glucose and other sugars. However, there did not appear to be clear differences in selection for core PTS components between the three different growth conditions. Independently of the PTS, pyruvate kinase (PK) can also generate pyruvate from PEP as the final step in glycolysis. Two PK activities have been identified in *E. coli*: PKI, encoded by *pykF*, and PKII, encoded by *pykA*. PykF is the most active form of PK and, according to TraDIS, *pykF* was selected exclusively during fermentative growth in glucose (Ponce *et al.*, 1995). This supports the hypothesis that there is a greater demand for glycolysis during fermentation compared to respiration.

5.3.1.2. Mixed acid fermentation

During mixed-acid fermentation, pyruvate is oxidised to lactate, formate, ethanol, succinate, and acetate resulting in the production of ATP (by substrate-level phosphorylation) and the regeneration of NAD^+ that has been consumed during glycolysis (Figure 5.5). TraDIS identified that all of the steps in mixed acid fermentation, with the exception of succinate production, were essential during anaerobic growth on glucose, but not essential during growth in the presence of nitrate. This is interesting considering the fact that measurable amounts of succinate were detected in supernatants of cultures grown under fermentative conditions in glucose (Figure 5.2B). Succinate is produced under anaerobic conditions in the presence of an alternative electron acceptor via fumarate reductase

(Unden & Bongaerts, 1997). However, succinate can also be produced via the TCA cycle by conversion of succinyl CoA to succinate by succinyl CoA synthetase (SucCD). In this study, *sucC* is scored as essential for anaerobic growth on glucose, suggesting that this may be the primary mechanism for succinate production during fermentative growth in glucose (Figure 5.4). The selection for *sucC* under these conditions has been demonstrated previously (see section 5.3.1.4, below). In addition, *ldhA,* encoding lactate dehydrogenase was also identified as an essential gene during anaerobic growth on glucose and, although lactate is produced during mixed-acid fermentation, a major fitness defect has not been previously observed for *ldhA* mutants during growth under these conditions (Mat-Jan *et al.*, 1989a). Deletion of *ldhA* has been shown to result in slightly reduced growth and glucose consumptions rates in glucose minimal medium under anaerobic conditions (Kabir *et al.*, 2005). Therefore, the identification of *ldhA* as an essential gene by TraDIS highlights the power of this technique in describing mutant phenotypes not normally detectable by standard approaches.

It is interesting to note that the secretion of lactate, succinate, and formate is known to be important in maintaining the essential proton motive force (pmf) under fermentative conditions (Trchounian & Trchounian, 2019). Therefore, the fact that all steps of the mixed-acid fermentation pathway that produce these compounds were essential in glucose, and that succinate was detectable in culture supernatants, would suggest that maintenance of the pmf is under significant selective pressure during competitive growth while cells are fermenting glucose.

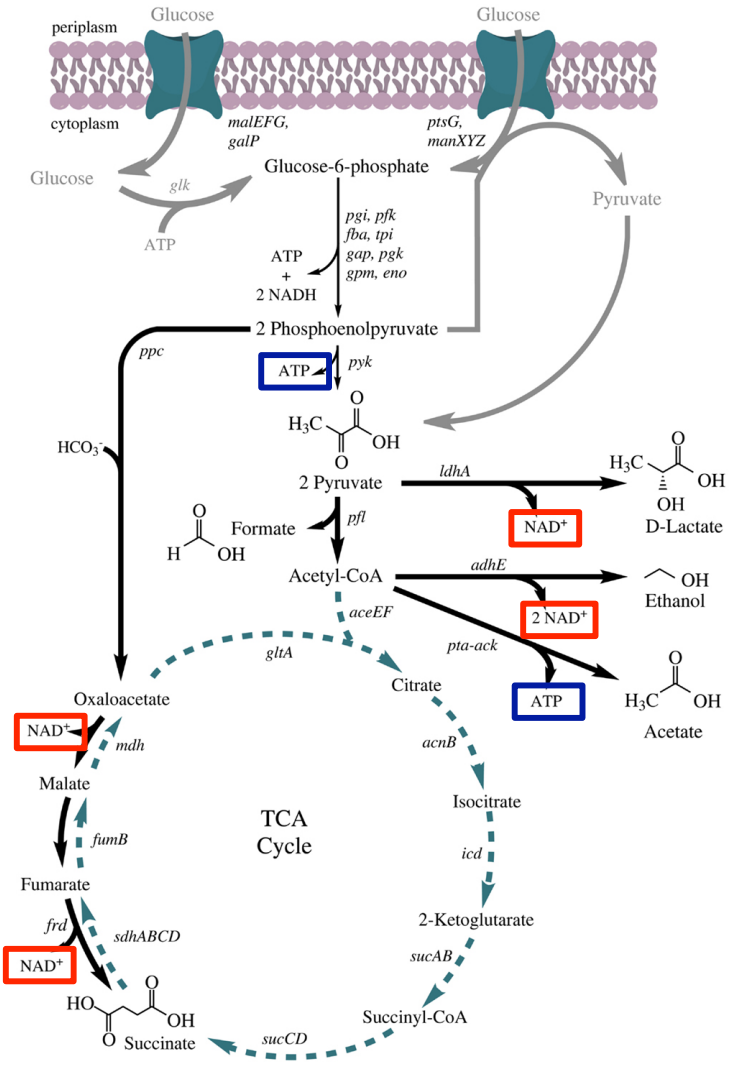
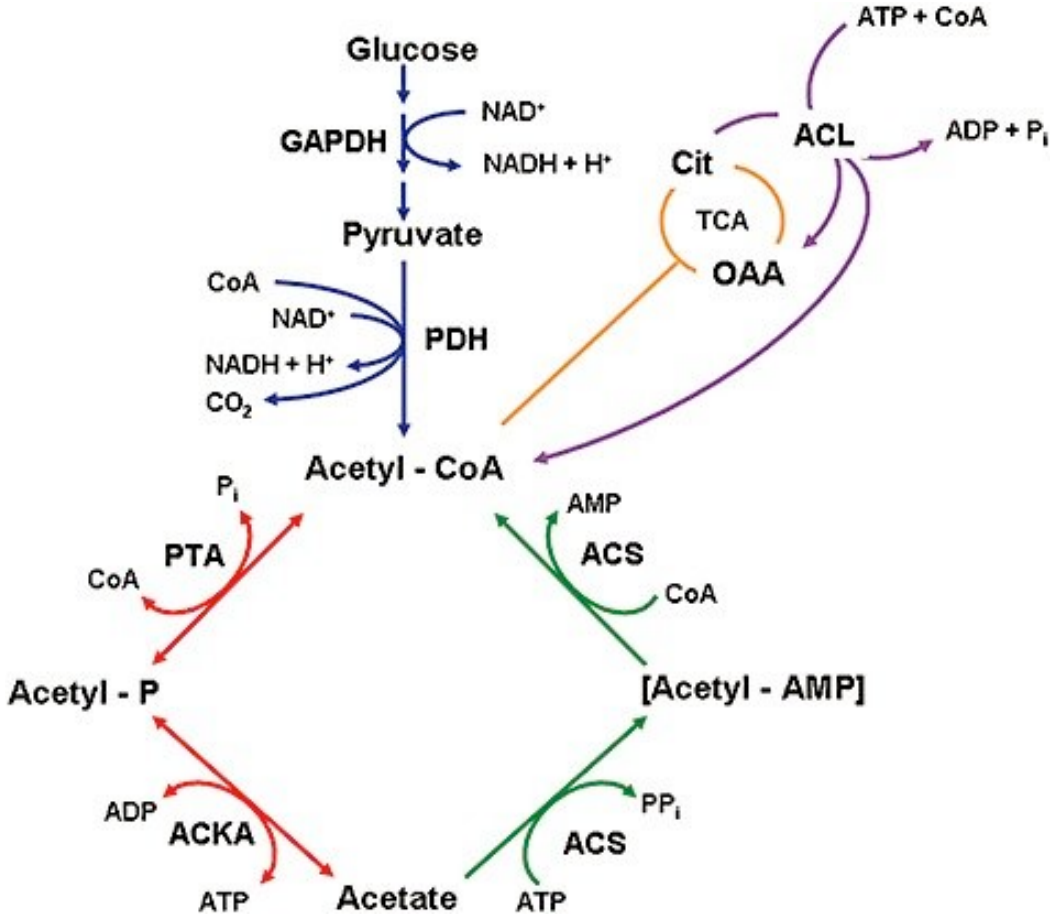


Figure 5.5. Mixed acid fermentation with ATP (blue boxes) and NAD^+ (red boxes) generation steps highlighted. Adapted from Förster and Gescher 2014.

A.


Glucose
GAPDH
NAD+
NADH + H+
Pyruvate
CoA
NAD+
NADH + H+
CO2
PDH
Acetyl - CoA
Cit
ACL
ATP + CoA
ADP + Pi
TCA
OAA
Pi
PTA
CoA
Acetyl - P
AMP
ACS
CoA
[Acetyl - AMP]
ADP
ACKA
ATP
Acetate
PPi
ACS
ATP

B.

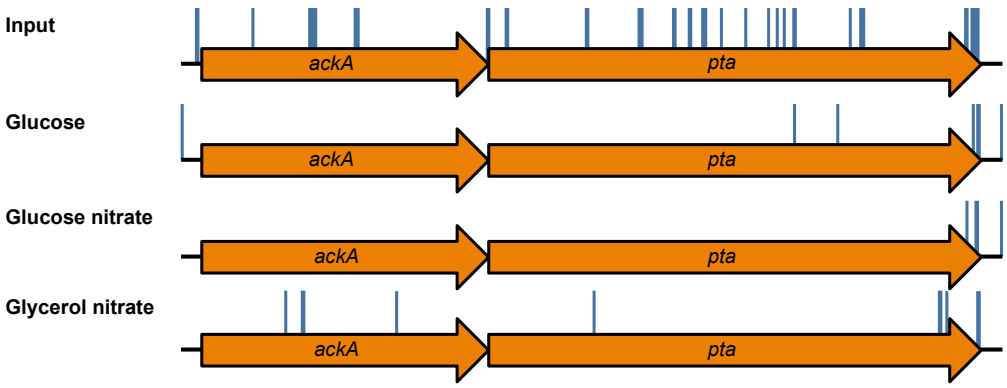

Input
ackA
pta
Glucose
ackA
pta
Glucose nitrate
ackA
pta
Glycerol nitrate
ackA
pta

Figure 5.6. (A) Acetate assimilation (green) and dissimilation (red), and its connection to other pathways of central metabolism (other colours). Acetate dissimilation involves *ackA* (encoding acetate kinase) and *pta* (encoding phosphate acetyltransferase). Acetyl CoA generates acetyl phosphate via Pta, and acetyl phosphate is then converted to acetate (secreted from the cell) and ATP by AckA. From (Hu *et al.*, 2010) (**B)** Transposon insertions in *ackA* and *pta* in the INPUT and three output libraries. Blue lines indicate transposon insertions (length not scaled to readcounts).

5.3.1.3. Acetate dissimilation

During growth on glucose the production of acetate is an important source of ATP (Figure 5.6A). Both *ackA* (encoding acetate kinase) and *pta* (encoding phosphate acetyltransferase) are labeled as essential in the INPUT and all output libraries (Figure 5.4). However, examination of transposon insertion plots reveals several insertions in these genes in the INPUT, which are subsequently lost to differing degrees under the different growth conditions (Figure 5.6B). Nevertheless, it would appear that there is a strong selection on these genes under the three different growth conditions, highlighting the importance of acetate dissimilation for ATP generation. However, the requirement for acetate dissimilation during growth in glycerol, as well as the simultaneous production of detectable levels of acetate under those conditions, was surprising. While the production of high levels of acetate by *E. coli* growing using glycerol has been reported previously, it was thought that neither *pta* nor *ackA* were required under those conditions (Chang *et al.*, 1999; Prohl *et al.*, 1998). *E. coli* excretes acetate under a number of different conditions, including when carbon flux exceeds the capacity of the TCA cycle or other central metabolic pathways, or when recycling of CoA is required to convert pyruvate to AcCoA (Wolfe, 2005). Therefore, acetate dissimilation may be playing one, or a combination of, these essential roles under the three different growth conditions.

5.3.1.4. TCA cycle

Under anaerobic conditions, the TCA cycle divides into oxidative (citrate to α-ketoglutarate) and reductive (oxaloacetate to succinate) branches. In this way it does not operate to generate energy, but, rather, provides precursor metabolites for biosynthesis. The oxidative branch was essential under all three growth conditions (Figure 5.4), presumably as it is required for the production of the key metabolite, α-ketoglutarate, required for e.g. the biosynthesis of several amino acids. In contrast, enzymes of the reductive branch were not under significant selection, with the exception of succinyl CoA synthetase, which cataylses the interconversion between succinyl CoA and succinate.

Succinyl CoA synthetase is encoded by *sucCD*, mutants of which were under greatest selection in glucose (logFC values of -5.2 and -5.7, respectively), followed by glucose nitrate (-1.7 and -1.9, respectively), with no selection under glycerol nitrate (Figure 5.7A). Greater amounts of succinate were produced by cultures grown in glucose only (Figure 5.2B), however this does not necessarily suggest that there was a higher level of succinyl CoA synthetase activity under these conditions, since succinate can be produced via several different mechanisms in *E. coli* (Yu *et al.*, 2011). Nevertheless, these results are in agreement with a previous study where it was shown that *sucCD* insertion mutants could not grow in M9 medium anaerobically in the presence of glucose, but could grow in M9 medium supplemented with glycerol and nitrate (Mat-Jan *et al.*, 1989b). Moreover, it has been demonstrated that MG1655

knockout mutants of *sucCD* produced lower biomass following growth in a glucose minimal medium compared to the wild type (Veit *et al.*, 2007).

As mentioned above (section 5.3.1.2), the selection on *sucC* is probably related to the role of succinate secretion to generate a pmf under anaerobic fermentation conditions (Trchounian & Trchounian, 2019). The selection may also be related to its role in amino acid biosynthesis (see section 5.3.2, below). However, it was notable that the end-point growth a Δ*sucC* mutant isolated from the Keio library was not significantly different compared to the wild type following growth under all three conditions (Figure 5.7B), suggesting that *sucC* mutants are capable of growing to high densities under the three growth conditions but are particularly sensitive to population stresses, likely related to the low growth rates displayed by *sucCD* mutants (Yu *et al.*, 2006).

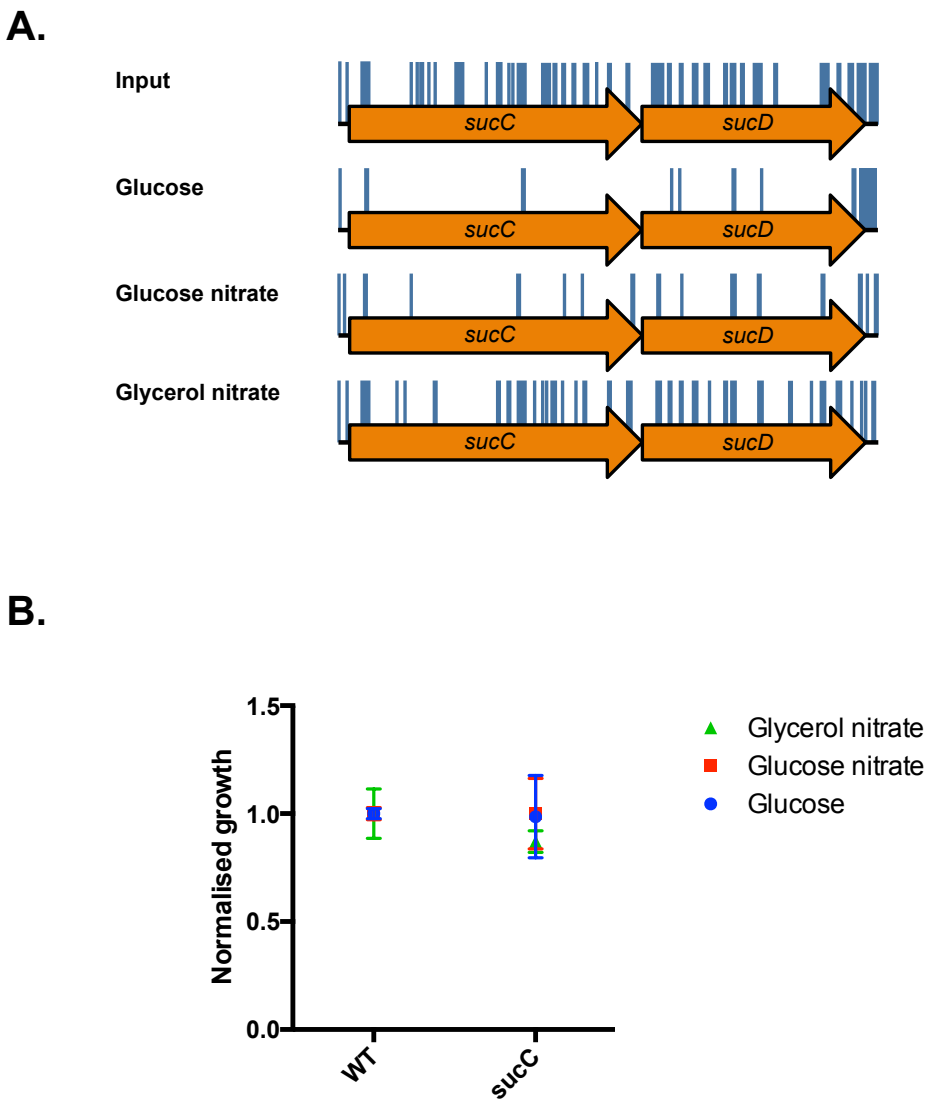


Figure 5.7. (A) Transposon insertions in *sucCD* in the INPUT and three output libraries. Blue lines indicate transpon insertions (length not scaled to readcounts). **(B)** Growth, normalised to the wild-type (line at y = 1), of a Keio Δ*sucC* mutant. Growth is compared to the wild type in M9 medium under anaerobic conditions, supplemented as indicated.

5.3.2. A metabolic model informs results in TraDIS

A model for the changes in flux through central metabolism and the respiratory chain associated with the addition of nitrate to M9 medium with glucose under anaerobic conditions was constructed using a combination of metabolomics and transcriptomics by Toya and colleagues (Toya *et al.*, 2012). Removal of oxygen had the effect of reducing metabolic flux through the respiratory chain and the TCA cycle, and this was correlated with a reduction in the expression of genes encoding proteins involved in these processes. Upon the addition of nitrate, the respiratory chain became active again; however, gene expression and flux through the TCA cycle remained repressed. Moreover, the model showed that, during fermentative growth with glucose, approximately 66% of ATP is generated through glycolysis, falling to approx. 33% upon the addition of nitrate. Mapping logFC values of non-essential genes generated during TraDIS onto central metabolism displayed some patterns similar to the flux model generated by Toya *et al*. (Figure 5.9). For example, flux from 3-phospho-glycerate to PEP was increased during fermentative growth compared to nitrate respiration (Toya *et al.*, 2012), which corresponds to the reduced fitness of *gpmA* and *pykF* mutants in glucose versus glucose nitrate. Similarly, flux to formate was greater in fermenting cells compared to respiring cells and this corresponds to TraDIS analysis identifying PFL as an essential enzyme in fermenting cells.

According to the metabolic model, flux from succinate to malate is decreased in fermenting cells, but increased in respiring cells (Toya *et al.*,

2012). However, in fermenting cells, there is also an increase in flux from succinate to other pathways e.g. diaminopimelate and L-methionine biosynthesis whilst there is no change in flux in respiring cells (Toya *et al.*, 2012). This suggests that succinyl CoA synthetase plays an important role in biosynthesis during fermentation, with a more prominent role in the TCA cycle during respiration. The importance of these distinct roles is reflected in the different logFC values for *sucCD* under these growth conditions. Together, these suggest that logFC values obtained in TraDIS can reflect degrees of metabolic flux through central metabolism.

However, TraDIS could also enhance the results of the metabolic model by outlining the importance of particular metabolic pathways for fitness, even in the instance where flux to those pathways was not particularly great. For example, there appeared to be relatively low flux to lactate during fermentation, despite the fact that *ldhA* was essential for growth in glucose (Figure 5.8). Therefore, even under low flux, lactate production under fermentative conditions was extremely important for cell survival. This may be related to the role of lactate secretion in the maintenance of the pmf, as previously mentioned (Trchounian & Trchounian, 2019), and would further highlight that cells are highly sensitive to disruption of the pmf during fermentative growth on glucose.

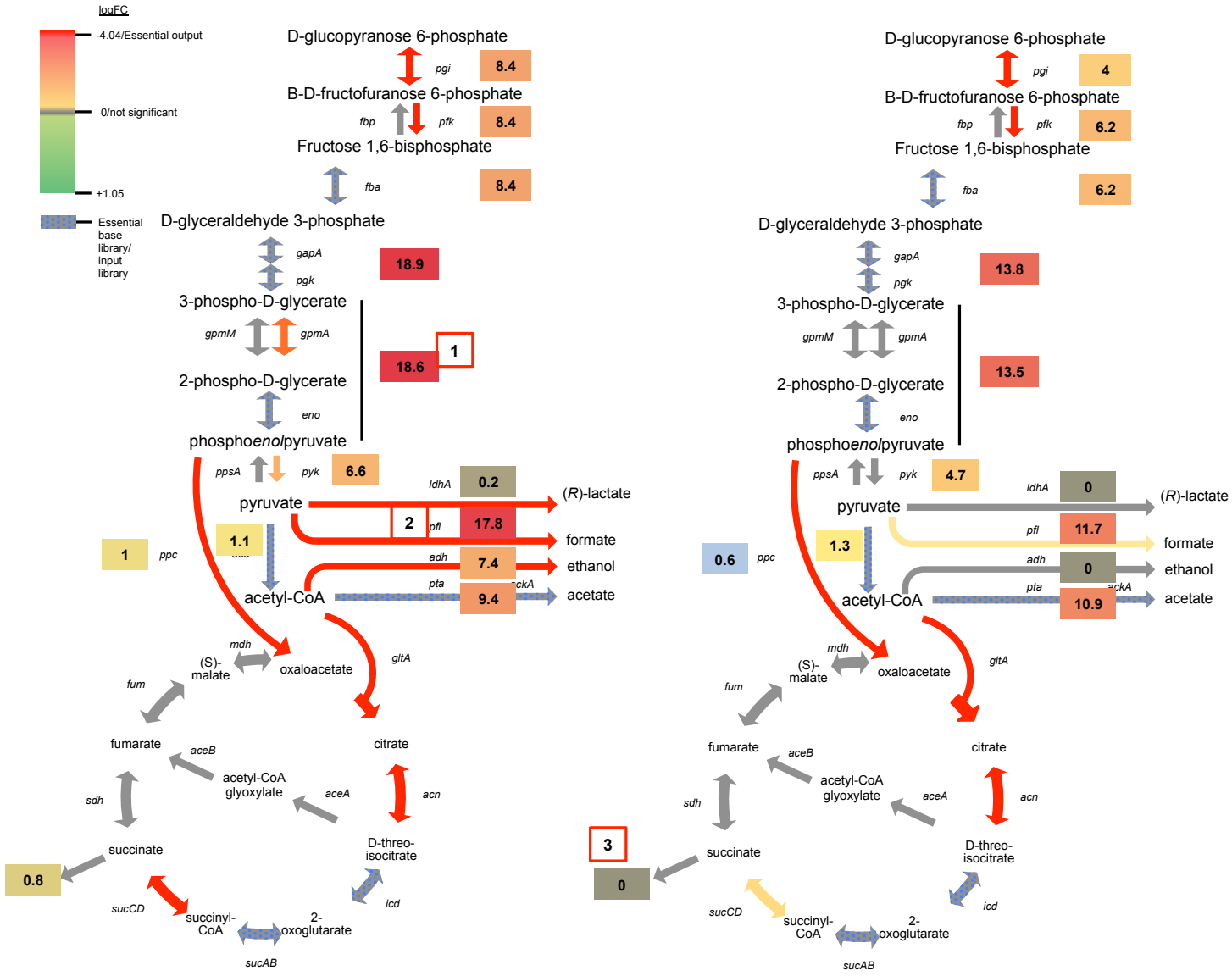

logFC
-4.04/Essential output
0/not significant
+1.05
Essential base library/ input library
D-glucopyranose 6-phosphate
pgi
8.4
B-D-fructofuranose 6-phosphate
fbp
pfk
8.4
Fructose 1,6-bisphosphate
fba
8.4
D-glyceraldehyde 3-phosphate
gapA
pgk
18.9
3-phospho-D-glycerate
gpmM
gpmA
18.6
1
2-phospho-D-glycerate
eno
phosphoenolpyruvate
ppsA
pyk
6.6
ldhA
0.2
(R)-lactate
pyruvate
2
pfl
17.8
formate
1.1
1
ppc
adh
7.4
ethanol
pta
ackA
9.4
acetate
acetyl-CoA
mdh
(S)-malate
oxaloacetate
gltA
fum
fumarate
aceB
citrate
acetyl-CoA glyoxylate
aceA
acn
sdh
D-threo-isocitrate
succinate
0.8
icd
sucCD
succinyl-CoA
2-oxoglutarate
sucAB
D-glucopyranose 6-phosphate
pgi
4
B-D-fructofuranose 6-phosphate
fbp
pfk
6.2
Fructose 1,6-bisphosphate
fba
6.2
D-glyceraldehyde 3-phosphate
gapA
pgk
13.8
3-phospho-D-glycerate
gpmM
gpmA
13.5
2-phospho-D-glycerate
eno
phosphoenolpyruvate
ppsA
pyk
4.7
ldhA
0
(R)-lactate
pyruvate
pfl
11.7
formate
0.6
ppc
1.3
adh
0
ethanol
pta
ackA
10.9
acetate
acetyl-CoA
mdh
(S)-malate
oxaloacetate
gltA
fum
fumarate
aceB
citrate
acetyl-CoA glyoxylate
aceA
acn
sdh
D-threo-isocitrate
3
succinate
0
icd
sucCD
succinyl-CoA
2-oxoglutarate
sucAB

Figure 5.8. Comparison between logFCs and flux through central metabolism. Central metabolism during growth on glucose (left) and glucose nitrate (right) with logFC values mapped onto equivalent pathway arrows. Degrees of pathway selection are indicated by a colour scale as per Figure 5.4. Flux values (in mmol $gDCW^{-1}$ h^{-1}) adapted from Toya *et al.* 2012 are highlighted in coloured boxes, with a greater number representing a higher degree of metabolic flux (Toya *et al.*, 2012). Reactions mentioned in the text are highlighted as follows: 1. 3-phosphoglycerate to PEP; 2. Formate production via PFL; 3. Succinate to biosynthetic pathways.

5.3.3. Nitrate respiration in glucose nitrate

Growth using nitrate as a terminal electron acceptor is well-characterized at the genetic and biochemical level (Cole & Richardson, 2013). The ability to respire with nitrate is tightly regulated and requires the appropriate expression of a large number of genes (Constantinidou *et al.*, 2006; Goh *et al.*, 2005). Important regulatory genes such as *fnr* (encoding FNR, a protein required for activation of anaerobic respiration, including the expression of nitrate reductase genes) and *narL* (encoding a protein important for regulating the response to high levels of nitrate) were strongly selected during growth with nitrate. Moreover, molybdopterin is an essential cofactor for the functioning of nitrate reductases and several genes encoding molybodpterin biosynthesis proteins e.g. *moa*, *moe*, *mog*, were also under strong selection during growth in the presence of nitrate (Figure 5.3).

However, there was no selection for the genes encoding nitrate reductases during growth in the presence of glucose and nitrate, suggesting that nitrate reduction per se may not be important for fitness during growth on glucose in the presence of nitrate. This is despite clear evidence of nitrate reduction to nitrite, and the absence of ethanol and lactate (indicating alternative forms of NAD^+ regeneration e.g. via respiratory pathways) in cell-free supernatants of the mutant library following growth in glucose with nitrate (Figure 5.2B, 5.2C). In contrast, there was a strong selection for *narG* and *narH*, encoding two of the three subunits of the NarGHI nitrate reductase A (NRA), and several *nuo* genes, encoding NADH dehydrogenase I, when the bacteria were

cultured in glycerol nitrate. This clearly suggests that nitrate respiration from NADH dehydrogenase I to NarGHI (Nitrate Reductase A, or NRA) was required for fitness in glycerol nitrate.

There are three nitrate reductases in *E. coli*, cytoplasmic NRA and NarZYX (Nitrate Reductase Z, or NRZ), and the periplasmic reductase, Nap. However, these three systems are not thought to be redundant, but rather complementary, operating according to the concentration of nitrate in the environment (Rodrigues *et al.*, 2006; Wang *et al.*, 1999; Wang & Gunsalus, 2000). Therefore, higher concentrations of nitrate induce the expression of NRA exclusively (as observed by selection of these genes during growth in glycerol nitrate), whereas lower concentrations allow for the expression of NRZ and Nap. Nonetheless, the lack of selection for NRA or any specific nitrate reductase-encoding genes during growth in glucose nitrate does suggest some functional redundancy in nitrate reductase activity in glucose nitrate.

To confirm that nitrate respiration is not required for fitness during growth in glucose nitrate, a Δ*moaC* mutant, deficient in molybdopterin biosynthesis and therefore containing no functional nitrate reductases (Johnson & Rajagopalan, 1987), was competed against wild type *E. coli* under similar conditions used during the mutant library screen (Figure 5.9A). In agreement with TraDIS, the Δ*moaC* mutant was less fit during growth on glycerol nitrate, but there was no obvious fitness defect during growth using glucose or another carbon source, gluconate. This was further confirmed with an end-point growth assay (Figure 5.9B), whereby the growth defects of mutants unable to produce molybdopterin were

much greater when cells were grown in glycerol compared to glucose. This suggests that nitrate respiration is dispensable during anaerobic growth using a fermentable carbon source such as glucose.

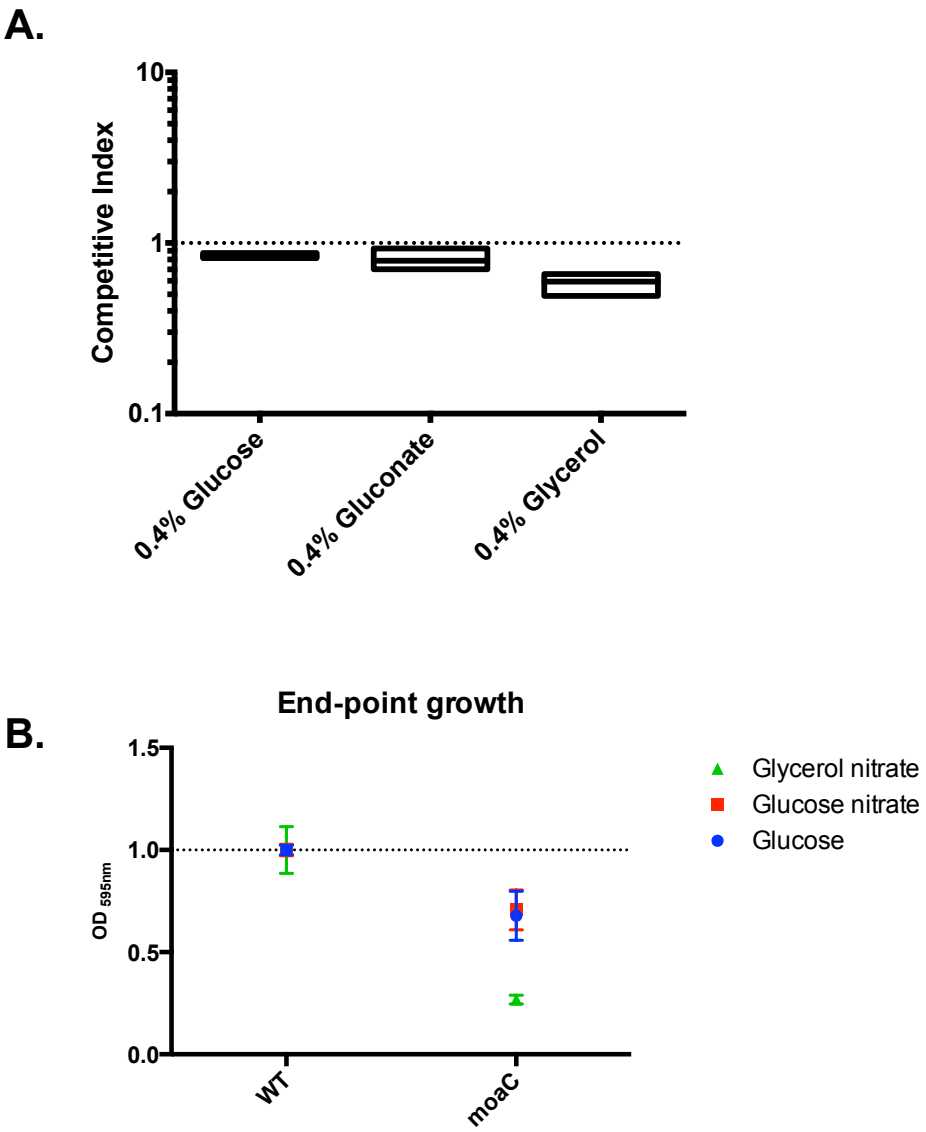


Figure 5.9. (A) Competition assays of a Keio Δ*moaC* mutant competed against the wild type in M9 minimal medium supplemented with the indicated concentrations of carbon sources. Wild-type competitiveness is indicated by a line at 1. **(B)** End-point growth of the Δ*moaC* mutant under the three conditions tested for TraDIS. Growth is normalised to the wild type figure of 1.

During nitrate respiration, NAD^+ can be regenerated via the activity of the NADH dehydrogenase, Nuo. Formate can also be generated in the absence of functional nitrate reductases, presenting another route by which NAD^+ can be regenerated (Kaiser & Sawers, 1994). ATP can be formed by oxidative phosphorylation during nitrate respiration; however, ATP can also be generated via acetate dissimilation (Wolfe, 2005). Therefore, it is likely that, in the absence of nitrate respiration, cultures growing in glucose nitrate can generate redox balance and energy in sufficient amounts via alternative mechanisms. Furthermore, a proton motive force can be produced during glucose fermentation via the activity of the F_1F_0 ATPase, which can hydrolyse ATP and pump H^+ ions out of the cell, and via the secretion of organic acids as previously mentioned (Trchounian & Trchounian, 2019). Therefore, cultures growing in glucose can display a high degree of metabolic flexibility. However, in carbon sources such as glycerol, the uptake of glycerol into the cell for gluconeogenesis and glycolysis requires functioning respiratory chains, therefore nitrate respiration is more essential for fitness during growth on this carbon source (Cole *et al.*, 1988; Schryvers *et al.*, 1978).

5.4. Gene category analysis - genes outside of central metabolism

Many genes encoding proteins predicted to participate in pathways or cellular processes other than central metabolism were also found in the TraDIS gene lists (Table A13). For example, within the 68 genes selected under all three growth conditions (see Figure 5.3) were gene functional groups such as cell division (e.g. *tolR, pal, lpp,* and *envC*), regulators

(e.g. *barA-uvrY, cra, csrD,* and *lysR*), and protein folding and processing mechanisms (e.g. *trxA, dsbA,* and *tatC*). However, many more genes and functional groups were selected under two or fewer growth conditions, highlighting that the addition of nitrate and/or carbon source exerts unique global selective pressures on the cell beyond central metabolism (see Figure 5.3).

5.4.1. Genes identified in nitrate-containing cultures

Non-central metabolism genes under selection uniquely in nitrate-containing cultures appeared to be primarily encoding functions directly related to the maintenance or regulation of the nitrate respiratory apparatus (see Table A18, Appendix). These included, for example, genes involved in molydbopterin biosynthesis (e.g. *moa, mog, moe, mod*), or the regulation of nitrate respiration (*fnr* and *narL*). Other genes in this category were also likely to play an indirect role in nitrate respiration, e.g. *clpXP* encodes the ClpXP serine protease, which has been shown to be involved in the degradation and turnover of FNR (Baker & Sauer, 2012; Mettert & Kiley, 2005). FNR contains an $[4Fe\text{-}4S]^{2+}$ cluster which maintains the protein in a dimerized form, allowing it to act as a transcription factor. O_2 converts the $[4Fe\text{-}4S]^{2+}$ cluster to $[2Fe\text{-}2S]^{2+}$, converting FNR to an inactive, monomeric form. ClpXP targets and degrades the monomeric form of FNR, allowing to maintain appropriate levels of active FNR during aerobic growth (Lazazzera *et al.*, 1996). The requirement for both *clpXP* and *fnr* in nitrate-containing cultures underpins the importance of optimally functioning FNR for fitness.

5.4.2. Functions unique to each growth condition

TraDIS demonstrated that each growth condition required unique functions or gene functional groups (see Figure 5.3 and Tables A15-A17). Some of these functional groups were expected, such as glycerol utilisation genes (e.g. *glpF, glpK,* and *glpD*) were unique to the glycerol nitrate list, however others were not, highlighting potential differences in environmental conditions induced by carbon source and/or nitrate. In addition, *gshA* and *gshB*, encoding both enzymes of the glutathione biosynthesis pathway, exhibited positive logFC values (+2.08 and +1.54, respectively) after growth in glucose. Glutathione (GSH) is an important component of the thiol redox system, responsible for controlling the redox state of cytoplasmic cysteine residues via thiol-disulfide exchange reactions. In *E. coli*, a variety of processes act as substrates to this system, ranging from DNA synthesis, to H_2O_2 metabolism, to sulfate assimilation, to arsenate reductase (Figure 5.10) (Toledano *et al.*, 2007).

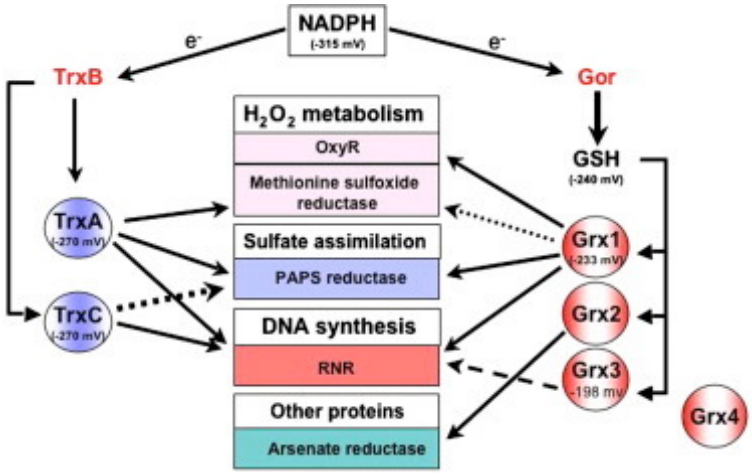


Figure 5.10. Components of the *E. coli* thioredoxin and glutaredoxin systems. From Toledano *et al*., 2007.

The *E. coli* thiol redox system contains two functionally redundant branches, the thioredoxin and glutaredoxin branches, and inactivation of either branch does not appear to affect normal aerobic growth (Toledano *et al.*, 2007). Therefore, it was interesting to see an essential requirement for *trxA* (encoding thioredoxin 1) under all three conditions, but a significant fitness advantage for *gshA* and *gshB* mutants under fermentative conditions only. It has been shown previously that GSH plays a role in the formation of the active, [4Fe-4S] cluster-containing form of FNR (Tran *et al.*, 2000). Also, deletion of *fnr* in *E. coli* BW25113 has been shown to result in increased flux through formate, acetate, and ethanol production, all shown to be essential processes for growth under the glucose condition in this study (Kargeti & Venkatesh, 2017). Therefore, mutation of *gshA* and *gshB* may have provided a growth advantage under fermentative conditions by indirectly increasing flux through mixed acid fermentation pathways.

5.4.3. Genes with unknown function (y genes)

In *E. coli*, y genes i.e. genes without an annotated function, represent approximately a third of the genome (Ghatak *et al.*, 2019). Y genes formed a significant portion of each gene list, but fewer y genes were present in the glucose or glycerol nitrate gene lists compared with the glucose nitrate list, possibly reflecting the greater stress placed on the glucose nitrate system (see Figure 5.11 and Table A19, Appendix). The selection for y genes highlights that many unknown processes play a role during growth under these conditions, illustrating the power of TraDIS in identifying and characterising novel biological processes.

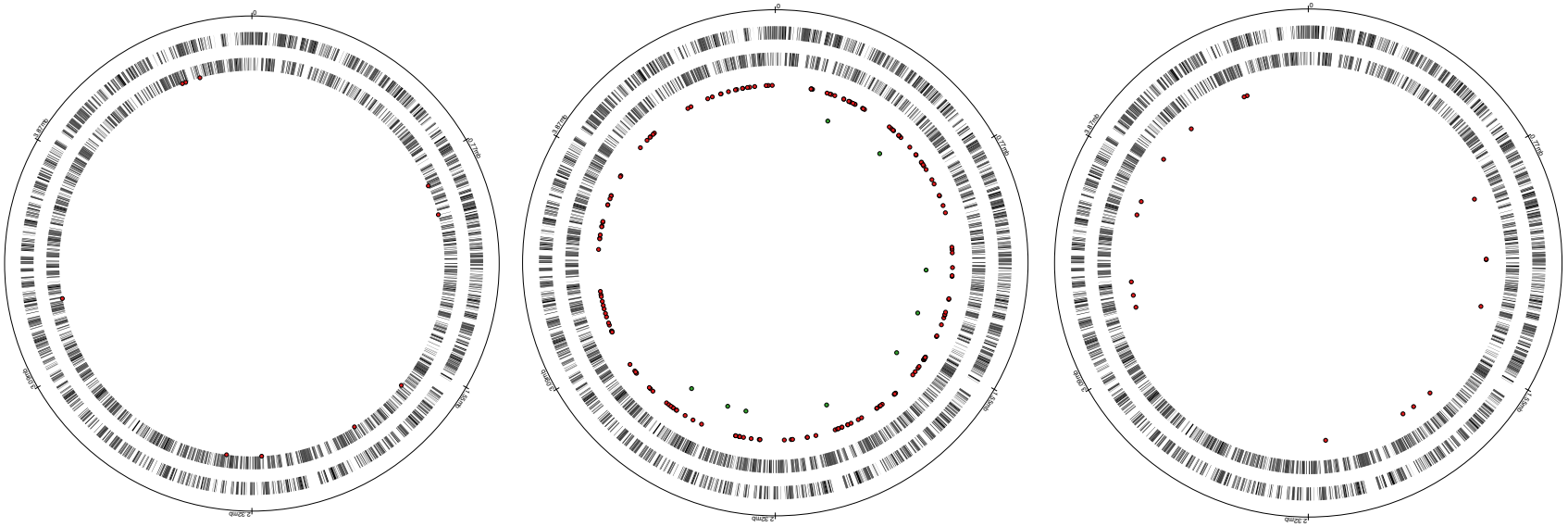


Figure 5.11. Location of y genes (red circles) on the MG1655 genome in glucose (left), glucose nitrate (middle), and glycerol nitrate (right).

Analysis of the predicted functions of proteins encoded by y genes resulted in some genes grouping into categories, including putative cell envelope-associated genes, putative regulators, fimbrial-like adhesin proteins, pseudogenes, and prophage genes (Table A19). Interestingly, an essential gene in glycerol nitrate, *yraP*, also displayed a reduced logFC value in the study of genetic requirements for bile (Chapter 4). As mentioned previously (see section 4.1.2, Chapter 4) YraP has been implicated in the activation of peptidoglycan hydrolase, AmiC (via NlpD) required for cell envelope invagination during cell division (Tsang *et al.*, 2017). *yraP* mutants are also more sensitive to SDS than the wild type, suggesting that *yraP* mutants exhibit defects in membrane integrity may have an increased fitness cost during growth under nitrate-respiring conditions (Onufryk *et al.*, 2005).

The inclusion of prophage genes solely in glucose nitrate (see Table 5.3) is notable, since prophages of *E. coli* K-12 are thought to contain mechanisms by which the bacterium can withstand stress (Wang *et al.*, 2010). It is also possible that that these genes were selected due to excision of prophages, which can occur upon encountering stressful conditions (Wang *et al.*, 2009). However, transposon insertion plots suggest that this probably did not occur, since many transposon insertions were present in specific genes (as opposed to being completely absent due to the absence of prophage DNA available for insertion). Therefore, prophages harboured genes with potentially important roles to play in fitness during anaerobic growth using glucose and nitrate.

Prophage	Gene(s) under selection
CP4-6	*yagL, yagM*
DLP12	*appY, ybcK, ybcL, ybcM, ybcY*
e14	*lit, mcrA, tfaP*
Rac	*trkG, ynaK*
Qin	*cspB, cspI, fixA, intQ, pinQ, ydfD, ydfJ*
CP4-44	-
CPS-53	*yfdK, yfdL, yfdQ*
CPZ-55	-
CP4-57	*alpA, yfjH, yfjI, yfjJ, yfjW*

Table 5.3. Prophage-associated genes under selection in glucose nitrate.

5.4.3.1. Y genes selected by TraDIS do not show a general growth defect

As illustrated in Chapter 4, end-point growth analysis of Keio mutants could potentially reveal some information behind logFC values obtained in TraDIS. Therefore, Keio library mutants of y genes that displayed the greatest logFC values in glucose nitrate were grown under the three different growth conditions (Figure 5.12). This revealed 3 different groups of genes: those that grew poorly under all three conditions (*yehA, yciB, ybcK*), mutants that grew poorly in nitrate cultures only (*yjbM, yceD*), or mutants that grew poorly in glycerol nitrate only (*ybaM, yhhZ, ygeH, yqeH, yhiL, yfjW, yfbN*). Therefore, there is no correlation between growth and fitness under the conditions tested.

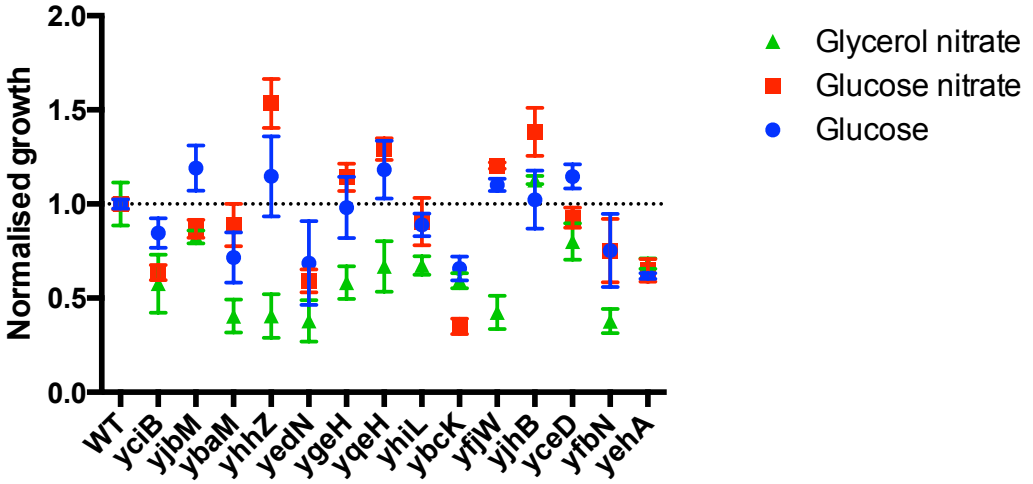


Figure 5.12. End-point growth of a selection of Keio y gene mutants following growth under the TraDIS assay conditions. The dotted line indicates the normalised wild type growth of 1.

Conclusions

In this study, TraDIS was employed to describe the genetic requirements for growth under anaerobic conditions in the presence of two different carbon sources and/or the addition of nitrate. In this way, it allowed the detailed description of global genetic requirements for growth under the conditions of anaerobic fermentation and anaerobic nitrate respiration. Of the three different growth conditions tested, it was shown that anaerobic growth in glucose with nitrate was the most stressful as demonstrated by the greater number of essential genes and genes with significantly reduced fitness values (as indicated by logFC values).

Genetic requirements for growth under the three different growth conditions were characterised based on differences in central metabolic

pathways and genes outside of central metabolism. In terms of central metabolism, differences were observed in each of the major pathways of central metabolism i.e. glycolysis, mixed acid fermentation, the TCA cycle, and nitrate respiration. Anaerobic growth in glucose without nitrate led to an increased flux requirement via glycolysis and an essential requirement for mixed acid fermentation. Anaerobic growth in glycerol and nitrate led to an essential requirement for glycerol uptake and nitrate respiration, but no fitness requirement for mixed acid fermentation pathways. However, anaerobic growth with glucose and nitrate allowed for a degree of metabolic flexibility, where neither mixed acid fermentation nor nitrate respiration were required for full fitness. Notably, when comparing fitness values from TraDIS to a metabolic flux model (Toya *et al.*, 2012), it was shown that fitness values were able to reflect degrees of metabolic flux through the pathways of central metabolism. Therefore, TraDIS could be applied to future metabolic flux analyses to further inform metabolic requirements for growth.

TraDIS also described genetic requirements for growth outside of central metabolic pathways, demonstrating unique physiological and regulatory requirements induced by carbon source and/or the addition of nitrate. These included a large number of unannotated y genes, indicating that much has yet to be understood about the effect of carbon source, addition of nitrate, and/or anaerobic growth on cell physiology.

Chapter 6 General discussion

E. coli takes part in a complex relationship with its host, with single strains potentially existing as commensals, pathogens, or pathobionts. The nature of this relationship is influenced by a variety of different factors, including the response of the *E. coli* strain to the environmental conditions within the GI tract, ranging from stomach acid, to bile, to inflammation. It is a testament to the versatility of the species that *E. coli* can adapt to, and even thrive within, each of these stressful conditions. However, with the fact that the *E. coli* pangenome is thought to contain upwards of 43,000 genes (Snipen *et al.*, 2009), and that various systems within the bacterial cell can be employed simultaneously in response to environmental stresses, there exists a significant challenge towards fully understanding how the species adapts to life within the GI tract. Addressing this challenge requires the use of high-throughput experimental approaches.

In addressing this challenge, this study employed TraDIS to describe the genetic requirements for growth in the presence of bile and under anaerobic conditions in the presence of nitrate, two conditions faced by *E. coli* during its existence within the GI tract. This study generated lists of candidate genes putatively required for fitness under these conditions, providing many new insights for further study. However, this study also explored the use of TraDIS in a number of novel ways, e.g in the analysis of mutants with increased fitness values, in addressing the challenge of validating TraDIS data, and in demonstrating how TraDIS can be combined with other analytical approaches.

Mutants are often described in terms of their 'loss-of-function', i.e. where mutation of a gene reduces growth or fitness of the mutant. However, this work placed particular emphasis on the analysis of mutants displaying increased fitness (e.g. see section 4.1.3, Chapter 4), not least due to the fact that genes with positive logFC values comprised a substantial proportion of the genes under selection (e.g. 36% in the case of genes under selection in bile). Previous TraDIS studies have generally not conducted a substantial analysis on mutants displaying increased fitness (see Table 1.2, Chapter 1; see Stocks *et al.* for a recent example (Stocks *et al.*, 2019)). However, this study has shown how investigating these mutants should be standard practice for fitness-related transposon sequencing studies. For example, mutants with positive fitness values can help to validate identify regulatory connections between genes and validate mutants with negative fitness values e.g. the positive selection on *acrR* in the presence of bile due to the derepression of *acrAB* in *acrR* mutants (see section 4.1.3, Chapter 4). Moreover, mutations causing increased fitness can offer unique insights into mechanisms of fitness that may not be identified in the list of negatively selected mutants e.g. the σ^E-encoding *rpoE* is essential in *E. coli* (Goodall *et al.*, 2018) and so mutant-phenotype associations cannot be determined for this gene, but the positive selection for *skp* mutants may be due to *skp* mutants having an induced σ^E regulon, suggesting some role for the sigma factor in bile tolerance (see section 4.1.3.2, Chapter 4).

In high-throughput experimental approaches such as RNA-seq, validation is necessary in the instance where the expression of specific

genes or groups of genes needs to be verified. Typically, validation requires comparison of RNA-seq data with gene transcript levels obtained using a different approach such as qRT-PCR (Fang & Cui, 2011). However, when using RNA-seq to compare global gene expression levels, validation by qRT-PCR is not considered to be necessary. Rather, validation should derive from the use of appropriate numbers of technical and biological replicates in the RNAseq experiment (Fang & Cui, 2011). This study used fitness data from TraDIS to analyse global genetic requirements for growth; therefore, library screens in bile and under anaerobic conditions were conducted in biological duplicate. However, the fitness requirements for specific genes and groups of genes were also validated using a number of different approaches. These included comparison with previous library screens (e.g. Chapter 3), end-point growth of equivalent whole-gene knockout mutants (e.g. see section 4.2, Chapter 4), and competition assays (e.g. Figure 4.4). No single validation method was completely congruent with the TraDIS fitness data, but this is to be expected since the selective pressures that exist within the pooled mutant library screen are different to other experiments, e.g. the population pressures exerted by the growth of mutant and/or the wild type alone vs. the population pressures exerted by a heterogeneous population of 900,000+ mutants in a pooled library. However, the combining of results from the different validation approaches could offer insights into why particular genes were selected during TraDIS. For example, in the presence of bile, Δ*dsbB* mutants showed no significant increase in competitiveness during competition assays against the wild-

type (Figure 4.3B), but did show significantly increased end-point growth (Table 4.5), suggesting that *dsbB* mutants may have been at a competitive advantage in the pooled library by being able to grow to higher cell densities. Therefore, this study contributes to future transposon sequencing studies by highlighting that the verification of specific genes or groups of genes can be done using a multi-method analysis of whole-gene deletion mutants.

Finally, this study illustrated that results from TraDIS could be combined with other types of data. Specifically, it was shown that logFC values obtained in TraDIS could map onto metabolic flux values (see section 5.3.2, Chapter 5). This allowed for greater insights into the phenotypes underlying logFC values as well as a description of the importance of metabolic fluxes for fitness, regardless of how great or small the flux was. This may be of particular use for metabolic engineering since it would allow for a global analysis of metabolic bottlenecks or other barriers to optimising yields of particular products due to poor cell survival. It would be interesting to conduct metabolomics and transposon sequencing simultaneously in order to both validate this observation and explore its potential in metabolic engineering applications.

In conclusion, this work outlines the use of TraDIS to describe, in detail, the genetic requirements of *E. coli* for growth under a variety of different conditions relevant to colonisation of the GI tract. In doing so, the power and versatility of TraDIS as an analytical technique was

highlighted, and novel uses and interpretations of TraDIS data that could be of significant use in future studies was described.

Bibliography

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721–5732.

Alhagamhmad, M. H., Day, A. S., Lemberg, D. A. & Leach, S. T. (2016). An overview of the bacterial contribution to Crohn disease pathogenesis. *J Med Microbiol* **65**, 1049–1059.

Amir, I., Konikoff, F. M., Oppenheim, M., Gophna, U. & Half, E. E. (2013). Gastric microbiota is altered in oesophagitis and Barrett's oesophagus and further modified by proton pump inhibitors. *Environ Microbiol* **16**, 2905–2914.

Apperloo-Renkema, H. Z., Van der Waaij, B. D. & Van der Waaij, D. (1990). Determination of colonization resistance of the digestive tract by biotyping of Enterobacteriaceae. *Epidemiol Infect* **105**, 355–361.

Aquino, P., Honda, B., Jaini, S., Lyubetskaya, A., Hosur, K., Chiu, J. G., Ekladious, I., Hu, D., Jin, L. & other authors. (2017). Coordinated regulation of acid resistance in *Escherichia coli*. *BMC Syst Biol* **11**, 1.

Asai, T., Zaporojets, D., Squires, C. & Squires, C. L. (1999). An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. *Proc Natl Acad Sci U S A* **96**, 1971–1976.

Audia, J. P., Webb, C. C. & Foster, J. W. (2001). Breaking through the acid barrier: An orchestrated response to proton stress by enteric bacteria. *Int J Med Microbiol* **291**, 97–106.

Van Avondt, K., Sorge, N. M. van & Meyaard, L. (2015). Bacterial Immune Evasion through Manipulation of Host Inhibitory Immune Signaling. *PLOS Pathog* **11**, e1004644.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. a, Tomita, M., Wanner, B. L. & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008.

Bachmann, V., Kostiuk, B., Unterweger, D., Diaz-Satizabal, L., Ogg, S. & Pukatzki, S. (2015). Bile Salts Modulate the Mucin-Activated Type VI Secretion System of Pandemic Vibrio cholerae. *PLoS Negl Trop Dis* **9**, e0004031.

Bailey, J. K., Pinyon, J. L., Anantham, S. & Hall, R. M. (2010). Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *J Med Microbiol* **59**, 1331–1339.

Baker, T. A. & Sauer, R. T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* **1823**, 15–28.

Baranova, N. & Nikaido, H. (2002). The baeSR two-component regulatory system activates transcription of the *yegMNOB* (*mdtABCD*) transporter gene cluster in *Escherichia coli* and

increases its resistance to novobiocin and deoxycholate. *J Bacteriol* **184**, 4168–4176.

Barquist, L., Boinett, C. J. & Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol* **10**, 1161–1169.

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., Langridge, G. C., Quail, M. A., Keane, J. A. & Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109–1111.

Baydaş, B., Uslu, H., Yavuz, İ., Ceylan, İ. & Dağsuyu, İ. M. (2007). Effect of a chronic nail-biting habit on the oral carriage of Enterobacteriaceae. *Oral Microbiol Immunol* **22**, 1–4.

Begley, M., Gahan, C. G. M. & Hill, C. (2005). The interaction between bacteria and bile. *FEMS Microbiol Rev* **29**, 625–651.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acids Res* **41**, D36-42.

Berg, R. D. (1995). Bacterial translocation from the gastrointestinal tract. *Trends Microbiol* **3**, 149–154.

Berg, R. D. (1996). The indigenous gastrointestinal microflora. *Trends Microbiol* **4**, 430–435.

Berkmen, M., Boyd, D. & Beckwith, J. (2005). The Nonconsecutive Disulfide Bond of *Escherichia coli* Phytase (AppA) Renders It

Dependent on the Protein-disulfide Isomerase, DsbC. *J Biol Chem* **280**, 11387–11394.

Bernstein, C., Bernstein, H., Payne, C. M., Beard, S. E. & Schneider, J. (1999). Bile Salt Activation of Stress Response Promoters in *Escherichia coli*. *Curr Microbiol* **39**, 68–72.

Bettelheim, K. & Lennox-King, S. (1976). The acquisition of *Escherichia coli* by new-born babies. *Infection* **4**, 174–179.

Birchenough, G. M. H., Johansson, M. E. V, Stabler, R. A., Dalgakiran, F., Hansson, G. C., Wren, B. W., Luzio, J. P. & Taylor, P. W. (2013). Altered Innate Defenses in the Neonatal Gastrointestinal Tract in Response to Colonization by Neuropathogenic *Escherichia coli*. *Infect Immun* **81**, 3264.

Blankenberg, D., Nekrutenko, A., Gordon, A., Von Kuster, G., Taylor, J. & Coraor, N. (2010). Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783–1785.

Brown, A., Fernandez, I. S., Gordiyenko, Y. & Ramakrishnan, V. (2016). Ribosome-dependent activation of stringent control. *Nature* **534**, 277–280.

Bulieris, P. V, Behrens, S., Holst, O. & Kleinschmidt, J. H. (2003). Folding and Insertion of the Outer Membrane Protein OmpA Is Assisted by the Chaperone Skp and by Lipopolysaccharide. *J Biol Chem* **278**, 9092– 9099.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. (2012). Artemis: an integrated platform for visualization and

analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469.

Chang, D. E., Shin, S., Rhee, J. S. & Pan, J. G. (1999). Acetate metabolism in a *pta* mutant of *Escherichia coli* W3110: importance of maintaining acetyl coenzyme A flux for growth and survival. *J Bacteriol* **181**, 6656–6663.

Chaudhuri, R. R., Morgan, E., Peters, S. E., Pleasance, S. J., Hudson, D. L., Davies, H. M., Wang, J., van Diemen, P. M., Buckley, A. M. & other authors. (2013). Comprehensive assignment of roles for *Salmonella typhimurium* genes in intestinal colonization of food-producing animals. *PLoS Genet* **9**, e1003456–e1003456.

Chipman, D. M. & Shaanan, B. (2001). The ACT domain family. *Curr Opin Struct Biol* **11**, 694–700.

Clements, A., Tull, D., Jenney, A. W., Farn, J. L., Kim, S.-H., Bishop, R. E., McPhee, J. B., Hancock, R. E. W., Hartland, E. L. & other authors. (2007). Secondary acylation of *Klebsiella pneumoniae* lipopolysaccharide contributes to sensitivity to antibacterial peptides. *J Biol Chem* **282**, 15569–15577.

Cole, J. A. & Richardson, D. J. (2013). Respiration of Nitrate and Nitrite. *EcoSal Plus.*

Cole, S. T., Eiglmeier, K., Ahmed, S., Honore, N., Elmes, L., Anderson, W. F. & Weiner, J. H. (1988). Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic sn-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *J Bacteriol* **170**, 2448–2456.

Conlon, M. A. & Bird, A. R. (2014). The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7**, 17–44.

Constantinidou, C., Hobman, J. L., Griffiths, L., Patel, M. D., Penn, C. W., Cole, J. A. & Overton, T. W. (2006). A Reassessment of the FNR Regulon and Transcriptomic Analysis of the Effects of Nitrate, Nitrite, NarXL, and NarQP as *Escherichia coli* K12 Adapts from Aerobic to Anaerobic Growth. *J Biol Chem* **281**, 4802–4815.

Conway, T., Krogfelt, K. A. & Cohen, P. S. (2013). The Life of Commensal *Escherichia coli* in the Mammalian Intestine. *EcoSal Plus*.

Conway, T. & Cohen, P. S. (2015). Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbiol Spectr* **3**, 1–15.

Cowley, L. A., Low, A. S., Pickard, D., Boinett, C. J., Dallman, T. J., Day, M., Perry, N., Gally, D. L., Parkhill, J. & other authors. (2018). Transposon Insertion Sequencing Elucidates Novel Gene Involvement in Susceptibility and Resistance to Phages T4 and T7 in *Escherichia coli* O157. *MBio* **9** e00705-18.

Crawford, R. W., Keestra, A. M., Winter, S. E., Xavier, M. N., Tsolis, R. M., Tolstikov, V. & Bäumler, A. J. (2012). Very Long O-antigen Chains Enhance Fitness during *Salmonella*-induced Colitis by Increasing Bile Resistance. *PLOS Pathog* **8**, e1002918.

Cremer, J., Segota, I., Yang, C.-Y., Arnoldini, M., Sauls, J. T., Zhang, Z., Gutierrez, E., Groisman, A. & Hwa, T. (2016). Effect of flow and peristaltic mixing on bacterial growth in a gut-like channel. *Proc Natl Acad Sci U S A* **113**, 11414–11419.

Cress, B. F., Englaender, J. A., He, W., Kasper, D., Linhardt, R. J. & Koffas, M. A. G. (2014). Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol Rev* **38**, 660–697.

D'Aldebert, E., Biyeyeme Bi Mve, M., Mergey, M., Wendum, D., Firrincieli, D., Coilly, A., Fouassier, L., Corpechot, C., Poupon, R. & other authors. (2009). Bile Salts Control the Antimicrobial Peptide Cathelicidin Through Nuclear Receptors in the Human Biliary Epithelium. *Gastroenterology* **136**, 1435–1443.

D'Mello, A. & Yotis, W. W. (1987). The action of sodium deoxycholate on *Escherichia coli*. *Appl Environ Microbiol* **53**, 1944 LP – 1946.

Dalebroux, Z. D., Edrozo, M. B., Pfuetzner, R. A., Ressl, S., Kulasekara, B. R., Blanc, M.-P. & Miller, S. I. (2015). Delivery of cardiolipins to the *Salmonella* outer membrane is necessary for survival within host tissues and virulence. *Cell Host Microbe* **17**, 441–451.

Dartigalongue, C., Missiakas, D. & Raina, S. (2001). Characterization of the *Escherichia coli* ςE Regulon. *J Biol Chem* **276**, 20866–20875.

Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640–6645.

Dautin, N. (2010). Serine protease autotransporters of enterobacteriaceae (SPATEs): biogenesis and function. *Toxins* **2**, 1179–1206.

Deckers, D., Vanlint, D., Callewaert, L., Aertsen, A. & Michiels, C. W. (2008). Role of the Lysozyme Inhibitor Ivy in Growth or Survival of *Escherichia coli* and *Pseudomonas aeruginosa* Bacteria in Hen Egg White and in Human Saliva and Breast Milk. *Appl Environ Microbiol* **74**, 4434–4439.

Deutscher, J., Aké, F. M. D., Derkaoui, M., Zébré, A. C., Cao, T. N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E. & Joyet, P. (2014). The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiol Mol Biol Rev* **78**, 231–256.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.- H., Lakshmanan, A. & Wade, W. G. (2010). The Human Oral Microbiome. *J Bacteriol* **192**, 5002–5017.

Ding, J. W., Andersson, R., Soltesz, V., Willén, R. & Bengmark, S. (1993). The Role of Bile and Bile Acids in Bacterial Translocation in Obstructive Jaundice in Rats. *Eur Surg Res* **25**, 11–19.

Dixit, S. M., Gordon, D. M., Wu, X.-Y., Chapman, T., Kailasapathy, K. & Chin, J. J.-C. (2004). Diversity analysis of commensal porcine *Escherichia coli* – associations between genotypes and habitat in the porcine gastrointestinal tract. *Microbiology* **150**, 1735–1740.

Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* **14**, 20–32.

Dougan, G. & Baker, S. (2014). *Salmonella enterica* Serovar Typhi and the Pathogenesis of Typhoid Fever. *Annu Rev Microbiol* **68**, 317–336.

Dykhuizen, R. S., Masson, J., McKnight, G., Mowat, A. N., Smith, C. C., Smith, L. M. & Benjamin, N. (1996). Plasma nitrate concentration in infective gastroenteritis and inflammatory bowel disease. *Gut* **39**, 393–395.

Eckert, S. E., Dziva, F., Chaudhuri, R. R., Langridge, G. C., Turner, D. J., Pickard, D. J., Maskell, D. J., Thomson, N. R. & Stevens, M. P. (2011). Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn*5*Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. *J Bacteriol* **193**, 1771–1776.

Eguchi, Y., Okada, T., Minagawa, S., Oshima, T., Mori, H., Yamamoto, K., Ishihama, A. & Utsumi, R. (2004). Signal transduction cascade between EvgA/EvgS and PhoP/PhoQ two-component systems of *Escherichia coli*. *J Bacteriol* **186**, 3006–3014.

Eicher, T., Brandstätter, L. & Pos, K. M. (2009). Structural and functional aspects of the multidrug efflux pump AcrB. *Biol Chem* **390**, 693–699.

Ellwood, M. & Nomura, M. (1980). Deletion of a ribosomal ribonucleic acid operon in *Escherichia coli*. *J Bacteriol* **143**, 1077–1080.

Enjalbert, B., Cocaign-Bousquet, M., Portais, J.-C. & Letisse, F. (2015). Acetate Exposure Determines the Diauxic Behavior of

Escherichia coli during the Glucose-Acetate Transition. *J Bacteriol* **197**, 3173–3181.

Evans Doyle J., J. & Evans, D. G. (1983). Classification of Pathogenic *Escherichia coli* According to Serotype and the Production of Virulence Factors, with Special Reference to Colonization-Factor Antigens. *Rev Infect Dis* **5**, S692–S701.

Faber, F. & Bäumler, A. J. (2014). The impact of intestinal inflammation on the nutritional environment of the gut microbiota. *Immunol Lett* **162**, 48–53.

Fabich, A. J., Jones, S. a, Chowdhury, F. Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J. W., Hightower, G. A., Smith, J. T. & other authors. (2008). Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect Immun* **76**, 1143–52.

Fagan, R. P., Lambert, M. A. & Smith, S. G. J. (2008). The *hek* outer membrane protein of *Escherichia coli* strain RS218 binds to proteoglycan and utilizes a single extracellular loop for adherence, invasion, and autoaggregation. *Infect Immun* **76**, 1135–1142.

Fang, Z. & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Brief Bioinform* **12**, 280–287.

Foster, J. W. (2004). *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2**, 898–907.

Foster, P. L. (2007). Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* **42**, 373–397.

Fraser, H. I., Kvaratskhelia, M. & White, M. F. (1999). The two analogous phosphoglycerate mutases of *Escherichia coli*. *FEBS Lett* **455**, 344–348.

Freed, N. E., Bumann, D. & Silander, O. K. (2016). Combining *Shigella* Tn-seq data with gold-standard *E. coli* gene deletion data suggests rare transitions between essential and non-essential gene functionality. *BMC Microbiol* **16**, 203.

Friman, V., Adlerberth, I., Connell, H., Svanborg, C., Hanson, L. A. & Wold, A. E. (1996). Decreased expression of mannose-specific adhesins by *Escherichia coli* in the colonic microflora of immunoglobulin A-deficient individuals. *Infect Immun* **64**, 2794 –2798.

Froelich, J. M., Tran, K. & Wall, D. (2006). A *pmrA* Constitutive Mutant Sensitizes *Escherichia coli* to Deoxycholic Acid. *J Bacteriol* **188**, 1180–1183.

Fröhlich, K. S., Papenfort, K., Berger, A. A. & Vogel, J. (2012). A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res* **40**, 3623–3640.

Fujisawa, T. & Mori, M. (1996). Influence of bile salts on β-glucuronidase activity of intestinal bacteria. *Lett Appl Microbiol* **22**, 271–274.

Fukiya, S., Mizoguchi, H., Tobe, T. & Mori, H. (2004). Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* Strains revealed by comparative genomic hybridization microarray. *J Bacteriol* **186**, 3911–3921.

Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E. & Tenaillon, O. (2007). Extraintestinal Virulence Is a Coincidental By-Product of Commensalism in B2 Phylogenetic Group *Escherichia coli* Strains. *Mol Biol Evol* **24**, 2373–2384.

Gausing, K. (1981). Rates of growth, ribosome synthesis and elongation factor synthesis in a *tufA* defective strain of *E. coli. Mol Gen Genet* **184**, 272–277.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I. & other authors. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185**, 5673–5684.

Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. (2019). The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res* **47**, 2446–2454.

Glode, M. P., Sutton, A., Robbins, J. B., McCracken, G. H., Gotschlich, E. C., Kaijser, B. & Hanson, L. A. (1977). Neonatal Meningitis Due to *Escherichia coli* K1. *J Infect Dis* **136**, s93–s97.

Goh, E.-B., Bledsoe, P. J., Chen, L.-L., Gyaneshwar, P., Stewart, V. & Igo, M. M. (2005). Hierarchical Control of Anaerobic Gene Expression in *Escherichia coli* K-12: the Nitrate-Responsive NarX-NarL Regulatory System Represses Synthesis of the Fumarate-Responsive DcuS-DcuR Regulatory System. *J Bacteriol* **187**, 4890–4899.

Goh, K. G. K., Phan, M.-D., Forde, B. M., Chong, T. M., Yin, W.-F., Chan, K.-G., Ulett, G. C., Sweet, M. J., Beatson, S. A. & Schembri, M. A. (2017). Genome-Wide Discovery of Genes Required for Capsule Production by Uropathogenic *Escherichia coli*. *MBio* **8**, e01558-17.

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A. & Henderson, I. R. (2018). The Essential Genome of *Escherichia coli* K-12. *MBio* **9**, e02096-17.

Gordon, D. M. (2004). The Influence of Ecological Factors on the Distribution and the Genetic Structure of *Escherichia coli*. *EcoSal Plus*.

Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R. & Reznikoff, W. S. (2000). Insertional transposon mutagenesis by electroporation of released Tn*5* transposition complexes. *Nat Biotechnol* **18**, 97–100.

Gray, A. N., Koo, B.-M., Shiver, A. L., Peters, J. M., Osadnik, H. & Gross, C. A. (2015). High-throughput bacterial functional genomics in the sequencing era. *Curr Opin Microbiol* **27**, 86–95.

Grenier, F., Matteau, D., Baby, V. & Rodrigue, S. (2014). Complete Genome Sequence of *Escherichia coli* BW25113. *Genome Announc* **2**, e01038-14.

Grenov, A. I. & Gerdes, S. Y. (2008). Modeling competitive outgrowth of mutant populations: why do essentiality screens yield divergent results? *Methods Mol Biol* **416**, 361–367.

Griess, J. P. & Bemerkungen, Z. A. H. H. (1879). Über einige Azoverbindungen. *Ber Deutch Chem Ges* **12**, 426–428.

Gruenheid, S. & Le Moual, H. (2012). Resistance to antimicrobial peptides in Gram-negative bacteria. *FEMS Microbiol Lett* **330**, 81–89.

Guaraldi, F. & Salvatori, G. (2012). Effect of breast and formula feeding on gut microbiota shaping in newborns. *Front Cell Infect Microbiol* 2, 94.

Guest, R. L., Wang, J., Wong, J. L. & Raivio, T. L. (2017). A Bacterial Stress Response Regulates Respiratory Protein Complexes To Control Envelope Stress Adaptation. *J Bacteriol* **199**, e00153-17.

Gutierrez, A., Laureti, L., Crussard, S., Abida, H., Rodríguez-Rojas, A., Blázquez, J., Baharoglu, Z., Mazel, D., Darfeuille, F. & other authors. (2013). β-Lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat Commun* **4**, 1610.

Hamner, S., McInnerney, K., Williamson, K., Franklin, M. J. & Ford, T. E. (2013). Bile salts affect expression of *Escherichia coli* O157:H7 genes for virulence and iron acquisition, and promote growth under iron limiting conditions. *PLoS One* **8**, e74647–e74647.

Hansen, A.-M., Qiu, Y., Yeh, N., Blattner, F. R., Durfee, T. & Jin, D. J. (2005). SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. *Mol Microbiol* **56**, 719–734.

Hay, A. J. & Zhu, J. (2016). In Sickness and in Health: The Relationships Between Bacteria and Bile in the Human Gut. *Adv Appl Microbiol*, **96,** 43–64.

Hernández, S. B., Cota, I., Ducret, A., Aussel, L. & Casadesús, J. (2012). Adaptation and Preadaptation of *Salmonella enterica* to Bile. *PLOS Genet* **8**, e1002459.

Hofmann, A. F. (1999). Bile Acids: The Good, the Bad, and the Ugly. *Physiology* **14**, 24–29.

Hollister, E. B., Gao, C. & Versalovic, J. (2014). Compositional and functional features of the gastrointestinal microbiome and their effects on human health. *Gastroenterology* **146**, 1449–1458.

Hu, L. I., Lima, B. P. & Wolfe, A. J. (2010). Bacterial protein acetylation: the dawning of a new age. *Mol Microbiol* **77**, 15–21.

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214.

Inagaki, T., Moschetta, A., Lee, Y.-K., Peng, L., Zhao, G., Downes, M., Yu, R. T., Shelton, J. M., Richardson, J. A. & other authors. (2006). Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proc Natl Acad Sci U S A* **103**, 3920–3925.

Islam, K. B. M. S., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T. & Yokota, A. (2011). Bile Acid Is a Host Factor That Regulates the Composition of the Cecal Microbiota in Rats. *Gastroenterology* **141**, 1773–1781.

Ito, M., Baba, T., Mori, H. & Mori, H. (2005). Functional analysis of 1440 *Escherichia coli* genes using the combination of knock-out library and phenotype microarrays. *Metab Eng* **7**, 318–327.

Jacobsen, L., Durso, L., Conway, T. & Nickerson, K. W. (2009). *Escherichia coli* O157:H7 and Other *E. coli* Strains Share Physiological Properties Associated with Intestinal Colonization. *Appl Environ Microbiol* **75**, 4633–4635.

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M. & Nageshwar Reddy, D. (2015). Role of the normal gut microbiota. *World J Gastroenterol* **21**, 8787–8803.

Jarchow, S., Lück, C., Görg, A. & Skerra, A. (2008). Identification of potential substrate proteins for the periplasmic *Escherichia coli* chaperone Skp. *Proteomics* **8**, 4987–4994.

Johnson, M. E. & Rajagopalan, K. V. (1987). Involvement of *chlA*, *E*, *M*, and *N* loci in *Escherichia coli* molybdopterin biosynthesis. *J Bacteriol* **169**, 117–125.

Jones, S. A., Gibson, T., Maltby, R. C., Chowdhury, F. Z., Stewart, V., Cohen, P. S. & Conway, T. (2011). Anaerobic Respiration of *Escherichia coli* in the Mouse Intestine. *Infect Immun* **79**, 4218-4226.

Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S. A., Palsson, B. O. & Agarwalla, S. (2006). Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J Bacteriol* **188**, 8259–8271.

Kaas, R. S., Friis, C., Ussery, D. W. & Aarestrup, F. M. (2012). Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* **13**, 577.

Kabir, M. M., Ho, P. Y. & Shimizu, K. (2005). Effect of *ldhA* gene deletion on the metabolism of *Escherichia coli* based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution. *Biochem Eng J* **26**, 1–11.

Kaiser, M. & Sawers, G. (1994). Pyruvate formate-lyase is not essential for nitrate respiration by *Escherichia coli*. *FEMS Microbiol Lett* **117**, 163-168.

Kakkanat, A., Phan, M.-D., Lo, A. W., Beatson, S. A. & Schembri, M. A. (2017). Novel genes associated with enhanced motility of *Escherichia coli* ST131. *PLoS One* **12**, e0176290.

Kaper, J. B., Nataro, J. P. & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**, 123–40.

Karami, N., Martner, A., Enne, V. I., Swerkersson, S., Adlerberth, I. & Wold, A. E. (2007). Transfer of an ampicillin resistance gene between two *Escherichia coli* strains in the bowel microbiota of an infant treated with antibiotics . *J Antimicrob Chemother* **60**, 1142–1145.

Kargeti, M. & Venkatesh, K. V. (2017). The effect of global transcriptional regulators on the anaerobic fermentative metabolism of *Escherichia coli*. *Mol Biosyst* **13**, 1388–1398.

Karow, M. & Georgopoulos, C. (1992). Isolation and characterization of the *Escherichia coli msbB* gene, a multicopy suppressor of null mutations in the high-temperature requirement gene *htrB*. *J Bacteriol* **174**, 702–710.

Kaya, M., Beştaş, R., Bacalan, F., Bacaksız, F., Arslan, E. G. & Kaplan, M. A. (2012). Microbial profile and antibiotic sensitivity pattern in bile cultures from endoscopic retrograde cholangiography patients. *World J Gastroenterol* **18**, 3585–3589.

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides- Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A. & other authors. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, 543–550.

Kim, H. J., Li, H., Collins, J. J. & Ingber, D. E. (2016). Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proc Natl Acad Sci U S A* **113**, 7–15.

Kim, K. S. (2003). Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury. *Nat Rev Neurosci* **4**, 376–385.

Kim, M., Ashida, H., Ogawa, M., Yoshikawa, Y., Mimuro, H. & Sasakawa, C. (2010). Bacterial Interactions with the Host Epithelium. *Cell Host Microbe* **8**, 20–35.

Kingsley, R. A., Langridge, G., Smith, S. E., Makendi, C., Fookes, M., Wileman, T. M., El Ghany, M. A., Keith Turner, A., Dyson, Z. A.

& other authors. (2018). Functional analysis of *Salmonella* Typhi adaptation to survival in water. *Environ Microbiol* **20**, 4079–4090.

Klein, G., Kobylak, N., Lindner, B., Stupak, A. & Raina, S. (2014). Assembly of lipopolysaccharide in *Escherichia coli* requires the essential LapB heat shock protein. *J Biol Chem* **289**, 14829–14853.

Van de Klundert, J. A., Van der Meide, P. H., Van de Putte, P. & Bosch, L. (1978). Mutants of *Escherichia coli* altered in both genes coding for the elongation factor Tu. *Proc Natl Acad Sci U S A* **75**, 4470–4473.

Korepanov, A. P., Gongadze, G. M., Garber, M. B., Court, D. L. & Bubunenko, M. G. (2007). Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in *Escherichia coli*. *J Mol Biol* **366**, 1199–1208.

Kotlowski, R., Bernstein, C. N., Sepehri, S. & Krause, D. O. (2007). High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* **56**, 669–675.

Kotnik, M., Anderluh, P. S. & Prezelj, A. (2007). Development of novel inhibitors targeting intracellular steps of peptidoglycan biosynthesis. *Curr Pharm Des* **13**, 2283–2309.

Kramer, V. C., Nickerson, K. W., Hamlett, N. V & O'Hara, C. (1984). Prevalence of extreme detergent resistance among the Enterobacteriaceae. *Can J Microbiol* **30**, 711–713.

Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V & Mushegian, A. (2010). A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**, 1481–1487.

Kus, J. V, Gebremedhin, A., Dang, V., Tran, S.-L., Serbanescu, A. & Barnett Foster, D. (2011). Bile salts induce resistance to polymyxin in enterohemorrhagic *Escherichia coli* O157:H7. *J Bacteriol* **193**, 4509–4515.

Lampe, D. J., Grant, T. E. & Robertson, H. M. (1998). Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*. *Genetics* **149**, 179–187.

Landeta, C., Boyd, D. & Beckwith, J. (2018). Disulfide bond formation in prokaryotes. *Nat Microbiol* **3**, 270–280.

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E. & other authors. (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* **19**, 2308–2316.

De Lay, N. R. & Cronan, J. E. (2008). Genetic interaction between the *Escherichia coli* AcpT phosphopantetheinyl transferase and the YejM inner membrane protein. *Genetics* **178**, 1327–1337.

Lazazzera, B. A., Beinert, H., Khoroshilova, N., Kennedy, M. C. & Kiley, P. J. (1996). DNA binding and dimerization of the Fe-S-

containing FNR protein from *Escherichia coli* are regulated by oxygen. *J Biol Chem* **271**, 2762–2768.

Lee, A., Mao, W., Warren, M. S., Mistry, A., Hoshino, K., Okumura, R., Ishida, H. & Lomovskaya, O. (2000). Interplay between Efflux Pumps May Provide Either Additive or Multiplicative Effects on Drug Resistance. *J Bacteriol* **182**, 3142–3150.

Leimbach, A., Hacker, J. & Dobrindt, U. (2013). *E. coli* as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. *Between Pathog Commensalism*, pp. 3–32.

Leverrier, P., Dimova, D., Pichereau, V., Auffray, Y., Boyaval, P. & Jan, G. (2003). Susceptibility and adaptive response to bile salts in *Propionibacterium freudenreichii*: physiological and proteomic analysis. *Appl Environ Microbiol* **69**, 3809–3818.

Li, G., Yang, M., Zhou, K., Zhang, L., Tian, L., Lv, S., Jin, Y., Qian, W., Xiong, H. & other authors. (2015). Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. *J Microbiol Biotechnol* **25**, 1136–1145.

Lin, J., Lee, I. S., Frey, J., Slonczewski, J. L. & Foster, J. W. (1995). Comparative analysis of extreme acid survival in *Salmonella typhimurium*, *Shigella flexneri*, and *Escherichia coli*. *J Bacteriol* **177**, 4097–4104.

Lin, J., Smith, M. P., Chapin, K. C., Baik, H. S., Bennett, G. N. & Foster, J. W. (1996). Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. *Appl Environ Microbiol* **62**, 3094–3100.

Long, C. P. & Antoniewicz, M. R. (2014). Metabolic flux analysis of *Escherichia coli* knockouts: lessons from the Keio collection and future outlook. *Curr Opin Biotechnol* **28**, 127–133.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S. H., Flint, H. J., Aldeguer, X. & Garcia-Gil, L. J. (2014). Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* **304**, 464–475.

Lopez, C. A., Winter, S. E., Rivera-Chávez, F., Xavier, M. N., Poon, V., Nuccio, S.-P., Tsolis, R. M. & Bäumler, A. J. (2012). Phage-Mediated Acquisition of a Type III Secreted Effector Protein Boosts Growth of *Salmonella* by Nitrate Respiration. *MBio* **3**, e00143-12.

Lorenzo-Zúñiga, V., Bartolí, R., Planas, R., Hofmann, A. F., Viñado, B., Hagey, L. R., Hernández, J. M., Mañé, J., Alvarez, M. A. & other authors. (2003). Oral bile acids reduce bacterial overgrowth, bacterial translocation, and endotoxemia in cirrhotic rats. *Hepatology* **37**, 551–557.

Ludden, C., Raven, K. E., Jamrozy, D., Gouliouris, T., Blane, B., Coll, F., de Goffau, M., Naydenova, P., Horner, C. & other authors. (2019). One Health Genomic Surveillance of *Escherichia coli* Demonstrates Distinct Lineages and Mobile Genetic Elements in Isolates from Humans versus Livestock. *MBio* **10** e02693-18.

Lukjancenko, O., Wassenaar, T. M. & Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* **60**, 708–720.

Lundell, A.-C., Björnsson, V., Ljung, A., Ceder, M., Johansen, S., Lindhagen, G., Törnhage, C.-J., Adlerberth, I., Wold, A. E. & Rudin, A. (2012). Infant B Cell Memory Differentiation and Early Gut Bacterial Colonization. *J Immunol* **188**, 4315–4322.

Mahalakshmi, S., Sunayana, M. R., SaiSree, L. & Reddy, M. (2014). *yciM* is an essential gene required for regulation of lipopolysaccharide synthesis in *Escherichia coli*. *Mol Microbiol* **91**, 145–157.

Maltby, R., Leatham-Jensen, M. P., Gibson, T., Cohen, P. S. & Conway, T. (2013). Nutritional Basis for Colonization Resistance by Human Commensal *Escherichia coli* Strains HS and Nissle 1917 against *E. coli* O157:H7 in the Mouse Intestine. *PLoS One* **8**, e53957.

Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D. & Garcia-Gil, J. L. (2006). Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. *Inflamm Bowel Dis* **12**, 1136–1145.

Martinsen, T. C., Bergh, K. & Waldum, H. L. (2005). Gastric Juice: A Barrier Against Infectious Diseases. *Basic Clin Pharmacol Toxicol* **96**, 94–102.

Martorana, A. M., Motta, S., Di Silvestre, D., Falchi, F., Dehò, G., Mauri, P., Sperandeo, P. & Polissi, A. (2014). Dissecting *Escherichia coli* Outer Membrane Biogenesis Using Differential Proteomics. *PLoS One* **9**, e100941.

Mason, K. L. & Huffnagle, G. B. (2009). Control of mucosal polymicrobial populations by innate immunity. *Cell Microbiol* **11**, 1297–1305.

Mat-Jan, F., Alam, K. Y. & Clark, D. P. (1989a). Mutants of *Escherichia coli* deficient in the fermentative lactate dehydrogenase. *J Bacteriol* **171**, 342–348.

Mat-Jan, F., Williams, C. R. & Clark, D. P. (1989b). Anaerobic growth defects resulting from gene fusions affecting succinyl-CoA synthetase in *Escherichia coli* K12. *Mol Gen Genet* **215**, 276–280.

McCarthy, A. J., Stabler, R. A. & Taylor, P. W. (2018). 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** e00698-17.

McGrath, M. E., Gillmor, S. A. & Fletterick, R. J. (1995). Ecotin: lessons on survival in a protease-filled world. *Protein Sci* **4**, 141–148.

Meador, J. P., Caldwell, M. E., Cohen, P. S. & Conway, T. (2014). *Escherichia coli* pathotypes occupy distinct niches in the mouse intestine. *Infect Immun* **82**, 1931–8.

Meredith, T. C. & Woodard, R. W. (2005). Identification of GutQ from *Escherichia coli* as a D-arabinose 5-phosphate isomerase. *J Bacteriol* **187**, 6936–6942.

Merritt, M. E. & Donaldson, J. R. (2009). Effect of bile salts on the DNA and membrane integrity of enteric bacteria. *J Med Microbiol* **58**, 1533–1541.

Mettert, E. L. & Kiley, P. J. (2005). ClpXP-dependent proteolysis of FNR upon loss of its O2-sensing [4Fe-4S] cluster. *J Mol Biol* **354**, 220–232.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. & Thomas, P. D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* **45**, 183–189.

Miajlovic, H., Cooke, N. M., Moran, G. P., Rogers, T. R. F. & Smith, S. G. (2014). Response of extraintestinal pathogenic *Escherichia coli* to human serum reveals a protective role for Rcs-regulated exopolysaccharide colanic acid. *Infect Immun* **82**, 298–305.

Minagawa, S., Ogasawara, H., Kato, A., Yamamoto, K., Eguchi, Y., Oshima, T., Mori, H., Ishihama, A. & Utsumi, R. (2003). Identification and molecular characterization of the Mg2+ stimulon of *Escherichia coli*. *J Bacteriol* **185**, 3696–3702.

Mirsepasi-Lauridsen, H. C., Vallance, B. A., Krogfelt, K. A. & Petersen, A. M. (2019). *Escherichia coli* Pathobionts Associated

with Inflammatory Bowel Disease. *Clin Microbiol Rev* **32**, e00060-18.

Missiakas, D., Betton, J.-M. & Raina, S. (1996). New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol Microbiol* **21**, 871–884.

Møller, A. K., Leatham, M. P., Conway, T., Nuijten, P. J. M., de Haan, L. A. M., Krogfelt, K. A. & Cohen, P. S. (2003). 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.

Monchois, V., Abergel, C., Sturgis, J., Jeudy, S. & Claverie, J.-M. (2001). *Escherichia coli ykfE* Gene Encodes a Potent Inhibitor of C-type Lysozyme. *J Biol Chem* **276**, 18437–18441.

Nalin, D., Rhead, J., Rennels, M., O'Donnell, S., Levine, M., Bergquist, E., Hughes, T. & Hornick, R. (1978). Cannabis, Hyphochlorhydria, and Cholera. *Lancet* **312**, 859–862.

Nicolaes, V., El Hajjaji, H., Davis, R. M., Van der Henst, C., Depuydt, M., Leverrier, P., Aertsen, A., Haufroid, V., Ollagnier de Choudens, S. & other authors. (2014). Insights into the function of YciM, a heat shock membrane protein required to maintain envelope integrity in *Escherichia coli*. *J Bacteriol* **196**, 300–309.

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol Mol Biol Rev* **67**, 593–656.

Noh, D. O. & Gilliland, S. E. (1993). Influence of Bile on Cellular Integrity and β-Galactosidase Activity of *Lactobacillus acidophilus*. *J Dairy Sci* **76**, 1253–1259.

Nolan, L. M., Whitchurch, C. B., Barquist, L., Katrib, M., Boinett, C. J., Mayho, M., Goulding, D., Charles, I. G., Filloux, A. & other authors. (2018). A global genomic approach uncovers novel components for twitching motility-mediated biofilm expansion in *Pseudomonas aeruginosa*. *Microb genomics* **4**.

Nonaka, G., Blankschien, M., Herman, C., Gross, C. A. & Rhodius, V. A. (2006). Regulon and promoter analysis of the *E. coli* heat-shock factor, sigma32, reveals a multifaceted cellular response to heat stress. *Genes Dev* **20**, 1776–1789.

O'Farrell, P. H. (1978). The suppression of defective translation by ppGpp and its role in the stringent response. *Cell* **14**, 545–557.

Ochman, H. & Selander, R. K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* **157**, 690–693.

Ochman, H. & Jones, I. B. (2000). Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J* **19**, 6637–6643.

Oliveira, A. M., Morais, M. B. de & Morais, T. B. (2012). A Novel and Potentially Valuable Exposure Measure: *Escherichia coli* in Oral Cavity and its Association with Child DayCare Center Attendance. *J Trop Pediatr* **58**, 517–520.

Onufryk, C., Crouch, M.-L., Fang, F. C. & Gross, C. A. (2005). Characterization of six lipoproteins in the sigmaE regulon. *J Bacteriol* **187**, 4552–4561.

van Opijnen, T. & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **11**, 435–42.

Overmars, L., van Hijum, S. A. F. T., Siezen, R. J. & Francke, C. (2015). CiVi: circular genome visualization with unique features to analyze sequence elements. *Bioinformatics* **31**, 2867–2869.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F. & Cerf-Bensussan, N. (2011). Trade-Off between Bile Resistance and Nutritional Competence Drives *Escherichia coli* Diversification in the Mouse Gut. *PLOS Genet* **7**, e1002107.

Park, Y.-H., Lee, C.-R., Choe, M. & Seok, Y.-J. (2013). HPr antagonizes the anti-sigma70 activity of Rsd in *Escherichia coli*. *Proc Natl Acad Sci U S A* **110**, 21142–21147.

Paul, S., Alegre, K. O., Holdsworth, S. R., Rice, M., Brown, J. A., McVeigh, P., Kelly, S. M. & Law, C. J. (2014). A single-component multidrug transporter of the major facilitator superfamily is part of a network that protects *Escherichia coli* from bile salt stress. *Mol Microbiol* **92**, 872–884.

Peleg, A. Y. & Hooper, D. C. (2010). Hospital-Acquired Infections Due to Gram-Negative Bacteria. *N Engl J Med* **362**, 1804–1813.

Pettengill, E. A., Pettengill, J. B. & Binet, R. (2016). Phylogenetic Analyses of *Shigella* and Enteroinvasive *Escherichia coli* for the

Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation. *Front Microbiol* **6**, 1573.

Phan, M.-D., Peters, K. M., Sarkar, S., Lukowski, S. W., Allsopp, L. P., Moriel, D. G., Achard, M. E. S., Totsika, M., Marshall, V. M. & other authors. (2013). The Serum Resistome of a Globally Disseminated Multidrug Resistant Uropathogenic *Escherichia coli* Clone. *PLOS Genet* **9**, e1003834.

Di Pilato, V., Freschi, G., Ringressi, M. N., Pallecchi, L., Rossolini, G. M. & Bechi, P. (2016). The esophageal microbiota in health and disease. *Ann N Y Acad Sci* **1381**, 21–33.

Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., Duarte, A. S. R., Black, R. E. & Angulo, F. J. (2015). Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. *PLoS One* **10**, e0142927.

Polissi, A., De Laurentis, W., Zangrossi, S., Briani, F., Longhi, V., Pesole, G. & Deho, G. (2003). Changes in *Escherichia coli* transcriptome during acclimatization at low temperature. *Res Microbiol* **154**, 573–580.

Ponce, E., Flores, N., Martinez, A., Valle, F. & Bolivar, F. (1995). Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. *J Bacteriol* **177**, 5719–5722.

Poulsen, L. K., Lan, F., Kristensen, C. S., Hobolth, P., Molin, S. & Krogfelt, K. A. (1994). Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA in situ hybridization. *Infect Immun* **62**, 5191–5194.

Prohl, C., Wackwitz, B., Vlad, D. & Unden, G. (1998). Functional citric acid cycle in an *arcA* mutant of *Escherichia coli* during growth with nitrate under anoxic conditions. *Arch Microbiol* **170**, 1–7.

Prouty, A. M., Brodsky, I. E., Manos, J., Belas, R., Falkow, S. & Gunn, J. S. (2006). Transcriptional regulation of *Salmonella enterica* serovar Typhimurium genes by bile. *FEMS Immunol Med Microbiol* **41**, 177–185.

Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. (2007). Lipid A Modification Systems in Gram-Negative Bacteria. *Annu Rev Biochem* **76**, 295–329.

Raghavan, R., Sage, A. & Ochman, H. (2011). Genome-wide identification of transcription start sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in *Escherichia coli*. *J Bacteriol* **193**, 2871–2874.

Ramos-Morales, F., Prieto, A. I., Beuzón, C. R., Holden, D. W. & Casadesús, J. (2003). Role for *Salmonella enterica* Enterobacterial Common Antigen in Bile Resistance and Virulence. *J Bacteriol* **185**, 5328 –5332.

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R. & other authors. (2008). The Pangenome Structure of

Escherichia coli: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. *J Bacteriol* **190**, 6881–6893.

Ray, M. C., Germon, P., Vianney, A., Portalier, R. & Lazzaroni, J. C. (2000). Identification by genetic suppression of *Escherichia coli* TolB residues important for TolB-Pal interaction. *J Bacteriol* **182**, 821–824.

Razaghi, M., Tajeddin, E., Ganji, L., Alebouyeh, M., Alizadeh, A. H. M., Sadeghi, A. & Zali, M. R. (2017). Colonization, resistance to bile, and virulence properties of *Escherichia coli* strains: Unusual characteristics associated with biliary tract diseases. *Microb Pathog* **111**, 262–268.

Renner, L. D. & Weibel, D. B. (2012). MinD and MinE interact with anionic phospholipids and regulate division plane formation in *Escherichia coli*. *J Biol Chem* **287**, 38835–38844.

Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. (2014). Bile acids and the gut microbiome. *Curr Opin Gastroenterol* **30**, 332–338.

Rincé, A., Le Breton, Y., Verneuil, N., Giard, J.-C., Hartke, A. & Auffray, Y. (2003). Physiological and molecular aspects of bile salt response in *Enterococcus faecalis*. *Int J Food Microbiol* **88**, 207–213.

Rodionova, I. A., Zhang, Z., Mehla, J., Goodacre, N., Babu, M., Emili, A., Uetz, P. & Saier, M. H. J. (2017). The phosphocarrier protein HPr of the bacterial phosphotransferase system globally regulates

energy metabolism by directly interacting with multiple enzymes in *Escherichia coli*. *J Biol Chem* **292**, 14250–14257.

Rodrigues, M. L., Oliveira, T. F., Pereira, I. A. C. & Archer, M. (2006). X-ray structure of the membrane-bound cytochrome c quinol dehydrogenase NrfH reveals novel haem coordination. *EMBO J* **25**, 5951–5960.

Ruiz, L., Sánchez, B., Ruas-Madiedo, P., De Los Reyes-Gavilán, C. G. & Margolles, A. (2007). Cell envelope changes in *Bifidobacterium animalis* ssp. *lactis* as a response to bile. *FEMS Microbiol Lett* **274**, 316–322.

Ruiz, N., Chng, S.-S., Hiniker, A., Kahne, D. & Silhavy, T. J. (2010). Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proc Natl Acad Sci U S A* **107**, 12245–12250.

Russell, D. W. (2003). The Enzymes, Regulation, and Genetics of Bile Acid Synthesis. *Annu Rev Biochem* **72**, 137–174.

Sabaté, M., Moreno, E., Pérez, T., Andreu, A. & Prats, G. (2006). Pathogenicity island markers in commensal and uropathogenic *Escherichia coli* isolates. *Clin Microbiol Infect* **12**, 880–886.

Sarff, L. D., McCracken, G. H., Schiffer, M. S., Glode, M. P., Robbins, J. B., Orskov, I. & Orskov, F. (1975). Epidemiology of *Escherichia coli* K1 in healthy and diseased newborns. *Lancet* **1**, 1099–1104.

Sarker, S. A. & Gyr, K. (1992). Non-immunological defence mechanisms of the gut. *Gut* **33**, 987–993.

Sassone-Corsi, M. & Raffatellu, M. (2015). No Vacancy: How Beneficial Microbes Cooperate with Immunity To Provide Colonization Resistance to Pathogens. *J Immunol* **194**, 4081–4087.

Savageau, M. A. (1983). *Escherichia coli* Habitats, Cell Types, and Molecular Mechanisms of Gene Control. *Am Nat* **122**, 732–744.

Schäfer, U., Beck, K. & Müller, M. (1999). Skp, a Molecular Chaperone of Gram-negative Bacteria, Is Required for the Formation of Soluble Periplasmic Intermediates of Outer Membrane Proteins. *J Biol Chem* **274**, 24567–24574.

Scheutz, F. & Strockbine, N. A. (2015). *Escherichia*. *Bergey's Man Syst Archaea Bact*.

Schryvers, A., Lohmeier, E. & Weiner, J. H. (1978). Chemical and functional properties of the native and reconstituted forms of the membrane-bound, aerobic glycerol-3-phosphate dehydrogenase of *Escherichia coli*. *J Biol Chem* **253**, 783–788.

Schwalm, J., Mahoney, T. F., Soltes, G. R. & Silhavy, T. J. (2013). Role for Skp in LptD Assembly in *Escherichia coli*. *J Bacteriol* **195**, 3734–3742.

Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. (2010). Gut Microbiota in Health and Disease. *Physiol Rev* **90**, 859–904.

Sezonov, G., Joseleau-Petit, D. & D'Ari, R. (2007). *Escherichia coli* Physiology in Luria-Bertani Broth. *J Bacteriol* **189**, 8746 LP – 8749.

Shouldice, S. R., Heras, B., Walden, P. M., Totsika, M., Schembri, M. A. & Martin, J. L. (2011). Structure and Function of DsbA, a Key

Bacterial Oxidative Folding Catalyst. *Antioxid Redox Signal* **14**, 1729–1760.

Sims, G. E. & Kim, S.-H. (2011). Whole-genome phylogeny of *Escherichia coli/Shigella* group by feature frequency profiles (FFPs). *Proc Natl Acad Sci U S A* **108**, 8329–8334.

Sistrunk, J. R., Nickerson, K. P., Chanin, R. B., Rasko, D. A. & Faherty, C. S. (2016). Survival of the Fittest: How Bacterial Pathogens Utilize Bile To Enhance Infection. *Clin Microbiol Rev* **29**, 819–836.

Skurnik, D., Ruimy, R., Andremont, A., Amorin, C., Rouquet, P., Picard, B. & Denamur, E. (2006). Effect of human vicinity on antimicrobial resistance and integrons in animal faecal *Escherichia coli*. *J Antimicrob Chemother* **57**, 1215–1219.

Snipen, L., Almoy, T. & Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* **10**, 385.

Sperandeo, P., Dehò, G. & Polissi, A. (2009). The lipopolysaccharide transport system of Gram-negative bacteria. *Biochim Biophys Acta - Mol Cell Biol Lipids* **1791**, 594–602.

Sperandeo, P., Villa, R., Martorana, A. M., Samalikova, M., Grandori, R., Dehò, G. & Polissi, A. (2011). New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *J Bacteriol* **193**, 1042–1053.

Stecher, B., Denzler, R., Maier, L., Bernet, F., Sanders, M. J., Pickard, D. J., Barthel, M., Westendorf, A. M., Krogfelt, K. A. & other authors. (2012). Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc Natl Acad Sci U S A* **109**, 1269–1274.

Stewart, V. (1993). Nitrate regulation of anaerobic respiratory gene expression in *Escherichia coli*. *Mol Microbiol* **9**, 425–434.

Stocks, C. J., Phan, M.-D., Achard, M. E. S., Nhu, N. T. K., Condon, N. D., Gawthorne, J. A., Lo, A. W., Peters, K. M., McEwan, A. G. & other authors. (2019). Uropathogenic *Escherichia coli* employs both evasion and resistance to subvert innate immune-mediated zinc toxicity for dissemination. *Proc Natl Acad Sci U S A* **116** (13), 6341-6350.

Su, C.-C., Rutherford, D. J. & Yu, E. W. (2007). Characterization of the multidrug efflux regulator AcrR from Escherichia coli. *Biochem Biophys Res Commun* **361**, 85–90.

Tam, C. & Missiakas, D. (2005). Changes in lipopolysaccharide structure induce the σE-dependent response of *Escherichia coli*. *Mol Microbiol* **55**, 1403–1412.

Tatusov, R. L., Koonin, E. V & Lipman, D. J. (1997). A genomic perspective on protein families. *Science* **278**, 631–637.

Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* **8**, 207–17.

Thanassi, D. G., Cheng, L. W. & Nikaido, H. (1997). Active efflux of bile salts by *Escherichia coli*. *J Bacteriol* **179**, 2512–2518.

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, 158–169.

Thomassin, J.-L., Brannon, J. R., Gibbs, B. F., Gruenheid, S. & Le Moual, H. (2012). OmpT outer membrane proteases of enterohemorrhagic and enteropathogenic *Escherichia coli* contribute differently to the degradation of human LL-37. *Infect Immun* **80**, 483–492.

Thompson, A. P., O'Neill, I., Smith, E. J., Catchpole, J., Fagan, A., Burgess, K. E. V, Carmody, R. J. & Clarke, D. J. (2016). Glycolysis and pyrimidine biosynthesis are required for replication of adherent–invasive *Escherichia coli* in macrophages. *Microbiology* **162**, 954–965.

Tokumoto, U. & Takahashi, Y. (2001). Genetic Analysis of the *isc* Operon in *Escherichia coli* Involved in the Biogenesis of Cellular Iron-Sulfur Proteins. *J Biochem* **130**, 63–71.

Toledano, M. B., Kumar, C., Le Moan, N., Spector, D. & Tacnet, F. (2007). The system biology of thiol redox system in *Escherichia coli* and yeast: differential functions in oxidative stress, iron metabolism and DNA synthesis. *FEBS Lett* **581**, 3598–3607.

Toya, Y., Nakahigashi, K., Tomita, M. & Shimizu, K. (2012). Metabolic regulation analysis of wild-type and *arcA* mutant *Escherichia coli* under nitrate conditions using different levels of omics data. *Mol Biosyst* **8**, 2593–2604.

Tran, Q. H., Arras, T., Becker, S., Holighaus, G., Ohlberger, G. & Unden, G. (2000). Role of glutathione in the formation of the active form of the oxygen sensor FNR ([4Fe-4S].FNR) and in the control of FNR function. *Eur J Biochem* **267**, 4817–4824.

Trchounian, A. & Trchounian, K. (2019). Fermentation Revisited: How Do Microorganisms Survive Under Energy-Limited Conditions? *Trends Biochem Sci.* **44 (5),** 391-400.

Tsang, M.-J., Yakhnina, A. A. & Bernhardt, T. G. (2017). NlpD links cell wall remodeling and outer membrane invagination during cytokinesis in *Escherichia coli*. *PLoS Genet* **13**, e1006888.

Unden, G. & Bongaerts, J. (1997). Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim Biophys Acta - Bioenerg* **1320**, 217–234.

de Valdez, G. F., Martos, G., Taranto, M. P., Lorca, G. L., Oliver, G. & de Ruiz Holgado, A. P. (1997). Influence of Bile on β-Galactosidase Activity and Cell Viability of *Lactobacillus reuteri* when Subjected to Freeze-Drying. *J Dairy Sci* **80**, 1955–1958.

VanOrsdel, C. E., Kelly, J. P., Burke, B. N., Lein, C. D., Oufiero, C. E., Sanchez, J. F., Wimmers, L. E., Hearn, D. J., Abuikhdair, F. J. & other authors. (2018). Identifying New Small Proteins in *Escherichia coli*. *Proteomics* **18**, e1700064.

Veit, A., Polen, T. & Wendisch, V. F. (2007). Global gene expression analysis of glucose overflow metabolism in *Escherichia coli* and

reduction of aerobic acetate formation. *Appl Microbiol Biotechnol* **74**, 406–421.

van Velkinburgh, J. C. & Gunn, J. S. (1999). PhoP-PhoQ-regulated loci are required for enhanced bile resistance in *Salmonella* spp. *Infect Immun* **67**, 1614–1622.

Vertommen, D., Depuydt, M., Pan, J., Leverrier, P., Knoops, L., Szikora, J.-P., Messens, J., Bardwell, J. C. A. & Collet, J.-F. (2008). The disulphide isomerase DsbC cooperates with the oxidase DsbA in a DsbD-independent manner. *Mol Microbiol* **67**, 336–349.

Vieira, G., Sabarly, V., Bourguignon, P.-Y., Durot, M., Le Fevre, F., Mornico, D., Vallenet, D., Bouvet, O., Denamur, E. & other authors. (2011). Core and panmetabolism in *Escherichia coli. J Bacteriol* **193**, 1461–1472.

Vogler, A. P., Trentmann, S. & Lengeler, J. W. (1989). Alternative route for biosynthesis of amino sugars in *Escherichia coli* K-12 mutants by means of a catabolic isomerase. *J Bacteriol* **171**, 6586–6592.

Wang, H. & Gunsalus, R. P. (2000). The *nrfA* and *nirB* nitrite reductase operons in *Escherichia coli* are expressed differently in response to nitrate than to nitrite. *J Bacteriol* **182**, 5813–5822.

Wang, H., Tseng, C. P. & Gunsalus, R. P. (1999). The *napF* and *narG* nitrate reductase operons in *Escherichia coli* are differentially expressed in response to submicromolar concentrations of nitrate but not nitrite. *J Bacteriol* **181**, 5303–5308.

Wang, X., Kim, Y. & Wood, T. K. (2009). Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J* **3**, 1164–1179.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M. & Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**, 147.

Waterman, S. R. & Small, P. L. (1998). Acid-sensitive enteric pathogens are protected from killing under extremely acidic conditions of pH 2.5 when they are inoculated onto certain solid food sources. *Appl Environ Microbiol* **64**, 3882–3886.

Weijland, A., Harmark, K., Cool, R. H., Anborgh, P. H. & Parmeggiani, A. (1992). Elongation factor Tu: a molecular switch in protein biosynthesis. *Mol Microbiol* **6**, 683–688.

Willcocks, S. J., Stabler, R. A., Atkins, H. S., Oyston, P. F. & Wren, B. W. (2018). High-throughput analysis of *Yersinia pseudotuberculosis* gene essentiality in optimised in vitro conditions, and implications for the speciation of *Yersinia pestis*. *BMC Microbiol* **18**, 46.

Williams, M. D., Ouyang, T. X. & Flickinger, M. C. (1994). Starvation-induced expression of SspA and SspB: the effects of a null mutation in *sspA* on *Escherichia coli* protein synthesis and survival during growth and prolonged starvation. *Mol Microbiol* **11**, 1029–1043.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., Tysk, C. & Jansson, J. K. (2009). Twin studies

reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* **15**, 653–660.

Winter, S. E., Winter, M. G., Xavier, M. N., Thiennimitr, P., Poon, V., Keestra, A. M., Laughlin, R. C., Gomez, G., Wu, J. & other authors. (2013). Host-Derived Nitrate Boosts Growth of *E. coli* in the Inflamed Gut. *Science* **339**, 708 – 711.

Witcomb, L. A., Collins, J. W., McCarthy, A. J., Frankel, G. & Taylor, P. W. (2015). Bioluminescent Imaging Reveals Novel Patterns of Colonization and Invasion in Systemic *Escherichia coli* K1 Experimental Infection in the Neonatal Rat. *Infect Immun* **83**, 4528–4540.

Wold, A. E., Caugant, D. A., Lidin-Janson, G., de Man, P. & Svanborg, C. (1992). Resident Colonic *Escherichia coli* Strains Frequently Display Uropathogenic Characteristics. *J Infect Dis* **165**, 46–52.

Wolfe, A. J. (2005). The acetate switch. *Microbiol Mol Biol Rev* **69**, 12–50.

Wright, M. L., Pendarvis, K., Nanduri, B., Edelmann, M. J., Jenkins, H. N., Reddy, J. S., Wilson, J. G., Ding, X., Broadway, P. R. & other authors. (2016). The Effect of Oxygen on Bile Resistance in *Listeria monocytogenes*. *J Proteomics Bioinform* **9**, 107–119.

Wu, H. C. & Wu, T. C. (1971). Isolation and characterization of a glucosamine-requiring mutant of *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. *J Bacteriol* **105**, 455–466.

Yamazaki, Y., Niki, H. & Kato, J. (2008). Profiling of *Escherichia coli* Chromosome database. *Methods Mol Biol* **416**, 385–389.

Yan, X., Zhao, C., Budin-Verneuil, A., Hartke, A., Rince, A., Gilmore, M. S., Auffray, Y. & Pichereau, V. (2009). The (p)ppGpp synthetase RelA contributes to stress adaptation and virulence in *Enterococcus faecalis* V583. *Microbiology* **155**, 3226–3237.

Yu, B. J., Sung, B. H., Lee, J. Y., Son, S. H., Kim, M. S. & Kim, S. C. (2006). *sucAB* and *sucCD* are mutually essential genes in *Escherichia coli*. *FEMS Microbiol Lett* **254**, 245–250.

Yu, C., Cao, Y., Zou, H. & Xian, M. (2011). Metabolic engineering of *Escherichia coli* for biotechnological production of high-value organic acids and alcohols. *Appl Microbiol Biotechnol* **89**, 573–583.

Zhou, J. & Rudd, K. E. (2012). EcoGene 3.0. *Nucleic Acids Res* **41**, 613–624.

Zhu, H., Hart, C. A., Sales, D. & Roberts, N. B. (2006). Bacterial killing in gastric juice – effect of pH and pepsin on *Escherichia coli* and *Helicobacter pylori. J Med Microbiol* **55**, 1265–1270.

Appendix Gene lists and analyses

Table A1. Genes identified as essential in MG1655 following growth on LB agar

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b0023	*rpsT*	20815	21078	-1	0	0	264	0
b0025	*ribF*	21407	22348	1	0	0	942	0
b0026	*ileS*	22391	25207	1	0	0	2817	0
b0027	*lspA*	25207	25701	1	0	0	495	0
b0029	*ispH*	26277	27227	1	0	0	951	0
b0031	*dapB*	28374	29195	1	0	0	822	0
b0048	*folA*	49823	50302	1	0	0	480	0
b0054	*lptD*	54755	57109	-1	0	0	2355	0
b0083	*ftsL*	91032	91397	1	0	0	366	0
b0084	*ftsI*	91413	93179	1	1	0.000628536	1767	1
b0085	*murE*	93166	94653	1	0	0	1488	0
b0086	*murF*	94650	96008	1	0	0	1359	0
b0087	*mraY*	96002	97084	1	0	0	1083	0
b0088	*murD*	97087	98403	1	0	0	1317	0
b0089	*ftsW*	98403	99647	1	0	0	1245	0
b0090	*murG*	99644	100711	1	0	0	1068	0
b0091	*murC*	100765	102240	1	0	0	1476	0
b0093	*ftsQ*	103155	103985	1	0	0	831	0
b0094	*ftsA*	103982	105244	1	0	0	1263	0
b0095	*ftsZ*	105305	106456	1	0	0	1152	0
b0096	*lpxC*	106557	107474	1	0	0	918	0
b0098	*secA*	108279	110984	1	0	0	2706	0
b0103	*coaE*	112599	113219	-1	4	0.003577818	621	2
b0116	*lpd*	127912	129336	1	4	0.003117693	1425	4
b0126	*can*	142008	142670	-1	0	0	663	0
b0154	*hemI*	173602	174882	-1	2	0.000867303	1281	1
b0156	*erpA*	176610	176954	1	0	0	345	0
b0159	*mtn*	178455	179153	-1	5	0.006349206	699	4
b0166	*dapD*	185123	185947	-1	1	0.001345895	825	1
b0168	*map*	188712	189506	-1	0	0	795	0
b4414	*tff*	189712	189847	1	0	0	136	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b0169	*rpsB*	189874	190599	1	0	0	726	0
b0170	*tsf*	190857	191708	1	0	0	852	0
b0171	*pyrH*	191855	192580	1	0	0	726	0
b0172	*frr*	192872	193429	1	0	0	558	0
b0173	*dxr*	193521	194717	1	2	0.000927644	1197	1
b0174	*ispU*	194903	195664	1	0	0	762	0
b0175	*cdsA*	195677	196534	1	0	0	858	0
b0176	*rseP*	196546	197898	1	0	0	1353	0
b0177	*bamA*	197928	200360	1	0	0	2433	0
b0179	*lpxD*	200971	201996	1	0	0	1026	0
b0180	*fabZ*	202101	202556	1	0	0	456	0
b0181	*lpxA*	202560	203348	1	0	0	789	0
b0182	*lpxB*	203348	204496	1	4	0.000966184	1149	1
b0184	*dnaE*	205126	208608	1	0	0	3483	0
b0185	*accA*	208621	209580	1	0	0	960	0
b0188	*tilS*	212331	213629	1	0	0	1299	0
b0194	*proS*	217057	218775	-1	0	0	1719	0
b0201	*rrsH*	223771	225312	1	5	0.001440922	1542	2
b0202	*ileV*	225381	225457	1	0	0	77	0
b0203	*alaV*	225500	225575	1	0	0	76	0
b0204	*rrlH*	225759	228662	1	6	0.001147666	2904	3
b0205	*rrfH*	228756	228875	1	0	0	120	0
b0206	*aspU*	228928	229004	1	0	0	77	0
b4586	*ykfM*	238257	238736	-1	2	0.002314815	480	1
b4688	*ykgS*	289653	289857	1	0	0	205	0
b0369	*hemB*	387977	388951	-1	2	0.001138952	975	1
b0408	*secD*	426871	428718	1	0	0	1848	0
b0409	*secF*	428729	429700	1	4	0.001142857	972	1
b0414	*ribD*	432679	433782	1	0	0	1104	0
b0415	*ribE*	433871	434341	1	0	0	471	0
b0417	*thiL*	434858	435835	1	3	0.002270148	978	2
b0420	*dxs*	437539	439401	-1	0	0	1863	0
b0421	*ispA*	439426	440325	-1	1	0.001234568	900	1

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b0455	*ffs*	475672	475785	1	0	0	114	0
b0470	*dnaX*	491316	493247	1	0	0	1932	0
b0474	*adk*	496399	497043	1	0	0	645	0
b0475	*hemH*	497279	498241	1	0	0	963	0
b0524	*lpxH*	552441	553163	-1	0	0	723	0
b0525	*ppiB*	553166	553660	-1	1	0.002242152	495	1
b0526	*cysS*	553834	555219	1	0	0	1386	0
b0529	*folD*	556098	556964	-1	0	0	867	0
b0536	*argU*	563946	564022	1	0	0	77	0
b0634	*mrdB*	664424	665536	-1	0	0	1113	0
b0635	*mrdA*	665539	667440	-1	0	0	1902	0
b0639	*nadD*	669154	669795	-1	0	0	642	0
b0640	*holA*	669797	670828	-1	0	0	1032	0
b0641	*lptE*	670828	671409	-1	0	0	582	0
b0642	*leuS*	671424	674006	-1	0	0	2583	0
b0657	*lnt*	688566	690104	-1	0	0	1539	0
b0659	*ybeY*	691097	691564	-1	3	0.002369668	468	1
b0662	*ubiF*	694324	695499	1	9	0.005665722	1176	6
b0666	*metU*	695887	695963	-1	0	0	77	0
b0668	*glnW*	695979	696053	-1	0	0	75	0
b0670	*glnU*	696088	696162	-1	0	0	75	0
b0672	*leuW*	696186	696270	-1	0	0	85	0
b0673	*metT*	696280	696356	-1	0	0	77	0
b0680	*glnS*	705316	706980	1	0	0	1665	0
b0684	*fldA*	710158	710688	-1	0	0	531	0
b0726	*sucA*	757929	760730	1	11	0.003568596	2802	9
b0727	*sucB*	760745	761962	1	2	0.000911577	1218	1
b0733	*cydA*	770681	772249	1	5	0.00353857	1569	5
b0734	*cydB*	772265	773404	1	2	0.001949318	1140	2
b0743	*lysT*	779777	779852	1	0	0	76	0
b0744	*valT*	779988	780063	1	0	0	76	0
b0745	*lysW*	780066	780141	1	0	0	76	0
b0746	*valZ*	780291	780366	1	0	0	76	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b0747	*lysY*	780370	780445	1	0	0	76	0
b0884	*infA*	925448	925666	-1	0	0	219	0
b0886	*cydC*	926697	928418	-1	6	0.001935484	1722	3
b0887	*cydD*	928419	930185	-1	4	0.002514142	1767	4
b0891	*lolA*	936595	937206	1	0	0	612	0
b0893	*serS*	938651	939943	1	0	0	1293	0
b0910	*cmk*	960424	961107	1	4	0.00487013	684	3
b0911	*rpsA*	961218	962891	1	1	0.00066357	1674	1
b0914	*msbA*	965844	967592	1	0	0	1749	0
b0915	*lpxK*	967589	968575	1	0	0	987	0
b0917	*ycaR*	969896	970078	1	0	0	183	0
b0918	*kdsB*	970075	970821	1	0	0	747	0
b0922	*mukF*	973542	974864	1	0	0	1323	0
b0923	*mukE*	974845	975549	1	0	0	705	0
b0924	*mukB*	975549	980009	1	0	0	4461	0
b0930	*asnS*	986808	988208	-1	0	0	1401	0
b0954	*fabA*	1015175	1015693	-1	0	0	519	0
b0971	*serT*	1030848	1030935	-1	0	0	88	0
b1054	*lpxL*	1114885	1115805	-1	0	0	921	0
b1069	*murJ*	1127062	1128597	1	0	0	1536	0
b1091	*fabH*	1147982	1148935	1	1	0.001164144	954	1
b1092	*fabD*	1148951	1149880	1	0	0	930	0
b1093	*fabG*	1149893	1150627	1	0	0	735	0
b1094	*acpP*	1150838	1151074	1	0	0	237	0
b1098	*tmk*	1154347	1154988	1	0	0	642	0
b1099	*holB*	1154985	1155989	1	0	0	1005	0
b1116	*lolC*	1174650	1175849	1	0	0	1200	0
b1117	*lolD*	1175842	1176543	1	0	0	702	0
b1118	*lolE*	1176543	1177787	1	0	0	1245	0
b1131	*purB*	1189839	1191209	-1	2	0.001620746	1371	2
b1145	*cohE*	1201482	1202156	-1	0	0	675	0
b1204	*pth*	1257152	1257736	-1	3	0.003795066	585	2
b1207	*prs*	1260151	1261098	-1	0	0	948	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b1208	*ispE*	1261249	1262100	-1	0	0	852	0
b1209	*lolB*	1262100	1262723	-1	0	0	624	0
b1210	*hemA*	1262937	1264193	1	1	0.000883392	1257	1
b1211	*prfA*	1264235	1265317	1	0	0	1083	0
b1212	*prmC*	1265317	1266150	1	0	0	834	0
b1215	*kdsA*	1267388	1268242	1	0	0	855	0
b4419	*ldrA*	1268391	1268498	-1	0	0	108	0
b4421	*ldrB*	1268926	1269033	-1	0	0	108	0
b4423	*ldrC*	1269461	1269568	-1	0	0	108	0
b1231	*tyrT*	1286761	1286845	-1	0	0	85	0
b1274	*topA*	1329072	1331669	1	11	0.002565199	2598	6
b1277	*ribA*	1336594	1337184	-1	0	0	591	0
b1279	*yciS*	1338267	1338575	1	1	0.003584229	309	1
b1280	*yciM*	1338582	1339751	1	0	0	1170	0
b4672	*ymiB*	1344820	1344924	1	0	0	105	0
b1288	*fabI*	1348275	1349063	-1	0	0	789	0
b4526	*ydaE*	1415862	1416032	-1	1	0.006493506	171	1
b1356	*racR*	1417789	1418265	-1	18	0.004651163	477	2
b1373	*tfaR*	1430435	1431010	1	8	0.005780347	576	3
b1375	*ynaE*	1432015	1432248	-1	0	0	234	0
b4674	*ynbG*	1463189	1463254	-1	0	0	66	0
b1455	*yncH*	1524964	1525176	1	1	0.005208333	213	1
b1457	*ydcD*	1527946	1528428	1	4	0.006896552	483	3
b1471	*yddK*	1542782	1543771	-1	11	0.006734007	990	6
b1500	*safA*	1581786	1581983	-1	0	0	198	0
b1508	*hipB*	1590200	1590466	-1	1	0.004149378	267	1
b1544	*ydfK*	1631096	1631329	1	0	0	234	0
b1546	*tfaQ*	1632334	1632909	-1	6	0.005780347	576	3
b1570	*dicA*	1645958	1646365	1	0	0	408	0
b1637	*tyrS*	1713972	1715246	-1	0	0	1275	0
b1652	*rnt*	1726371	1727018	1	5	0.005136986	648	3
b1665	*valV*	1744459	1744535	1	0	0	77	0
b1713	*pheT*	1793581	1795968	-1	0	0	2388	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b1714	*pheS*	1795983	1796966	-1	0	0	984	0
b1715	*pheM*	1797250	1797294	-1	0	0	45	0
b1716	*rplT*	1797417	1797773	-1	0	0	357	0
b1717	*rpmI*	1797826	1798023	-1	0	0	198	0
b1718	*infC*	1798120	1798662	-1	0	0	543	0
b1719	*thrS*	1798666	1800594	-1	0	0	1929	0
b1740	*nadE*	1820482	1821309	1	2	0.002680965	828	2
b1779	*gapA*	1860795	1861790	1	0	0	996	0
b1807	*yeaZ*	1888596	1889291	-1	4	0.003189793	696	2
b4433	*ryeB*	1921188	1921308	-1	0	0	121	0
b4677	*yobI*	1944139	1944204	-1	0	0	66	0
b1866	*aspS*	1946774	1948546	-1	0	0	1773	0
b1876	*argS*	1958086	1959819	1	0	0	1734	0
b1909	*leuZ*	1989839	1989925	-1	0	0	87	0
b1910	*cysT*	1989938	1990011	-1	0	0	74	0
b1911	*glyW*	1990066	1990141	-1	0	0	76	0
b1912	*pgsA*	1990293	1990841	-1	0	0	549	0
b4582	*yoeA*	2066659	2068498	1	51	0.006642512	1840	11
b2114	*metG*	2192322	2194355	1	25	0.004915347	2034	9
b2153	*folE*	2241006	2241674	-1	0	0	669	0
b2185	*rplY*	2280539	2280823	1	0	0	285	0
b2198	*ccmD*	2293399	2293608	-1	20	0.005291005	210	1
b2231	*gyrA*	2334815	2337442	-1	0	0	2628	0
b2232	*ubiG*	2337589	2338311	1	1	0.001536098	723	1
b2234	*nrdA*	2342887	2345172	1	0	0	2286	0
b2235	*nrdB*	2345406	2346536	1	0	0	1131	0
b2311	*ubiX*	2426079	2426648	-1	1	0.001949318	570	1
b2315	*folC*	2429696	2430964	-1	0	0	1269	0
b2316	*accD*	2431034	2431948	-1	0	0	915	0
b2323	*fabB*	2438407	2439627	-1	0	0	1221	0
b4643	*pawZ*	2474606	2474620	1	0	0	15	0
b2396	*alaX*	2516063	2516138	-1	0	0	76	0
b2397	*alaW*	2516178	2516253	-1	0	0	76	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b2400	*gltX*	2517279	2518694	-1	0	0	1416	0
b2401	*valU*	2518953	2519028	1	0	0	76	0
b2403	*valY*	2519195	2519270	1	0	0	76	0
b2404	*lysV*	2519275	2519350	1	0	0	76	0
b2411	*ligA*	2526183	2528198	-1	0	0	2016	0
b2412	*zipA*	2528269	2529255	-1	0	0	987	0
b2415	*ptsH*	2531786	2532043	1	0	0	258	0
b2472	*dapE*	2589629	2590756	1	0	0	1128	0
b2478	*dapA*	2596904	2597782	-1	0	0	879	0
b2496	*hda*	2616097	2616798	-1	2	0.003164557	702	2
b2511	*der*	2633906	2635378	-1	0	0	1473	0
b2514	*hisS*	2637323	2638597	-1	0	0	1275	0
b2515	*ispG*	2638708	2639826	-1	0	0	1119	0
b2525	*fdx*	2654770	2655105	-1	1	0.00330033	336	1
b2527	*hscB*	2656974	2657489	-1	15	0.006451613	516	3
b2529	*iscU*	2657925	2658311	-1	2	0.005730659	387	2
b2530	*iscS*	2658339	2659553	-1	6	0.004570384	1215	5
b2533	*suhB*	2661464	2662267	1	0	0	804	0
b2551	*glyA*	2682276	2683529	-1	7	0.00088574	1254	1
b2559	*tadA*	2695376	2695879	-1	1	0.002202643	504	1
b2563	*acpS*	2698640	2699020	-1	0	0	381	0
b2566	*era*	2700503	2701408	-1	0	0	906	0
b2568	*lepB*	2702357	2703331	-1	0	0	975	0
b2573	*rpoE*	2707459	2708034	-1	0	0	576	0
b2585	*pssA*	2720749	2722104	1	0	0	1356	0
b2588	*rrfG*	2724091	2724210	-1	0	0	120	0
b2589	*rrlG*	2724303	2727206	-1	0	0	2904	0
b2590	*gltW*	2727391	2727466	-1	0	0	76	0
b2591	*rrsG*	2727638	2729179	-1	6	0.001440922	1542	2
b2594	*rluD*	2733053	2734033	-1	1	0.001132503	981	1
b2595	*bamD*	2734168	2734905	1	0	0	738	0
b2606	*rplS*	2742205	2742552	-1	1	0.003184713	348	1
b2607	*trmD*	2742594	2743361	-1	0	0	768	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b2608	*rimM*	2743392	2743940	-1	1	0.002020202	549	1
b2609	*rpsP*	2743959	2744207	-1	0	0	249	0
b2610	*ffh*	2744456	2745817	-1	0	0	1362	0
b2614	*grpE*	2748137	2748730	-1	1	0.001869159	594	1
b2615	*nadK*	2748853	2749731	1	0	0	879	0
b2693	*argY*	2816220	2816296	-1	0	0	77	0
b2694	*argV*	2816495	2816571	-1	0	0	77	0
b2695	*serV*	2816575	2816667	-1	0	0	93	0
b2696	*csrA*	2816983	2817168	-1	0	0	186	0
b2697	*alaS*	2817403	2820033	-1	0	0	2631	0
b2746	*ispF*	2869323	2869802	-1	0	0	480	0
b2747	*ispD*	2869802	2870512	-1	0	0	711	0
b2779	*eno*	2904665	2905963	-1	0	0	1299	0
b2780	*pyrG*	2906051	2907688	-1	0	0	1638	0
b2827	*thyA*	2962383	2963177	-1	0	0	795	0
b2828	*lgt*	2963184	2964059	-1	0	0	876	0
b2851	*ygeG*	2989290	2989781	1	7	0.006772009	492	3
b2890	*lysS*	3031679	3033196	-1	3	0.002194587	1518	3
b2891	*prfB*	3033206	3034304	-1	0	0	1099	0
b2898	*ygfZ*	3039335	3040315	1	8	0.003397508	981	3
b2907	*ubiH*	3050362	3051540	-1	3	0.00094162	1179	1
b4665	*ibsC*	3054912	3054971	-1	0	0	60	0
b2925	*fbaA*	3068187	3069266	-1	0	0	1080	0
b2926	*pgk*	3069481	3070644	-1	0	0	1164	0
b2942	*metK*	3084728	3085882	1	0	0	1155	0
b2949	*yqgF*	3091522	3091938	1	4	0.002659574	417	1
b3018	*plsC*	3160766	3161503	-1	4	0.003007519	738	2
b3019	*parC*	3161737	3163995	-1	2	0.000491642	2259	1
b3030	*parE*	3171526	3173418	-1	0	0	1893	0
b3041	*ribB*	3181835	3182488	-1	0	0	654	0
b4666	*ibsE*	3193163	3193222	1	0	0	60	0
b3056	*cca*	3199913	3201151	1	2	0.000896057	1239	1
b3058	*folB*	3202243	3202611	-1	0	0	369	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b3064	*ygjD*	3207552	3208565	-1	0	0	1014	0
b3065	*rpsU*	3208803	3209018	1	0	0	216	0
b3066	*dnaG*	3209129	3210874	1	0	0	1746	0
b3067	*rpoD*	3211069	3212910	1	0	0	1842	0
b3069	*ileX*	3213620	3213695	1	0	0	76	0
b3123	*rnpB*	3268238	3268614	-1	0	0	377	0
b3164	*pnp*	3307055	3309190	-1	26	0.005200208	2136	10
b3165	*rpsO*	3309437	3309706	-1	0	0	270	0
b3168	*infB*	3311364	3314036	-1	7	0.001246883	2673	3
b3169	*nusA*	3314061	3315548	-1	8	0.002985075	1488	4
b3176	*glmM*	3320755	3322092	-1	0	0	1338	0
b3177	*folP*	3322085	3322933	-1	2	0.00130719	849	1
b3178	*ftsH*	3323023	3324957	-1	0	0	1935	0
b3179	*rlmE*	3325057	3325686	-1	3	0.005291005	630	3
b3183	*obgE*	3328604	3329776	-1	11	0.00094697	1173	1
b3185	*rpmA*	3330884	3331141	-1	0	0	258	0
b3186	*rplU*	3331162	3331473	-1	0	0	312	0
b3187	*ispB*	3331732	3332703	1	0	0	972	0
b3189	*murA*	3333257	3334516	-1	0	0	1260	0
b3199	*lptC*	3340858	3341433	1	8	0.005780347	576	3
b3200	*lptA*	3341402	3341959	1	0	0	558	0
b3201	*lptB*	3341966	3342691	1	0	0	726	0
b3230	*rpsI*	3375837	3376229	-1	1	0.002824859	393	1
b3231	*rplM*	3376245	3376673	-1	0	0	429	0
b3235	*degS*	3380222	3381289	1	7	0.005197505	1068	5
b3249	*mreD*	3396409	3396897	-1	0	0	489	0
b3250	*mreC*	3396897	3398000	-1	0	0	1104	0
b3251	*mreB*	3398066	3399109	-1	0	0	1044	0
b3255	*accB*	3403458	3403928	1	0	0	471	0
b3256	*accC*	3403939	3405288	1	0	0	1350	0
b3273	*thrV*	3421602	3421677	-1	0	0	76	0
b3274	*rrfD*	3421690	3421809	-1	0	0	120	0
b3275	*rrlD*	3421902	3424805	-1	0	0	2904	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b3276	*alaU*	3424980	3425055	-1	0	0	76	0
b3277	*ileU*	3425098	3425174	-1	0	0	77	0
b3278	*rrsD*	3425243	3426784	-1	0	0	1542	0
b3282	*rimN*	3428865	3429437	-1	0	0	573	0
b3287	*def*	3431712	3432221	1	0	0	510	0
b3288	*fmt*	3432236	3433183	1	0	0	948	0
b3294	*rplQ*	3437638	3438021	-1	0	0	384	0
b3295	*rpoA*	3438062	3439051	-1	0	0	990	0
b3296	*rpsD*	3439077	3439697	-1	0	0	621	0
b3297	*rpsK*	3439731	3440120	-1	0	0	390	0
b3298	*rpsM*	3440137	3440493	-1	0	0	357	0
b3299	*rpmJ*	3440640	3440756	-1	0	0	117	0
b3300	*secY*	3440788	3442119	-1	0	0	1332	0
b3301	*rplO*	3442127	3442561	-1	0	0	435	0
b3302	*rpmD*	3442565	3442744	-1	0	0	180	0
b3303	*rpsE*	3442748	3443251	-1	0	0	504	0
b3304	*rplR*	3443266	3443619	-1	0	0	354	0
b3305	*rplF*	3443629	3444162	-1	0	0	534	0
b3306	*rpsH*	3444175	3444567	-1	0	0	393	0
b3307	*rpsN*	3444601	3444906	-1	0	0	306	0
b3308	*rplE*	3444921	3445460	-1	0	0	540	0
b3309	*rplX*	3445475	3445789	-1	0	0	315	0
b3310	*rplN*	3445800	3446171	-1	0	0	372	0
b3311	*rpsQ*	3446336	3446590	-1	0	0	255	0
b3312	*rpmC*	3446590	3446781	-1	0	0	192	0
b3313	*rplP*	3446781	3447191	-1	0	0	411	0
b3314	*rpsC*	3447204	3447905	-1	0	0	702	0
b3315	*rplV*	3447923	3448255	-1	0	0	333	0
b3316	*rpsS*	3448270	3448548	-1	0	0	279	0
b3317	*rplB*	3448565	3449386	-1	0	0	822	0
b3318	*rplW*	3449404	3449706	-1	0	0	303	0
b3319	*rplD*	3449703	3450308	-1	0	0	606	0
b3320	*rplC*	3450319	3450948	-1	0	0	630	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b3321	*rpsJ*	3450981	3451292	-1	0	0	312	0
b3339	*tufA*	3468167	3469351	-1	3	0.001874414	1185	2
b3340	*fusA*	3469422	3471536	-1	0	0	2115	0
b3341	*rpsG*	3471564	3472103	-1	1	0.002057613	540	1
b3342	*rpsL*	3472200	3472574	-1	0	0	375	0
b3357	*crp*	3484142	3484774	1	4	0.001754386	633	1
b3384	*trpS*	3510656	3511660	-1	0	0	1005	0
b3398	*yrfF*	3524491	3526626	1	3	0.000520021	2136	1
b3433	*asd*	3571798	3572901	-1	0	0	1104	0
b3461	*rpoH*	3597952	3598806	-1	2	0.002597403	855	2
b3464	*ftsY*	3600773	3602266	-1	0	0	1494	0
b3559	*glyS*	3720351	3722420	-1	0	0	2070	0
b3560	*glyQ*	3722430	3723341	-1	0	0	912	0
b3593	*rhsA*	3760206	3764339	1	68	0.005912389	4134	22
b3608	*gpsA*	3780665	3781684	-1	1	0.001089325	1020	1
b3609	*secB*	3781684	3782151	-1	2	0.002369668	468	1
b3633	*waaA*	3806563	3807840	1	2	0.00086881	1278	1
b3634	*coaD*	3807848	3808327	1	0	0	480	0
b3637	*rpmB*	3809461	3809697	-1	0	0	237	0
b3639	*dfp*	3810754	3811974	1	0	0	1221	0
b3640	*dut*	3811955	3812410	1	0	0	456	0
b3648	*gmk*	3819451	3820074	1	0	0	624	0
b3699	*gyrB*	3875728	3878142	-1	0	0	2415	0
b3701	*dnaN*	3879244	3880344	-1	0	0	1101	0
b3702	*dnaA*	3880349	3881752	-1	0	0	1404	0
b3703	*rpmH*	3882359	3882499	1	0	0	141	0
b3704	*rnpA*	3882516	3882875	1	0	0	360	0
b3705	*yidC*	3883099	3884745	1	0	0	1647	0
b3730	*glmU*	3911853	3913223	-1	0	0	1371	0
b3756	*rrsC*	3939831	3941372	1	0	0	1542	0
b3757	*gltU*	3941458	3941533	1	0	0	76	0
b3758	*rrlC*	3941727	3944630	1	2	0.000382555	2904	1
b3759	*rrfC*	3944723	3944842	1	0	0	120	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b3760	*aspT*	3944895	3944971	1	0	0	77	0
b3761	*trpT*	3944980	3945055	1	0	0	76	0
b3783	*rho*	3964440	3965699	1	6	0.000881834	1260	1
b3793	*wzyE*	3976624	3977976	1	0	0	1353	0
b3796	*argX*	3980398	3980474	1	0	0	77	0
b3797	*hisR*	3980532	3980608	1	0	0	77	0
b3799	*proM*	3980758	3980834	1	0	0	77	0
b3804	*hemD*	3987111	3987851	-1	1	0.00149925	741	1
b3805	*hemC*	3987848	3988789	-1	0	0	942	0
b3835	*ubiB*	4018249	4019889	1	0	0	1641	0
b3843	*ubiD*	4023011	4024504	1	7	0.002973978	1494	4
b3850	*hemG*	4032631	4033176	1	2	0.004065041	546	2
b3851	*rrsA*	4033554	4035095	1	0	0	1542	0
b3852	*ileT*	4035164	4035240	1	0	0	77	0
b3853	*alaT*	4035283	4035358	1	0	0	76	0
b3854	*rrlA*	4035542	4038446	1	7	0.000764818	2905	2
b3863	*polA*	4044989	4047775	1	18	0.003188521	2787	8
b3865	*yihA*	4048156	4048788	-1	0	0	633	0
b3935	*priA*	4122635	4124833	-1	23	0.006565657	2199	13
b3936	*rpmE*	4125036	4125248	1	2	0.005208333	213	1
b3967	*murI*	4163451	4164308	1	4	0.001293661	858	1
b3968	*rrsB*	4164682	4166223	1	0	0	1542	0
b3969	*gltT*	4166395	4166470	1	0	0	76	0
b3970	*rrlB*	4166664	4169567	1	0	0	2904	0
b3971	*rrfB*	4169660	4169779	1	0	0	120	0
b3972	*murB*	4170080	4171108	1	0	0	1029	0
b3973	*birA*	4171105	4172070	1	54	0.003448276	966	3
b3974	*coaA*	4172099	4173049	-1	3	0.002336449	951	2
b3976	*thrU*	4173411	4173486	1	0	0	76	0
b3978	*glyT*	4173696	4173770	1	0	0	75	0
b3980	*tufB*	4173967	4175151	1	9	0.005623243	1185	6
b3981	*secE*	4175381	4175764	1	0	0	384	0
b3982	*nusG*	4175766	4176311	1	0	0	546	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b3983	*rplK*	4176470	4176898	1	0	0	429	0
b3984	*rplA*	4176902	4177606	1	0	0	705	0
b3985	*rplJ*	4178019	4178516	1	0	0	498	0
b3986	*rplL*	4178583	4178948	1	0	0	366	0
b3987	*rpoB*	4179268	4183296	1	0	0	4029	0
b3988	*rpoC*	4183373	4187596	1	1	0.000263019	4224	1
b3997	*hemE*	4195739	4196803	1	0	0	1065	0
b4007	*rrsE*	4206170	4207711	1	0	0	1542	0
b4008	*gltV*	4207797	4207872	1	0	0	76	0
b4009	*rrlE*	4208066	4210969	1	0	0	2904	0
b4010	*rrfE*	4211063	4211182	1	0	0	120	0
b4040	*ubiA*	4251039	4251911	1	2	0.001272265	873	1
b4041	*plsB*	4252066	4254489	-1	2	0.00091659	2424	2
b4043	*lexA*	4255138	4255746	1	0	0	609	0
b4052	*dnaB*	4262337	4263752	1	0	0	1416	0
b4059	*ssb*	4272148	4272684	1	0	0	537	0
b4134	*pheU*	4360574	4360649	-1	0	0	76	0
b4142	*groS*	4368711	4369004	1	0	0	294	0
b4143	*groL*	4369048	4370694	1	0	0	1647	0
b4160	*psd*	4387415	4388383	-1	2	0.002290951	969	2
b4162	*orn*	4389627	4390172	1	0	0	546	0
b4164	*glyX*	4390495	4390570	1	0	0	76	0
b4168	*yjeE*	4393608	4394069	1	0	0	462	0
b4200	*rpsF*	4423141	4423536	1	0	0	396	0
b4201	*priB*	4423543	4423857	1	0	0	315	0
b4202	*rpsR*	4423862	4424089	1	2	0.004854369	228	1
b4226	*ppa*	4447145	4447675	-1	0	0	531	0
b4258	*valS*	4479005	4481860	-1	0	0	2856	0
b4261	*lptF*	4484241	4485341	1	0	0	1101	0
b4262	*lptG*	4485341	4486423	1	0	0	1083	0
b4361	*dnaC*	4598261	4598998	-1	0	0	738	0
b4362	*dnaT*	4599001	4599540	-1	1	0.002057613	540	1
b4370	*leuQ*	4604338	4604424	-1	0	0	87	0

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b4372	*holD*	4605826	4606239	1	0	0	414	0

Table A2. Genes identified as ambiguous in MG1655 following growth on LB agar

locus_tag	gene_name	start	end	strand	read_count	ins_index	gene_length	ins_count
b4662	*sgrT*	77388	77519	1	1	0.008403361	132	1
b0082	*rsmH*	90094	91035	1	11	0.009433962	942	8
b0418	*pgpA*	435813	436331	1	12	0.008547009	519	4
b0439	*lon*	458112	460466	1	34	0.00990566	2355	21
b0542	*renD*	565907	567470	1	45	0.009232955	1564	13
b0631	*ybeD*	661602	661865	-1	2	0.008403361	264	2
b4514	*ybfQ*	735668	735907	1	10	0.009259259	240	2
b4515	*ybgT*	773419	773532	1	1	0.009708738	114	1
b0956	*matP*	1017708	1018160	1	10	0.009803922	453	4
b1138	*ymfE*	1196756	1197460	-1	7	0.007874016	705	5
b1174	*minE*	1223502	1223768	-1	4	0.008298755	267	2
b1662	*ribC*	1740625	1741266	-1	6	0.006920415	642	4
b4537	*yecJ*	1985531	1985782	-1	11	0.008810573	252	2
b2276	*nuoN*	2388070	2389527	-1	32	0.009139375	1458	12
b2285	*nuoE*	2399574	2400074	-1	9	0.00886918	501	4
b2416	*ptsI*	2532088	2533815	1	17	0.007069409	1728	11
b2507	*guaA*	2628980	2630557	-1	23	0.007037298	1578	10
b2526	*hscA*	2655107	2656957	-1	34	0.008403361	1851	14
b2567	*rnc*	2701405	2702085	-1	19	0.008156607	681	5
b2583	*yfiP*	2717245	2717943	1	46	0.00952381	699	6
b2748	*ftsB*	2870531	2870842	-1	6	0.007117438	312	2
b4682	*yqcG*	2903579	2903719	1	1	0.007874016	141	1
b3167	*rbfA*	3310799	3311200	-1	8	0.008287293	402	3
b3198	*kdsC*	3340295	3340861	1	15	0.009784736	567	5
b3343	*tusB*	3472700	3472987	-1	3	0.007692308	288	2
b3482	*rhsB*	3617215	3621450	1	159	0.009441385	4236	36
b3607	*cysE*	3779764	3780585	-1	8	0.009459459	822	7
b3833	*ubiE*	4016878	4017633	1	5	0.007342144	756	5
b3855	*rrfA*	4038540	4038659	1	1	0.009259259	120	1
b4655	*ythA*	4504471	4504596	1	4	0.00877193	126	1
b4294	*insA*	4516550	4516825	1	8	0.008032129	276	2

Table A3. Comparative analysis with Goodall *et al.*, 2018.
Essential genes shared by all studies

locus_tag	gene_name	Function
b0185	*accA*	acetyl-CoA carboxylase
b3255	*accB*	acetyl CoA carboxylase
b3256	*accC*	acetyl-CoA carboxylase
b2316	*accD*	acetyl-CoA carboxylase
b1094	*acpP*	acyl carrier protein (ACP)
b2563	*acpS*	holo-[acyl-carrier-protein] synthase 1
b0474	*adk*	adenylate kinase
b1876	*argS*	arginyl-tRNA synthetase
b3433	*asd*	aspartate-semialdehyde dehydrogenase
b0930	*asnS*	asparaginyl tRNA synthetase
b1866	*aspS*	aspartyl-tRNA synthetase
b0177	*bamA*	outer membrane protein assembly factor
b2595	*bamD*	lipoprotein required for OM biogenesis
b3973	*birA*	bifunctional biotin-[acetylCoA carboxylase] holoenzyme synthetase/ DNA-binding transcriptional repressor
b0126	*can*	carbonic anhydrase
b3056	*cca*	fused tRNA nucleotidyl transferase/2'3'-cyclic phosphodiesterase/2'nucleotidase and phosphatase
b0175	*cdsA*	CDP-diglyceride synthase
b3634	*coaD*	pantetheine-phosphate adenylyltransferase
b2696	*csrA*	pleiotropic regulatory protein for carbon source metabolism
b0526	*cysS*	cysteinyl-tRNA synthetase
b2478	*dapA*	dihydrodipicolinate synthase
b0031	*dapB*	dihydrodipicolinate reductase
b0166	*dapD*	tetrahydrodipicolinate succinylase
b2472	*dapE*	N-succinyl-diaminopimelate deacylase
b3287	*def*	peptide deformylase
b2511	*der*	GTPase; multicopy suppressor of ftsJ
b3639	*dfp*	fused 4'-phosphopantothenoylcysteine decarboxylase/phosphopantothenoylcysteine synthetase
b3702	*dnaA*	chromosomal replication initiator protein DnaA
b4052	*dnaB*	replicative DNA helicase
b4361	*dnaC*	DNA biosynthesis protein

locus_tag	gene_name	Function
b0184	*dnaE*	DNA polymerase III alpha subunit
b3701	*dnaN*	DNA polymerase III
b0470	*dnaX*	DNA polymerase III/DNA elongation factor III
b3640	*dut*	deoxyuridinetriphosphatase
b0173	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase
b0420	*dxs*	1-deoxyxylulose-5-phosphate synthase
b2779	*eno*	enolase
b2566	*era*	membrane-associated
b0156	*erpA*	iron-sulfur cluster insertion protein
b0954	*fabA*	beta-hydroxydecanoyl thioester dehydrase
b2323	*fabB*	3-oxoacyl-[acyl-carrier-protein] synthase I
b1092	*fabD*	malonyl-CoA-[acyl-carrier-protein] transacylase
b1093	*fabG*	3-oxoacyl-[acyl-carrier-protein] reductase
b1288	*fabI*	enoyl-[acyl-carrier-protein] reductase
b0180	*fabZ*	(3R)-hydroxymyristol acyl carrier protein dehydratase
b2925	*fbaA*	fructose-bisphosphate aldolase
b2610	*ffh*	Signal Recognition Particle (SRP) component with 4.5S RNA (ffs)
b0684	*fldA*	flavodoxin 1
b3288	*fmt*	10-formyltetrahydrofolate:L-methionyl-tRNA(fMet) N-formyltransferase
b0048	*folA*	dihydrofolate reductase
b2315	*folC*	bifunctional folylpolyglutamate synthase/ dihydrofolate synthase
b0529	*folD*	bifunctional 5
b2153	*folE*	GTP cyclohydrolase I
b0172	*frr*	ribosome recycling factor
b0094	*ftsA*	ATP-binding cell division protein involved in recruitment of FtsK to Z ring
b3178	*ftsH*	protease
b0084	*ftsI*	transpeptidase involved in septal peptidoglycan synthesis (penicillin-binding protein 3)
b0083	*ftsL*	membrane bound cell division protein at septum containing leucine zipper motif
b0093	*ftsQ*	Divisome assembly protein
b0089	*ftsW*	Lipid II flippase; integral membrane protein involved in stabilizing FstZ ring during cell division
b3464	*ftsY*	Signal Recognition Particle (SRP) receptor
b0095	*ftsZ*	GTP-binding tubulin-like cell division protein
b3340	*fusA*	protein chain elongation factor EF-G

locus_tag	gene_name	Function
b1779	*gapA*	glyceraldehyde-3-phosphate dehydrogenase A
b3730	*glmU*	fused N-acetyl glucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyl transferase
b0680	*glnS*	glutamyl-tRNA synthetase
b2400	*gltX*	glutamyl-tRNA synthetase
b3560	*glyQ*	glycine tRNA synthetase
b3648	*gmk*	guanylate kinase
b3608	*gpsA*	glycerol-3-phosphate dehydrogenase (NAD+)
b4142	*groS*	Cpn10 chaperonin GroES
b2614	*grpE*	heat shock protein
b2231	*gyrA*	DNA gyrase (type II topoisomerase)
b3699	*gyrB*	DNA gyrase
b1210	*hemA*	glutamyl tRNA reductase
b0369	*hemB*	5-aminolevulinate dehydratase (porphobilinogen synthase)
b3805	*hemC*	hydroxymethylbilane synthase
b3804	*hemD*	uroporphyrinogen III synthase
b3850	*hemG*	protoporphyrin oxidase
b0475	*hemH*	ferrochelatase
b0154	*hemL*	glutamate-1-semialdehyde aminotransferase (aminomutase)
b2514	*hisS*	histidyl tRNA synthetase
b0640	*holA*	DNA polymerase III
b1099	*holB*	DNA polymerase III
b0884	*infA*	translation initiation factor IF-1
b3168	*infB*	fused protein chain initiation factor 2
b1718	*infC*	protein chain initiation factor IF-3
b0421	*ispA*	geranyltranstransferase
b3187	*ispB*	octaprenyl diphosphate synthase
b2747	*ispD*	4-diphosphocytidyl-2C-methyl-D-erythritol synthase
b1208	*ispE*	4-diphosphocytidyl-2-C-methylerythritol kinase
b2746	*ispF*	2C-methyl-D-erythritol 2
b2515	*ispG*	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase
b0029	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
b0174	*ispU*	undecaprenyl pyrophosphate synthase
b1215	*kdsA*	3-deoxy-D-manno-octulosonate 8-phosphate synthase

locus_tag	gene_name	Function
b0918	*kdsB*	3-deoxy-manno-octulosonate cytidylyltransferase
b2568	*lepB*	leader peptidase (signal peptidase I)
b0642	*leuS*	leucyl-tRNA synthetase
b4043	*lexA*	DNA-binding transcriptional repressor of SOS regulon
b2828	*lgt*	phosphatidylglycerol-prolipoprotein diacylglyceryl transferase
b2411	*ligA*	DNA ligase
b0657	*lnt*	apolipoprotein N-acyltransferase
b0891	*lolA*	chaperone for lipoproteins
b1209	*lolB*	OM lipoprotein required for localization of lipoproteins
b1116	*lolC*	lipoprotein-releasing system transmembrane protein
b1117	*lolD*	outer membrane-specific lipoprotein transporter subunit
b1118	*lolE*	lipoprotein-releasing system transmembrane protein
b3200	*lptA*	periplasmic LPS-binding protein
b0054	*lptD*	LPS assembly OM complex LptDE
b0641	*lptE*	LPS assembly OM complex LptDE
b4261	*lptF*	lipopolysaccharide export ABC permease of the LptBFGC export complex
b4262	*lptG*	lipopolysaccharide export ABC permease of the LptBFGC export complex
b0181	*lpxA*	UDP-N-acetylglucosamine acetyltransferase
b0182	*lpxB*	tetraacyldisaccharide-1-P synthase
b0096	*lpxC*	UDP-3-O-acyl N-acetylglucosamine deacetylase
b0179	*lpxD*	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase
b0524	*lpxH*	UDP-2
b0915	*lpxK*	lipid A 4'kinase
b0027	*lspA*	prolipoprotein signal peptidase (signal peptidase II)
b0168	*map*	methionine aminopeptidase
b2114	*metG*	methionyl-tRNA synthetase
b2942	*metK*	S-adenosylmethionine synthetase
b0087	*mraY*	phospho-N-acetylmuramoyl-pentapeptide transferase
b0635	*mrdA*	transpeptidase involved in peptidoglycan synthesis (penicillin-binding protein 2)
b0634	*mrdB*	cell wall shape-determining protein
b3251	*mreB*	cell wall structural complex MreBCD
b3250	*mreC*	cell wall structural complex MreBCD transmembrane component MreC
b3249	*mreD*	cell wall structural complex MreBCD transmembrane component MreD

locus_tag	gene_name	Function
b0914	*msbA*	fused lipid transporter subunits of ABC superfamily: membrane component/ATP-binding component
b0924	*mukB*	chromosome condensin MukBEF
b0923	*mukE*	chromosome condensin MukBEF
b0922	*mukF*	chromosome condensin MukBEF
b3189	*murA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
b3972	*murB*	UDP-N-acetylenolpyruvoylglucosamine reductase
b0091	*murC*	UDP-N-acetylmuramate:L-alanine ligase
b0088	*murD*	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase
b0085	*murE*	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-diaminopimelate ligase
b0086	*murF*	UDP-N-acetylmuramoyl-tripeptide:D-alanyl-D-alanine ligase
b0090	*murG*	N-acetylglucosaminyl transferase
b3967	*murI*	glutamate racemase
b1069	*murJ*	probable peptidoglycan lipid II flippase required for murein synthesis
b0639	*nadD*	nicotinic acid mononucleotide adenylyltransferase
b1740	*nadE*	NAD synthetase
b2615	*nadK*	NAD kinase
b2234	*nrdA*	ribonucleoside-diphosphate reductase 1
b2235	*nrdB*	ribonucleoside-diphosphate reductase 1
b3169	*nusA*	transcription termination/antitermination L factor
b3982	*nusG*	transcription termination factor
b3183	*obgE*	GTPase involved in cell partioning and DNA repair
b4162	*orn*	oligoribonuclease
b3030	*parE*	DNA topoisomerase IV
b2926	*pgk*	phosphoglycerate kinase
b1912	*pgsA*	phosphatidylglycerophosphate synthetase
b1714	*pheS*	phenylalanine tRNA synthetase
b1713	*pheT*	phenylalanine tRNA synthetase
b4041	*plsB*	glycerol-3-phosphate O-acyltransferase
b3018	*plsC*	1-acyl-sn-glycerol-3-phosphate acyltransferase
b4226	*ppa*	inorganic pyrophosphatase
b1211	*prfA*	peptide chain release factor RF-1
b1212	*prmC*	N5-glutamine methyltransferase
b0194	*proS*	prolyl-tRNA synthetase

locus_tag	gene_name	Function
b1207	*prs*	phosphoribosylpyrophosphate synthase
b4160	*psd*	phosphatidylserine decarboxylase
b2585	*pssA*	phosphatidylserine synthase (CDP-diacylglycerol-serine O-phosphatidyltransferase)
b1204	*pth*	peptidyl-tRNA hydrolase
b2780	*pyrG*	CTP synthetase
b0171	*pyrH*	uridylate kinase
b1277	*ribA*	GTP cyclohydrolase II
b0414	*ribD*	fused diaminohydroxyphosphoribosylaminopyrimidine deaminase and 5-amino-6-(5-phosphoribosylamino) uracil reductase
b0415	*ribE*	riboflavin synthase beta chain
b0025	*ribF*	bifunctional riboflavin kinase/FAD synthetase
b3282	*rimN*	tRNA(ANN) t(6)A37 threonylcarbamoyladenosine modification protein
b3704	*rnpA*	protein C5 component of RNase P
b3317	*rplB*	50S ribosomal subunit protein L2
b3320	*rplC*	50S ribosomal subunit protein L3
b3319	*rplD*	50S ribosomal subunit protein L4
b3308	*rplE*	50S ribosomal subunit protein L5
b3305	*rplF*	50S ribosomal subunit protein L6
b3985	*rplJ*	50S ribosomal subunit protein L10
b3986	*rplL*	50S ribosomal subunit protein L7/L12
b3231	*rplM*	50S ribosomal subunit protein L13
b3310	*rplN*	50S ribosomal subunit protein L14
b3301	*rplO*	50S ribosomal subunit protein L15
b3313	*rplP*	50S ribosomal subunit protein L16
b3294	*rplQ*	50S ribosomal subunit protein L17
b3304	*rplR*	50S ribosomal subunit protein L18
b2606	*rplS*	50S ribosomal subunit protein L19
b1716	*rplT*	50S ribosomal subunit protein L20
b3186	*rplU*	50S ribosomal subunit protein L21
b3315	*rplV*	50S ribosomal subunit protein L22
b3318	*rplW*	50S ribosomal subunit protein L23
b3309	*rplX*	50S ribosomal subunit protein L24
b3185	*rpmA*	50S ribosomal subunit protein L27
b3637	*rpmB*	50S ribosomal subunit protein L28

locus_tag	gene_name	Function
b3312	*rpmC*	50S ribosomal subunit protein L29
b3302	*rpmD*	50S ribosomal subunit protein L30
b3703	*rpmH*	50S ribosomal subunit protein L34
b3295	*rpoA*	RNA polymerase
b3987	*rpoB*	RNA polymerase
b3988	*rpoC*	RNA polymerase
b3461	*rpoH*	RNA polymerase
b0911	*rpsA*	30S ribosomal subunit protein S1
b0169	*rpsB*	30S ribosomal subunit protein S2
b3314	*rpsC*	30S ribosomal subunit protein S3
b3296	*rpsD*	30S ribosomal subunit protein S4
b3303	*rpsE*	30S ribosomal subunit protein S5
b3341	*rpsG*	30S ribosomal subunit protein S7
b3306	*rpsH*	30S ribosomal subunit protein S8
b3230	*rpsI*	30S ribosomal subunit protein S9
b3321	*rpsJ*	30S ribosomal subunit protein S10
b3297	*rpsK*	30S ribosomal subunit protein S11
b3342	*rpsL*	30S ribosomal subunit protein S12
b3298	*rpsM*	30S ribosomal subunit protein S13
b3307	*rpsN*	30S ribosomal subunit protein S14
b2609	*rpsP*	30S ribosomal subunit protein S16
b3311	*rpsQ*	30S ribosomal subunit protein S17
b4202	*rpsR*	30S ribosomal subunit protein S18
b3316	*rpsS*	30S ribosomal subunit protein S19
b0098	*secA*	preprotein translocase subunit
b3981	*secE*	preprotein translocase membrane subunit
b3300	*secY*	preprotein translocase membrane subunit
b0893	*serS*	seryl-tRNA synthetase
b4059	*ssb*	single-stranded DNA-binding protein
b2533	*suhB*	inositol monophosphatase
b0417	*thiL*	thiamin-monophosphate kinase
b1719	*thrS*	threonyl-tRNA synthetase
b0188	*tilS*	tRNA(Ile)-lysidine synthetase

locus_tag	gene_name	Function
b1098	*tmk*	thymidylate kinase
b1274	*topA*	DNA topoisomerase I
b2607	*trmD*	tRNA m(1)G37 methyltransferase
b3384	*trpS*	tryptophanyl-tRNA synthetase
b0170	*tsf*	protein chain elongation factor EF-Ts
b1637	*tyrS*	tyrosyl-tRNA synthetase
b4040	*ubiA*	p-hydroxybenzoate octaprenyltransferase
b4258	*valS*	valyl-tRNA synthetase
b3633	*waaA*	3-deoxy-D-manno-octulosonic-acid transferase (KDO transferase)
b1807	*yeaZ*	protease specific for Gcp(YgjD)
b3064	*ygjD*	t(6)A tRNA modification protein; glycation-binding protein; genome maintenance protein
b3705	*yidC*	membrane protein insertase
b3865	*yihA*	GTP-binding protein required for normal cell division
b4168	*yjeE*	ADP-binding protein essential for nucleoid integrity
b2949	*yqgF*	predicted Holliday junction resolvase
b3398	*yrfF*	inner membrane protein
b2412	*zipA*	cell division protein involved in Z ring assembly

Table A4. Comparative analysis with Goodall *et al.*, 2018.
Essential genes shared by MG1655 and BW25113 TraDIS studies

locus_tag	gene_name	Category	Growth in Baba *et al.,* 2006	Potential polarity?
b0116	*lpd*	2-oxoglutarate dehydrogenase complex	Growth	N
b0726	*sucA*	2-oxoglutarate dehydrogenase complex	Growth	N
b0727	*sucB*	2-oxoglutarate dehydrogenase complex	Growth	N
b1054	*lpxL*	Cell envelope	Growth	N
b1091	*fabH*	Cell envelope	Growth	Y - *fabD*
b1279	*yciS (lapA)*	Cell envelope	Growth	Y - *yciM* (*lapB*)
b2551	*glyA*	Cofactor metabolic process	Growth	N
b3177	*folP*	Cofactor metabolic process	No growth	Y - *glmM*/*ftsH*
b3997	*hemE*	Cofactor metabolic process	No growth	N
b2525	*fdx*	Fe-S cluster assembly	Growth	N
b2529	*iscU*	Fe-S cluster assembly	Growth	N
b2530	*iscS*	Fe-S cluster assembly	Growth	N
b2898	*ygfZ*	Fe-S cluster assembly	Growth	N
b0734	*cydB*	Misc	Growth	N
b0887	*cydD*	Misc	Growth	N
b1508	*hipB*	Misc	Growth	N
b1652	*rnt*	Misc	Growth	N
b2594	*rluD*	Misc	Growth	N
b4372	*holD*	Misc	Growth	N
b2827	*thyA*	Misc	Growth	Y - *lgt*
b1715	*pheM*	Misc	Growth	Y - *pheS*
b1500	*safA*	Misc	N/A	N
b4201	*priB*	Misc	No growth	Y - *rpsR*, *rpsF*
b2608	*rimM*	Translation	Growth	Y - *trmD*

locus_tag	gene_name	Category	Growth in Baba *et al.,* 2006	Potential polarity?
b0659	*ybeY*	Translation	Growth	N
b2890	*lysS*	Translation	Growth	Y - *prfB*
b2185	*rplY*	Translation	Growth	N
b3984	*rplA*	Translation	Growth	Y - *rplK*
b1717	*rpmI*	Translation	Growth	Y - *rplT*, *infC*
b0023	*rpsT*	Translation	Indeterminate	N
b3165	*rpsO*	Translation	Indeterminate	N
b4200	*rpsF*	Translation	Indeterminate	Y - *priB*
b3065	*rpsU*	Translation	No growth	Y - *dnaG*
b3983	*rplK*	Translation	No growth	Y - *rplA*
b2232	*ubiG*	Ubiquinone	Growth	N
b2311	*ubiX*	Ubiquinone	Growth	N
b2907	*ubiH*	Ubiquinone	Growth	N
b1455	*yncH*	Y genes	Growth	N
b1457	*ydcD*	Y genes	Growth	N
b2851	*ygeG*	Y genes	Growth	N
b4526	*ydaE*	Y genes	Growth	N
b4586	*ykfM*	Y genes	N/A	N
b0917	*ycaR*	Y genes	Growth	Y - *kdsB*
b4672	*ymiB*	Y genes	N/A	N
b4674	*ynbG*	Y genes	N/A	N
b4677	*yobI*	Y genes	N/A	N

Table A5. Comparative analysis with Goodall *et al.*, 2018.
Essential genes unique to MG1655

locus_tag	gene_name	Function	Insertions in CDS?	Comment	Likely essential?
b2415	*ptsH*	phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)(*ptsH*)	No	No insertions	Yes
b3339	*tufA*	translation elongation factor EF-Tu 1(*tufA*)	Yes	Insertions mostly at extreme 3' end	Yes
b1280	*yciM*	LPS regulatory protein; putative modulator of LpxC proteolysis(*yciM*)	No	yciS essential in MG and BW tradis?	Yes
b1544	*ydfK*	cold shock protein, function unknown, Qin prophage(*ydfK*)	No	Surrounding genes have many insertions	Yes
b4688	*ykgS*	CP4-6 prophage; protein YkgS	Yes	Insertions mostly at extreme 3' end	Yes
b1375	*ynaE*	cold shock protein, Rac prophage(*ynaE*)	No	Surrounding genes have many insertions	Yes
b2198	*ccmD*	cytochrome c biogenesis protein; heme export ABC transporter holo-CcmE release factor(*ccmD*)	Yes	Insertions at extreme 5' end	Indeterminate
b4665	*ibsC*	toxic membrane protein(*ibsC*)	No	Gene length <75bp	Indeterminate
b4666	*ibsE*	toxic membrane protein(*ibsE*)	No	Gene length <75bp	Indeterminate
b4419	*ldrA*	toxic polypeptide, small(*ldrA*)	No	Surrounding genes have many insertions	Indeterminate
b4421	*ldrB*	toxic polypeptide, small(*ldrB*)	No	Surrounding genes have many insertions	Indeterminate
b4423	*ldrC*	toxic polypeptide, small(*ldrC*)	No	Surrounding genes have many insertions	Indeterminate
b3936	*rpmE*	50S ribosomal subunit protein L31(*rpmE*)	Yes	Growth rate?	Indeterminate
b3299	*rpmJ*	50S ribosomal subunit protein L36(*rpmJ*)	No	Growth rate?	Indeterminate
b3980	*tufB*	translation elongation factor EF-Tu 2(*tufB*)	Yes	Insertions mostly at extreme 5' end	Indeterminate
b0910	*cmk*	cytidylate kinase(*cmk*)	Yes	Contains insertions	No
b3357	*crp*	cAMP-activated global transcription factor, mediator of catabolite repression(*crp*)	Yes	Contains insertions	No
b2527	*hscB*	HscA co-chaperone, J domain-containing protein Hsc56; IscU-specific chaperone HscAB(*hscB*)	Yes	Contains insertions	No

locus_tag	gene_name	Function	Insertions in CDS?	Comment	Likely essential?
b0159	*mtn*	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase(*mtn*)	Yes	Insertions throughout gene	No
b3164	*pnp*	polynucleotide phosphorylase/polyadenylase(*pnp*)	Yes	Insertions throughout gene	No
b3863	*polA*	DNA polymerase I	Yes	Contains insertions	No
b0525	*ppiB*	peptidyl-prolyl cis-trans isomerase B (rotamase B)(*ppiB*)	No	Polar on downstream lpxH	No
b3593	*rhsA*	Rhs protein with putative toxin 55 domain; putative polysaccharide synthesis/export protein; putative neighboring cell growth inhibitor(*rhsA*)	Yes	Insertions mostly at extreme 3' end but some in middle of gene. Gene length 4133 bp	No
b3179	*rlmE*	23S rRNA U2552 2'-O-ribose methyltransferase, SAM-dependent(*rlmE*)	Yes	Insertions in one orientation, away from upstream essential ftsH	No
b4433	*ryeB*	small regulatory RNA SdsR	No	Only 104 bp long	No
b3609	*secB*	protein export chaperone(*secB*)	No	Downstream grxC (essential LB) has many insertions; upstream gpsA has few.	No
b1546	*tfaQ*	Qin prophage; putative tail fibre assembly protein(*tfaQ*)	Yes	Surrounding genes have few insertions.	No
b1373	*tfaR*	Rac prophage; putative tail fiber assembly protein(*tfaR*)	Yes	Surrounding genes have few insertions.	No
b4414	*tff*	ncRNA(*tff*)	No	Upstream of essential 30S subunit gene rpsB.	No
b0662	*ubiF*	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol oxygenase(*ubiF*)	Yes	Contains insertions	No
b1471	*yddK*	pseudo(*yddK*)	Yes	Contains insertions	No
b1995	*yoeA*	pseudo(*yoeA*)	Yes	Contains insertions	No
b3853	*alaT*	tRNA(*alaT*)	No	Gene length <75bp	No
b3276	*alaU*	tRNA(*alaU*)	No	Gene length <75bp	No
b0203	*alaV*	tRNA(*alaV*)	No	Gene length <75bp	No
b2397	*alaW*	tRNA(*alaW*)	No	Gene length <75bp	No
b2396	*alaX*	tRNA(*alaX*)	No	Gene length <75bp	No
b2694	*argV*	tRNA(*argV*)	No	Gene length <75bp	No
b2693	*argY*	tRNA(*argY*)	No	Gene length <75bp	No

locus_tag	gene_name	Function	Insertions in CDS?	Comment	Likely essential?
b3760	*aspT*	tRNA(*aspT*)	No	Gene length <75bp	No
b0206	*aspU*	tRNA(*aspU*)	No	Gene length <75bp	No
b0670	*glnU*	tRNA(*glnU*)	No	Gene length <75bp	No
b0668	*glnW*	tRNA(*glnW*)	No	Gene length <75bp	No
b3969	*gltT*	tRNA(*gltT*)	No	Gene length <75bp	No
b3757	*gltU*	tRNA(*gltU*)	No	Gene length <75bp	No
b4008	*gltV*	tRNA(*gltV*)	No	Gene length <75bp	No
b2590	*gltW*	tRNA(*gltW*)	No	Gene length <75bp	No
b1911	*glyW*	tRNA(*glyW*)	No	Gene length <75bp	No
b4164	*glyX*	tRNA(*glyX*)	No	Gene length <75bp	No
b3852	*ileT*	tRNA(*ileT*)	No	Gene length <75bp	No
b3277	*ileU*	tRNA(*ileU*)	No	Gene length <75bp	No
b0202	*ileV*	tRNA(*ileV*)	No	Gene length <75bp	No
b3069	*ileX*	tRNA(*ileX*)	No	Surrounding genes have many insertions	No
b4370	*leuQ*	tRNA(*leuQ*)	No	Gene length <75bp	No
b0743	*lysT*	tRNA(*lysT*)	No	Next to other lysine tRNAs who have no insertions.	No
b2404	*lysV*	tRNA(*lysV*)	No	Next to valine tRNAs with no insertions.	No
b0745	*lysW*	tRNA(*lysW*)	No	Next to other lysine tRNAs who have no insertions.	No
b0747	*lysY*	tRNA(*lysY*)	No	Next to other lysine tRNAs who have no insertions.	No
b0673	*metT*	tRNA(*metT*)	No	Next to other glutamine tRNAs who have no insertions.	No
b0666	*metU*	tRNA(*metU*)	No	Next to other lysine tRNAs who have no insertions.	No
b4643	*pawZ*	tRNA(*pawZ*)	No	Gene length <75bp. Pseudogene.	No
b4134	*pheU*	tRNA(*pheU*)	No	Surrounding genes have many insertions	No
b3971	*rrfB*	5S ribosomal RNA of *rrnB* operon(*rrfB*)	No	Surrounding genes have many insertions. 5S rRNA.	No

locus_tag	gene_name	Function	Insertions in CDS?	Comment	Likely essential?
b3759	*rrfC*	5S ribosomal RNA of *rrnC* operon(*rrfC*)	No	Surrounding genes have many insertions. 5S rRNA.	No
b3274	*rrfD*	5S ribosomal RNA of *rrnD* operon(*rrfD*)	No	Surrounding genes have many insertions. 5S rRNA.	No
b4010	*rrfE*	5S ribosomal RNA of *rrnE* operon(*rrfE*)	No	Surrounding genes have many insertions. 5S rRNA.	No
b3788	*rrfG*	5S ribosomal RNA of *rrnG* operon(*rrfG*)	No	Surrounding genes have many insertions. 5S rRNA.	No
b0205	*rrfH*	5S ribosomal RNA of *rrnH* operon(*rrfH*)	No	Surrounding genes have many insertions. 5S rRNA.	No
b3854	*rrlA*	23S ribosomal RNA of *rrnA* operon(*rrlA*)	Yes	Surrounding genes have many insertions. Investigate region with insertions.	No
b3970	*rrlB*	23S ribosomal RNA of *rrnB* operon(*rrlB*)	No	Surrounding genes have many insertions	No
b3758	*rrlC*	23S ribosomal RNA of *rrnC* operon(*rrlC*)	Yes	Surrounding genes have many insertions. Investigate region with insertions.	No
b3275	*rrlD*	23S ribosomal RNA of *rrnD* operon(*rrlD*)	No	Surrounding genes have many insertions	No
b4009	*rrlE*	23S ribosomal RNA of *rrnE* operon(*rrlE*)	No	Surrounding genes have many insertions	No
b2589	*rrlG*	23S ribosomal RNA of *rrnG* operon(*rrlG*)	No	Surrounding genes have many insertions	No
b0204	*rrlH*	23S ribosomal RNA of *rrnH* operon(*rrlH*)	Yes	Surrounding genes have many insertions.	No
b3851	*rrsA*	16S ribosomal RNA of *rrnA* operon(*rrsA*)	No	Surrounding genes have many insertions	No
b3968	*rrsB*	16S ribosomal RNA of *rrnB* operon(*rrsB*)	No	Surrounding genes have many insertions	No
b3756	*rrsC*	16S ribosomal RNA of *rrnC* operon(*rrsC*)	No	Surrounding genes have many insertions	No

locus_tag	gene_name	Function	Insertions in CDS?	Comment	Likely essential?
b3278	*rrsD*	16S ribosomal RNA of *rrnD* operon(*rrsD*)	No	Surrounding genes have many insertions	No
b4007	*rrsE*	16S ribosomal RNA of *rrnE* operon(*rrsE*)	No	Surrounding genes have many insertions	No
b2591	*rrsG*	16S ribosomal RNA of *rrnG* operon(*rrsG*)	No	Surrounding genes have many insertions	No
b0201	*rrsH*	16S ribosomal RNA of *rrnH* operon(*rrsH*)	Yes	Surrounding genes have many insertions	No
b3273	*thrV*	tRNA(*thrV*)	No	Gene length <75bp	No
b1231	*tyrT*	tRNA(*tyrT*)	No	Gene length <75bp	No
b0744	*valT*	tRNA(*valT*)	No	Gene length <75bp	No
b2401	*valU*	tRNA(*valU*)	No	Gene length <75bp	No
b1665	*valV*	tRNA(*valV*)	No	Gene length <75bp	No
b2403	*valY*	tRNA(*valY*)	No	Gene length <75bp	No
b0746	*valZ*	tRNA(*valZ*)	No	Gene length <75bp	No

Table A6. Comparative analysis with Goodall *et al.*, 2018.
Essential genes unique to BW25113

locus_tag	gene_name	Comment
b0115	*aceF*	
b4515	*cydX*	
b3809	*dapF*	
b2065	*dcd*	
b2507	*guaA*	Ambiguous in MG1655
b3082	*higA*	
b2526	*hscA*	Ambiguous in MG1655
b1712	*ihfA*	
b1160	*iraM*	
b0628	*lipA*	
b1638	*pdxH*	
b2416	*ptsI*	Ambiguous in MG1655
b3167	*rbfA*	Ambiguous in MG1655
b1564	*relB*	
b3386	*rpe*	
b1089	*rpmF*	
b2935	*tktA*	
b1252	*tonB*	
b1265	*trpL*	
b4638	*ttcc*	
b0969	*tusE*	
b3833	*ubiE*	Ambiguous in MG1655
b1357	*ydaS*	
b1472	*yddL*	
b1549	*ydfO*	
b1648	*ydhL*	
b1935	*yedM*	
b2450	*yffS*	
b2850	*ygeF*	
b2858	*ygeN*	
b4621	*yjbS*	
b1138	*ymfE*	Ambiguous in MG1655

locus_tag	gene_name	Comment
b4682	*yqcG*	Ambiguous in MG1655
b4683	*yqeL*	

Table A7. Genes with significant logFC values following growth in bile

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b0463	*acrA*	multidrug efflux system	-13.45762217	8.107981948	2.38E-147	1.99E-144
b1129	*phoQ*	sensory histidine kinase in two-component regulatory system with PhoP	-11.49034613	6.154695365	1.76E-28	2.23E-26
b3035	*tolC*	transport channel	-11.04748427	5.716194272	1.60E-18	1.49E-16
b1130	*phoP*	response regulator in two-component regulatory system with PhoQ	-10.99210588	5.664022857	1.50E-17	1.31E-15
b3026	*qseC*	quorum sensing sensory histidine kinase in two-component regulatory system with QseB	-10.6742087	5.351889666	5.36E-15	3.95E-13
b3625	*waaY*	lipopolysaccharide core biosynthesis protein	-10.57156247	7.565681444	2.45E-101	1.29E-98
b3632	*waaQ*	lipopolysaccharide core biosynthesis protein	-10.53616871	9.294558163	2.65E-236	1.11E-232
b2283	*nuoG*	NADH:ubiquinone oxidoreductase, chain G	-10.51441405	5.193699119	2.71E-14	1.93E-12
b0462	*acrB*	multidrug efflux system protein	-10.45658313	8.828509649	1.84E-206	3.87E-203
b2284	*nuoF*	NADH:ubiquinone oxidoreductase, chain F	-9.991423561	4.690209138	2.24E-09	1.13E-07
b2286	*nuoC*	NADH:ubiquinone oxidoreductase, fused CD subunit	-9.710277668	4.416953636	2.95E-10	1.65E-08
b2287	*nuoB*	NADH:ubiquinone oxidoreductase, chain B	-9.408109621	4.129102475	7.82E-09	3.72E-07
b1677	*lpp*	murein lipoprotein	-9.360655383	4.08062559	1.41E-09	7.42E-08
b2277	*nuoM*	NADH:ubiquinone oxidoreductase, membrane subunit M	-9.358471301	4.084179683	1.27E-07	5.06E-06
b2280	*nuoJ*	NADH:ubiquinone oxidoreductase, membrane subunit J	-9.127948557	3.856570574	1.76E-08	8.12E-07
b2278	*nuoL*	NADH:ubiquinone oxidoreductase, membrane subunit L	-9.099262266	3.835265076	2.23E-08	1.01E-06

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2282	*nuoH*	NADH:ubiquinone oxidoreductase, membrane subunit H	-9.077033141	3.808122554	3.15E-08	1.36E-06
b2276	*nuoN*	NADH:ubiquinone oxidoreductase, membrane subunit N	-9.064111238	3.7966679	2.36E-08	1.06E-06
b3790	*wecD*	TDP-fucosamine acetyltransferase	-8.995135148	3.7328504	2.55E-08	1.13E-06
b0632	*dacA*	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 5)	-8.778923104	6.631207866	1.06E-49	3.18E-47
b2742	*nlpD*	activator of AmiC murein hydrolase activity, lipoprotein	-8.597085736	8.630673946	2.86E-174	3.00E-171
b1279	*lapA*	lipopolysaccharide assembly protein A	-8.385769457	3.178239246	1.04E-05	0.000297558
b3836	*tatA*	TatABCE protein translocation system subunit	-8.359027723	3.143940262	9.28E-07	3.32E-05
b3631	*waaG*	UDP-glucose:(heptosyl)lipopolysaccharide alpha-1,3-glucosyltransferase; lipopolysaccharide core biosynthesis protein; lipopolysaccharide glucosyltransferase I	-8.33236493	6.716792399	2.44E-52	7.86E-50
b3839	*tatC*	TatABCE protein translocation system subunit	-8.317076347	3.115936475	1.48E-05	0.000410424
b2281	*nuoI*	NADH:ubiquinone oxidoreductase, chain I	-8.31115119	3.105603277	2.07E-06	6.97E-05
b3627	*waaO*	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1,3-D-galactosyltransferase	-8.058461143	7.132861909	7.29E-68	3.06E-65
b1782	*mipA*	scaffolding protein for murein synthesizing machinery	-7.875083818	6.649558327	2.32E-47	6.08E-45
b3791	*wecE*	TDP-4-oxo-6-deoxy-D-glucose transaminase	-7.35178421	5.228702309	9.48E-13	6.31E-11
b3229	*sspA*	stringent starvation protein A, phage P1 late gene activator, RNAP-associated acid-resistance protein, inactive glutathione S-transferase homolog	-7.338341577	7.070122812	4.18E-63	1.60E-60
b3613	*envC*	activator of AmiB,C murein hydrolases, septal ring factor	-6.864721514	4.75839634	5.86E-10	3.19E-08
b0214	*rnhA*	ribonuclease HI, degrades RNA of DNA-RNA hybrids	-6.466952214	3.568802069	8.39E-07	3.03E-05

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b3784	*wecA*	UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase	-6.345422176	6.971113853	1.78E-41	3.56E-39
b3388	*damX*	cell division protein that binds to the septal ring	-6.295504459	6.047208622	2.25E-19	2.14E-17
b3387	*dam*	DNA adenine methyltransferase	-6.191379902	3.324487575	1.80E-05	0.000477657
b2741	*rpoS*	RNA polymerase, sigma S (sigma 38) factor	-6.148878632	10.2161087	1.11E-141	7.75E-139
b3786	*wecB*	UDP-N-acetyl glucosamine-2-epimerase	-6.065042639	6.498622476	5.37E-30	7.26E-28
b0739	*tolA*	membrane anchored protein in TolA-TolQ-TolR complex	-5.930362556	3.072178617	9.36E-05	0.002121891
b3628	*waaB*	lipopolysaccharide 1,6-galactosyltransferase; UDP-D-galactose:(glucosyl)lipopolysaccharide-1, 6-D-galactosyltransferase	-5.552127939	7.58817588	3.37E-58	1.18E-55
b3842	*rfaH*	transcription antitermination protein	-5.489032561	3.46552811	5.20E-06	0.000162573
b2346	*mlaA*	ABC transporter maintaining OM lipid asymmetry, OM lipoprotein component	-5.249988062	7.297589513	3.30E-34	5.54E-32
b3194	*mlaE*	ABC transporter maintaining OM lipid asymmetry, inner membrane permease protein	-5.17713926	6.375473821	3.82E-26	4.33E-24
b3195	*mlaF*	ABC transporter maintaining OM lipid asymmetry, ATP-binding protein	-5.142729773	8.162446773	6.25E-34	1.01E-31
b2904	*gcvH*	glycine cleavage complex lipoylprotein	-4.868416243	3.758523678	9.41E-06	0.000274105
b1856	*mepM*	murein DD-endopeptidase, space-maker hydrolase, septation protein	-4.86038371	6.728665286	3.67E-32	5.50E-30
b4313	*fimE*	tyrosine recombinase/inversion of on/off regulator of fimA	-4.74222024	13.59601667	9.56E-176	1.34E-172
b3912	*cpxR*	response regulator in two-component regulatory system with CpxA	-4.732320293	4.853807006	1.01E-07	4.07E-06
b1830	*prc*	carboxy-terminal protease for penicillin-binding protein 3	-4.721561537	6.380489668	7.08E-18	6.32E-16

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2435	*amiA*	N-acetylmuramoyl-l-alanine amidase I	-4.666784349	6.584599099	7.20E-30	9.43E-28
b3649	*rpoZ*	RNA polymerase, omega subunit	-4.525412657	6.094936654	1.31E-16	1.06E-14
b1048	*opgG*	OPG biosynthetic periplasmic beta-1,6 branching glycosyltransferase	-4.428565763	6.826877857	2.66E-26	3.10E-24
b2188	*yejM*	essential inner membrane DUF3413 domain-containing protein; lipid A production and membrane permeability factor	-4.379544394	7.013388293	4.34E-28	5.35E-26
b3191	*mlaB*	ABC transporter maintaining OM lipid asymmetry, cytoplasmic STAS component	-4.371585805	6.957828281	2.91E-38	5.54E-36
b3150	*yraP*	outer membrane lipoprotein	-4.292027217	5.092370556	1.22E-06	4.28E-05
b2288	*nuoA*	NADH:ubiquinone oxidoreductase, membrane subunit A	-4.275454887	3.958877556	0.00013782	0.002994899
b1193	*emtA*	lytic murein endotransglycosylase E	-4.172616843	5.067818758	0.000411007	0.007695375
b3787	*wecC*	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	-4.07503752	6.544759978	2.62E-13	1.77E-11
b3192	*mlaC*	ABC transporter maintaining OM lipid asymmetry, periplasmic binding protein	-4.063917144	5.896084057	5.36E-08	2.25E-06
b2817	*amiC*	N-acetylmuramoyl-L-alanine amidase	-3.972138737	7.573107333	7.72E-48	2.16E-45
b2174	*lpxT*	lipid A 1-diphosphate synthase; undecaprenyl pyrophosphate:lipid A 1-phosphate phosphotransferase	-3.659447131	7.970115437	2.97E-47	7.34E-45
b3785	*wzzE*	Entobacterial Common Antigen (ECA) polysaccharide chain length modulation protein	-3.513526463	7.082187762	4.35E-15	3.32E-13
b1049	*opgH*	OPG biosynthetic ACP-dependent transmembrane UDP-glucose beta-1,2 glycosyltransferase; nutrient-dependent cell size regulator, FtsZ assembly antagonist	-3.467307771	6.508859021	1.88E-19	1.83E-17
b3207	*yrbL*	Mg(2+)-starvation-stimulated protein	-3.46436912	6.972337621	1.10E-25	1.21E-23
b0957	*ompA*	outer membrane protein A (3a;II*;G;d)	-3.370508115	5.847651978	3.98E-07	1.53E-05

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b1604	*ydgH*	DUF1471 family periplasmic protein	-3.270032231	7.533634092	1.11E-37	2.03E-35
b2531	*iscR*	isc operon transcriptional repressor; suf operon transcriptional activator; oxidative stress-and iron starvation-inducible; autorepressor	-3.204867193	7.79130598	2.99E-44	6.61E-42
b2317	*dedA*	DedA family inner membrane protein	-3.159649916	6.877252335	1.24E-21	1.24E-19
b4232	*fbp*	fructose-1,6-bisphosphatase I	-3.126913274	7.639618964	9.88E-36	1.73E-33
b2313	*cvpA*	colicin V production protein	-3.088368482	5.963595406	2.87E-08	1.25E-06
b0877	*ybjX*	DUF535 family protein	-3.040792035	7.416931525	6.47E-28	7.76E-26
b3630	*waaP*	kinase that phosphorylates core heptose of lipopolysaccharide	-2.962906875	5.136343118	9.01E-05	0.002065253
b1097	*yceG (mltG)*	septation protein, ampicillin sensitivity	-2.934249077	6.30641219	1.63E-10	9.47E-09
b3193	*mlaD*	OM lipid asymmetry maintenance protein; membrane-anchored ABC family periplasmic binding protein	-2.824128901	6.128215486	1.72E-05	0.00046237
b3753	*rbsR*	transcriptional repressor of ribose metabolism	-2.741790731	9.331324783	1.94E-45	4.51E-43
b3163	*nlpI*	lipoprotein involved in osmotic sensitivity and filamentation	-2.692187914	5.787844191	2.85E-05	0.000736581
b2865	*ygeR*	LysM domain-containing M23 family putative peptidase; septation lipoprotein	-2.677389398	7.196411928	4.19E-17	3.52E-15
b0436	*tig*	peptidyl-prolyl cis/trans isomerase (trigger factor)	-2.547921418	7.093621214	3.43E-15	2.66E-13
b0592	*fepB*	ferrienterobactin ABC transporter periplasmic binding protein	-2.371796342	5.934981076	7.34E-06	0.000221538
b0209	*yafD*	endo/exonuclease/phosphatase family protein	-2.334268969	5.685463045	0.000178755	0.003786366
b2299	*yfcD*	putative NUDIX hydrolase	-2.249695085	5.971067154	5.36E-06	0.000165417

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2215	*ompC*	outer membrane porin protein C	-2.113010415	6.246736468	0.0002236	0.0046889
b0889	*lrp*	leucine-responsive global transcriptional regulator	-2.111372737	6.634139281	2.06E-10	1.17E-08
b0759	*galE*	UDP-galactose-4-epimerase	-2.072225453	5.904218461	0.000225232	0.004699616
b2144	*sanA*	DUF218 superfamily vancomycin high temperature exclusion protein	-2.042948636	9.448720281	3.02E-24	3.16E-22
b2784	*relA*	(p)ppGpp synthetase I/GTP pyrophosphokinase	-1.913792193	7.394577113	2.64E-12	1.68E-10
b0920	*elyC*	envelope biogenesis factor; DUF218 superfamily protein	-1.839529726	6.577463941	5.86E-07	2.17E-05
b2813	*mltA*	membrane-bound lytic murein transglycosylase A	-1.791409174	7.21764081	1.78E-06	6.12E-05
b0585	*fes*	enterobactin/ferrienterobactin esterase	-1.783776862	6.491962774	2.00E-05	0.000525291
b1176	*minC*	inhibitor of FtsZ ring polymerization	-1.769927978	6.303814662	1.98E-05	0.000522218
b2593	*yfiH (pgeF)*	UPF0124 family protein	-1.764334883	6.29011286	1.51E-05	0.000416091
b2366	*dsdA*	D-serine dehydratase	-1.762039446	7.588581876	2.86E-12	1.79E-10
b2903	*gcvP*	glycine decarboxylase, PLP-dependent, subunit P of glycine cleavage complex	-1.727409141	7.680437843	1.40E-09	7.42E-08
b0434	*yajG*	putative lipoprotein	-1.695713511	6.862551325	9.46E-09	4.46E-07
b0633	*rlpA*	septal ring protein, suppressor of prc, minor lipoprotein	-1.687315928	6.502535855	1.23E-06	4.29E-05
b3493	*pitA*	phosphate transporter, low-affinity; tellurite importer	-1.620857596	6.14873557	0.000231361	0.004756512
b1961	*dcm*	DNA cytosine methyltransferase	-1.608714021	7.760474999	2.26E-07	8.92E-06

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2905	*gcvT*	aminomethyltransferase, tetrahydrofolate-dependent, subunit (T protein) of glycine cleavage complex	-1.573287896	6.900182401	1.18E-06	4.20E-05
b3017	*ftsP*	septal ring component that protects the divisome from stress; multicopy suppressor of ftsI(Ts)	-1.530552824	7.709805247	5.38E-11	3.23E-09
b3147	*lpoA*	OM lipoprotein stimulator of MrcA transpeptidase	-1.519435943	6.357341874	4.87E-05	0.001188517
b2494	*bepA*	OM protein maintenance and assembly metalloprotease and chaperone, periplasmic	-1.505296613	7.134477922	4.23E-08	1.79E-06
b2767	*ygcO*	putative 4Fe-4S cluster-containing protein	-1.468248473	6.898021691	5.04E-06	0.000159049
b0433	*ampG*	muropeptide transporter	-1.3638939	8.516976105	1.48E-13	1.03E-11
b1954	*dsrA*		-1.348764807	6.672725134	3.53E-05	0.000880266
b1831	*proQ*	RNA chaperone, putative ProP translation regulator	-1.29201986	7.590029178	3.51E-06	0.000111406
b3940	*metL*	Bifunctional aspartokinase/homoserine dehydrogenase 2	-1.268421934	7.934006134	1.23E-08	5.75E-07
b3624	*waaZ*	lipopolysaccharide KdoIII transferase; lipopolysaccharide core biosynthesis protein	-1.242886553	7.85455316	3.76E-09	1.84E-07
b0949	*uup*	replication regulatory ABC-F family DNA-binding ATPase	-1.167099607	7.272010934	0.000301079	0.005819002
b3626	*waaR*	lipopolysaccharide 1,2-glucosyltransferase; UDP-glucose:(glucosyl)LPS alpha-1,2-glucosyltransferase	-1.164278643	9.035383338	1.97E-12	1.27E-10
b1878	*flhE*	proton seal during flagellar secretion	-1.104199313	6.960348587	8.05E-05	0.001853937
b2958	*yggN*	DUF2884 family putative periplasmic protein	-1.062743664	7.534634164	2.43E-06	8.08E-05
b0240	*crl*	pseudogene	-1.059985162	8.25727622	1.95E-09	9.96E-08
b2624	*alpA*	CP4-57 prophage; DNA-binding transcriptional activator	-1.023915082	6.935118865	0.000236948	0.004814513
b1504	*ydeS*	putative fimbrial-like adhesin protein	-1.02006805	7.402301035	0.000100203	0.002247342
b3623	*waaU*	lipopolysaccharide core biosynthesis	-1.01767509	8.008665355	6.55E-07	2.41E-05
b3755	*yieP*	putative transcriptional regulator	-0.98741311	8.420198838	9.88E-08	4.02E-06

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b3590	*selB*	selenocysteinyl-tRNA-specific translation factor	-0.981505802	7.512376007	7.06E-05	0.001654914
b1811	*yoaH*	UPF0181 family protein	-0.980819156	7.028728975	0.000283544	0.005559586
b2837	*galR*	galactose-inducible d-galactose regulon transcriptional repressor; autorepressor	-0.978401643	7.697924396	1.69E-05	0.000459134
b3813	*uvrD*	DNA-dependent ATPase I and helicase II	-0.978270544	7.518025927	3.44E-05	0.000864776
b0908	*aroA*	5-enolpyruvylshikimate-3-phosphate synthetase	-0.966731593	7.435982604	6.46E-05	0.001522289
b3182	*dacB*	D-alanyl-D-alanine carboxypeptidase	-0.924773893	7.611253084	4.19E-05	0.001032932
b1283	*osmB*	osmotically and stress inducible lipoprotein	-0.901281405	8.805236501	4.68E-07	1.77E-05
b4169	*amiB*	N-acetylmuramoyl-l-alanine amidase II	-0.870078577	7.419245527	0.000159306	0.003443962
b0951	*pqiB*	paraquat-inducible, SoxRS-regulated MCE domain protein	-0.866118046	7.815328994	4.16E-05	0.001032932
b1805	*fadD*	acyl-CoA synthetase (long-chain-fatty-acid--CoA ligase)	-0.854875141	7.736589524	5.25E-05	0.001266234
b3129	*prlF*	antitoxin of the SohA(PrlF)-YhaV toxin-antitoxin system	-0.852319926	7.237348752	0.000532467	0.009667387
b2829	*ptsP*	PEP-protein phosphotransferase enzyme I; GAF domain containing protein	-0.82713541	8.551114631	6.19E-05	0.001466029
b1422	*ydcI*	putative DNA-binding transcriptional regulator	-0.804020447	8.173811116	1.14E-05	0.000321011
b3260	*dusB*	tRNA-dihydrouridine synthase B	-0.799394108	8.570638918	3.44E-06	0.000110086
b3622	*waaL*	O-antigen ligase	-0.776520848	9.142252999	9.25E-06	0.000271351
b1688	*ydiK*	UPF0118 family inner membrane protein	-0.775073608	7.856567079	0.000122918	0.002727606

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2040	*rfbD*	dTDP-L-rhamnose synthase, NAD(P)-dependent dTDP-4-dehydrorhamnose reductase subunit	-0.676910575	9.419089567	7.65E-06	0.000229033
b4017	*arpA*	ankyrin repeat protein	-0.668534076	8.169577211	0.000397844	0.007516031
b1329	*mppA*	murein tripeptide (L-ala-gamma-D-glutamyl-meso-DAP) transporter subunit	-0.640538419	10.70740394	8.78E-06	0.000260969
b4396	*rob*	right oriC-binding transcriptional activator, AraC family	-0.639000143	8.398526648	0.000256711	0.005078523
b1498	*ydeN*	putative Ser-type periplasmic non-aryl sulfatase	-0.625264807	9.296650257	0.000253191	0.005032623
b0948	*rlmI*	23S rRNA m(2)G2445 and m(7)G2069 methyltransferases, SAM-dependent	-0.555243132	8.549400296	0.000494122	0.009049562
b3872	*yihL*	putative DNA-binding transcriptional regulator	0.579156648	8.507917711	0.000300425	0.005819002
b4049	*dusA*	tRNA-dihydrouridine synthase A	0.580112164	8.540347956	0.000439225	0.008187146
b4159	*mscM*	mechanosensitive channel protein, miniconductance	0.592516104	9.20800717	0.000284333	0.005559586
b1784	*yeaH*	UPF0229 family protein	0.604413991	8.771918993	0.000137613	0.002994899
b2216	*rcsD*	phosphotransfer intermediate protein in two-component regulatory system with RcsBC	0.624163483	9.208025496	0.000113085	0.00252276
b2501	*ppk*	polyphosphate kinase, component of RNA degradosome	0.628703055	8.328575294	0.00051024	0.009304108
b1642	*slyA*	global transcriptional regulator	0.649660641	8.928393489	0.000191696	0.004040076
b3970	*rrlB*		0.654451948	8.31917915	0.000160946	0.003445548
b0431	*cyoB*	cytochrome o ubiquinol oxidase subunit I	0.698985055	8.196075032	0.000459033	0.008480995
b3275	*rrlD*		0.709145476	8.320174819	4.95E-05	0.001200805
b2589	*rrlG*		0.710873751	8.298183781	5.78E-05	0.001386148
b3758	*rrlC*		0.717869755	8.344692451	9.95E-05	0.002243892
b1188	*ycgB*	SpoVR family stationary phase protein	0.717903835	8.653744351	8.84E-06	0.000260969
b3347	*fkpA*	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)	0.724046899	7.778160848	0.000406791	0.007650596
b3084	*rlmG*	23S rRNA m(2)G1835 methyltransferase, SAM-dependent	0.753226202	8.00114335	8.03E-05	0.001853937
b1162	*bluR*	repressor of blue light-responsive genes	0.766774299	7.708999428	0.00024483	0.004936614

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b4089	*alsR*	d-allose-inducible als operon transcriptional repressor; autorepressor; repressor of rpiR	0.771258201	9.391007031	3.26E-06	0.000105268
b3854	*rrlA*		0.782361672	8.363212017	5.23E-06	0.000162573
b3516	*gadX*	acid resistance regulon transcriptional activator; autoactivator	0.784541293	12.07560767	7.33E-07	2.67E-05
b1629	*rsxC*	SoxR iron-sulfur cluster reduction factor component; putative membrane-associated NADH oxidoreductase of electron transport complex	0.796891837	8.214745947	9.61E-06	0.000277882
b0824	*ybiY*	putative pyruvate formate lyase activating enzyme	0.798649928	7.785494564	0.000237626	0.004814513
b2094	*gatA*	galactitol-specific enzyme IIA component of PTS	0.801465726	9.099413514	4.40E-07	1.68E-05
b4009	*rrlE*		0.808532047	8.328767036	3.12E-06	0.000101592
b0204	*rrlH*		0.843219071	8.406456906	5.45E-07	2.04E-05
b4138	*dcuA*	C4-dicarboxylate antiporter	0.849357626	8.239717544	9.89E-06	0.000284173
b2095	*gatZ*	D-tagatose 1,6-bisphosphate aldolase 2, subunit	0.85137098	9.478773181	2.40E-08	1.07E-06
b3546	*eptB*	KDO phosphoethanolamine transferase, Ca(2+)-inducible	0.852124756	9.646212106	7.33E-08	3.01E-06
b3057	*bacA*	undecaprenyl pyrophosphate phosphatase	0.884538375	7.508418141	0.000247223	0.004961012
b1380	*ldhA*	fermentative D-lactate dehydrogenase, NAD-dependent	0.892783767	7.597092528	0.000229131	0.004756512
b4323	*uxuB*	D-mannonate oxidoreductase, NAD-dependent	0.901673842	7.798371995	6.00E-05	0.001428813
b1379	*hslJ*	heat-inducible lipoprotein involved in novobiocin resistance	0.909638125	7.21767908	0.000377907	0.007171685
b2677	*proV*	glycine betaine/proline ABC transporter periplasmic binding protein	0.90969709	10.91370345	3.72E-09	1.83E-07
b0819	*ldtB*	L,D-transpeptidase linking Lpp to murein	0.948952036	8.403290307	6.35E-08	2.64E-06
b3506	*slp*	outer membrane lipoprotein	0.958544376	11.13616288	5.82E-11	3.44E-09
b3740	*rsmG*	16S rRNA m(7)G527 methyltransferase, SAM-dependent; glucose-inhibited cell-division protein	0.97906325	9.392426229	3.36E-10	1.86E-08
b2517	*rlmN*	dual specificity 23S rRNA m(2)A2503, tRNA m(2)A37 methyltransferase, SAM-dependent	1.008919181	7.499823008	3.24E-05	0.000819146
b0406	*tgt*	tRNA-guanine transglycosylase	1.063667114	8.726711664	4.92E-11	2.99E-09
b2893	*dsbC*	protein disulfide isomerase II	1.083088733	7.787870984	2.42E-07	9.50E-06
b3455	*livG*	branched-chain amino acid ABC transporter ATPase	1.090469655	6.973007819	7.23E-05	0.001684905

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b2678	*proW*	glycine betaine/proline ABC transporter permease	1.095003683	8.163904505	3.94E-08	1.69E-06
b3507	*dctR*	Putative LuxR family repressor for dicarboxylate transport	1.09545967	11.2125188	2.17E-13	1.49E-11
b1632	*rsxE*	SoxR iron-sulfur cluster reduction factor component; electron transport inner membrane NADH-quinone reductase	1.127212934	7.036095554	0.000161022	0.003445548
b3166	*truB*	tRNA pseudouridine synthase B: tRNA pseudouridine(55) synthase and putative tmRNA pseudouridine(342) synthase	1.144623942	7.409559236	5.72E-06	0.00017525
b3075	*ebgR*	transcriptional repressor	1.151368264	6.932887096	0.000285005	0.005559586
b1627	*rsxA*	SoxR iron-sulfur cluster reduction factor component; inner membrane protein of electron transport complex	1.169947357	7.792863806	1.83E-06	6.23E-05
b2587	*kgtP*	alpha-ketoglutarate transporter	1.180160956	6.953663593	2.04E-05	0.000532458
b0405	*queA*	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	1.187835883	7.01016126	1.70E-05	0.000459134
b3210	*arcB*	aerobic respiration control sensor histidine protein kinase, cognate to two-component response regulators ArcA and RssB	1.229681142	9.307796892	1.16E-09	6.22E-08
b3512	*gadE*	gad regulon transcriptional activator	1.27174923	8.275450062	2.77E-09	1.38E-07
b2064	*asmA*	suppressor of OmpF assembly mutants; putative outer membrane protein assembly factor; inner membrane-anchored periplasmic protein	1.273247699	8.923374869	2.52E-15	2.00E-13
b1630	*rsxD*	SoxR iron-sulfur cluster reduction factor component; putative membrane protein of electron transport complex	1.278368009	7.264470177	1.05E-05	0.000297558
b0423	*thiI*	tRNA s(4)U8 sulfurtransferase	1.365058738	6.668968298	0.000452952	0.008405661
b1928	*yedD*	lipoprotein	1.365059787	7.782160016	1.75E-10	1.01E-08
b0092	*ddlB*	D-alanine:D-alanine ligase	1.450737591	7.380420321	1.53E-09	7.95E-08
b4381	*deoC*	2-deoxyribose-5-phosphate aldolase, NAD(P)-linked	1.462385413	6.610761752	0.000230644	0.004756512
b3961	*oxyR*	oxidative and nitrosative stress transcriptional regulator	1.478100436	8.236178346	5.26E-15	3.94E-13
b0464	*acrR*	transcriptional repressor	1.513325462	8.393222345	1.48E-14	1.07E-12
b3175	*secG*	preprotein translocase membrane subunit	1.531438013	6.209670516	0.000249808	0.004989028

locus_tag	gene_name	function	logFC	logCPM	PValue	q.value (cutoff of 0.001510574)
b3196	*yrbG*	putative calcium/sodium:proton antiporter	1.601064112	9.21371762	1.14E-23	1.16E-21
b2314	*dedD*	membrane-anchored periplasmic protein involved in septation	1.608640931	7.47363505	2.68E-11	1.65E-09
b1249	*clsA*	cardiolipin synthase 1	1.697342793	8.530837488	2.32E-24	2.50E-22
b0699	*ybfA*	DUF2517 family protein	1.907445419	9.049417885	3.83E-32	5.55E-30
b0015	*dnaJ*	chaperone Hsp40, DnaK co-chaperone	1.943807542	8.585002726	5.75E-32	8.04E-30
b3237	*argR*	l-arginine-responsive arginine metabolism regulon transcriptional regulator	2.10636033	6.475511609	6.03E-09	2.91E-07
b1185	*dsbB*	oxidoreductase that catalyzes reoxidation of DsbA protein disulfide isomerase I	2.188350739	7.309277684	7.08E-17	5.82E-15
b3933	*ftsN*	essential cell division protein	2.213102993	8.392813398	2.05E-18	1.87E-16
b3963	*fabR*	transcriptional repressor of fabA and fabB	2.37776147	8.704663251	2.43E-43	5.09E-41
b3197	*kdsD*	D-arabinose 5-phosphate isomerase	2.458790821	8.567237116	1.63E-17	1.39E-15
b0178	*skp*	periplasmic chaperone	2.510734765	6.023270046	1.45E-06	5.04E-05
b2318	*truA*	tRNA pseudouridine(38-40) synthase	2.721837882	6.541443942	1.80E-12	1.18E-10
b0014	*dnaK*	chaperone Hsp70, with co-chaperone DnaJ	2.981012405	6.115589191	3.92E-07	1.52E-05
b4368	*leuV*		3.021861591	5.198315806	3.13E-05	0.000796542
b3860	*dsbA*	periplasmic protein disulfide isomerase I	3.154766263	9.69454157	6.16E-80	2.87E-77
b3198	*kdsC*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	3.544988468	5.154735084	0.000135511	0.002991218
b3199	*lptC*	periplasmic membrane-anchored LPS-binding protein; LPS export protein	3.883968253	7.019669623	3.40E-33	5.29E-31
b1447	*ydcZ*	DUF606 family inner membrane protein	4.463060881	10.73044663	2.08E-06	6.97E-05
b1855	*lpxM*	myristoyl-acyl carrier protein (ACP)-dependent acyltransferase	6.304596163	8.445517501	1.24E-138	7.41E-136

Table A8. Functional enrichment analysis of genes with significant logFC values following growth in bile (n = 93 genes included)

Gene	Category
ompA	Cell envelope
bepA	Cell envelope
skp	Cell envelope
asmA	Cell envelope
clsA	Cell envelope
yejM	Cell envelope
mlaA	Cell envelope
mlaB	Cell envelope
mlaC	Cell envelope
mlaD	Cell envelope
mlaE	Cell envelope
mlaF	Cell envelope
fabR	Cell envelope
yajG	Cell envelope
yedD	Cell envelope
waaB	LPS
waaG	LPS
waaL	LPS
waaO	LPS
waaP	LPS
waaQ	LPS
waaR	LPS
waaU	LPS
waaY	LPS
waaZ	LPS
kdsC	LPS
kdsD	LPS
lptC	LPS
lpxM	LPS
ybjX	LPS
sanA	PG synthesis and turnover
emtA	PG synthesis and turnover
mltA	PG synthesis and turnover
mipA	PG synthesis and turnover
amiA	PG synthesis and turnover
amiB	PG synthesis and turnover
dacA	PG synthesis and turnover
amiC	PG synthesis and turnover
elyC	PG synthesis and turnover
opgG	Osmoregulation
opgH	Osmoregulation
proQ	Osmoregulation
proW	Osmoregulation
proV	Osmoregulation
nuoG	NADH dehydrogenase
nuoF	NADH dehydrogenase
nuoC	NADH dehydrogenase
nuoB	NADH dehydrogenase
nuoM	NADH dehydrogenase
nuoJ	NADH dehydrogenase
nuoL	NADH dehydrogenase
nuoH	NADH dehydrogenase
nuoN	NADH dehydrogenase
nuoI	NADH dehydrogenase
nuoA	NADH dehydrogenase
phoP	Bile tolerance
phoQ	Bile tolerance
acrA	Bile tolerance
acrB	Bile tolerance
tolC	Bile tolerance

Gene	Category
rfaH	LPS
lpxT	LPS
yrbG	LPS
wecA	ECA
wecB	ECA
wecC	ECA
wecD	ECA
wecE	ECA
wzzE	ECA
yraP	Cell division
envC	Cell division
minC	Cell division
ftsN	Cell division
tolA	Cell division
nlpD	Cell division
ftsP	Cell division
ygeR	Cell division
nlpI	Cell division
prc	Cell division
dedD	Cell division
damX	Cell division

Gene	Category
acrR	Bile tolerance
pqiB	Oxidative stress
rsxC	Oxidative stress
rsxE	Oxidative stress
rsxA	Oxidative stress
rsxD	Oxidative stress
dam	DNA replication/modification
rnhA	DNA replication/modification
dcm	DNA replication/modification
uup	DNA replication/modification
uvrD	DNA replication/modification

Table A9. End-point growth of Keio library mutants vs TraDIS logFCs

Gene	logFC	Gene	logFC	Key
acrA	-13.45762217	*yihL*	0.579156648	Significantly greater OD than plate average/positive logFC
phoQ	-11.49034613	*dusA*	0.580112164	Significantly lower OD than plate average/negative logFC
tolC	-11.04748427	*mscM*	0.592516104	
phoP	-10.99210588	*yeaH*	0.604413991	
qseC	-10.6742087	*rcsD*	0.624163483	
waaY	-10.57156247	*ppk*	0.628703055	
waaQ	-10.53616871	*slyA*	0.649660641	
nuoG	-10.51441405	*rrlB*	0.654451948	
acrB	-10.45658313	*cyoB*	0.698985055	
nuoF	-9.991423561	*rrlD*	0.709145476	
nuoC	-9.710277668	*rrlG*	0.710873751	
nuoB	-9.408109621	*rrlC*	0.717869755	
lpp	-9.360655383	*ycgB*	0.717903835	
nuoM	-9.358471301	*fkpA*	0.724046899	
nuoJ	-9.127948557	*rlmG*	0.753226202	
nuoL	-9.099262266	*bluR*	0.766774299	
nuoH	-9.077033141	*alsR*	0.771258201	
nuoN	-9.064111238	*rrlA*	0.782361672	
wecD	-8.995135148	*gadX*	0.784541293	
dacA	-8.778923104	*rsxC*	0.796891837	
nlpD	-8.597085736	*ybiY*	0.798649928	
lapA	-8.385769457	*gatA*	0.801465726	
tatA	-8.359027723	*rrlE*	0.808532047	
waaG	-8.33236493	*rrlH*	0.843219071	
tatC	-8.317076347	*dcuA*	0.849357626	
nuoI	-8.31115119	*gatZ*	0.85137098	
waaO	-8.058461143	*eptB*	0.852124756	
mipA	-7.875083818	*bacA*	0.884538375	
wecE	-7.35178421	*ldhA*	0.892783767	
sspA	-7.338341577	*uxuB*	0.901673842	

Gene	logFC	Gene	logFC	Key
envC	-6.864721514	*hslJ*	0.909638125	Significantly greater OD than plate average/positive logFC
rnhA	-6.466952214	*proV*	0.90969709	Significantly lower OD than plate average/negative logFC
wecA	-6.345422176	*ldtB*	0.948952036	
damX	-6.295504459	*slp*	0.958544376	
dam	-6.191379902	*rsmG*	0.97906325	
rpoS	-6.148878632	*rlmN*	1.008919181	
wecB	-6.065042639	*tgt*	1.063667114	
tolA	-5.930362556	*dsbC*	1.083088733	
waaB	-5.552127939	*livG*	1.090469655	
rfaH	-5.489032561	*proW*	1.095003683	
mlaA	-5.249988062	*dctR*	1.09545967	
mlaE	-5.17713926	*rsxE*	1.127212934	
mlaF	-5.142729773	*truB*	1.144623942	
gcvH	-4.868416243	*ebgR*	1.151368264	
mepM	-4.86038371	*rsxA*	1.169947357	
fimE	-4.74222024	*kgtP*	1.180160956	
cpxR	-4.732320293	*queA*	1.187835883	
prc	-4.721561537	*arcB*	1.229681142	
amiA	-4.666784349	*gadE*	1.27174923	
rpoZ	-4.525412657	*asmA*	1.273247699	
opgG	-4.428565763	*rsxD*	1.278368009	
yejM	-4.379544394	*thiI*	1.365058738	
mlaB	-4.371585805	*yedD*	1.365059787	
yraP	-4.292027217	*ddlB*	1.450737591	
nuoA	-4.275454887	*deoC*	1.462385413	
emtA	-4.172616843	*oxyR*	1.478100436	
wecC	-4.07503752	*acrR*	1.513325462	
mlaC	-4.063917144	*secG*	1.531438013	
amiC	-3.972138737	*yrbG*	1.601064112	
lpxT	-3.659447131	*dedD*	1.608640931	
wzzE	-3.513526463	*clsA*	1.697342793	

Gene	logFC	Gene	logFC	Key
				Significantly greater OD than plate average/positive logFC
opgH	-3.467307771	*ybfA*	1.907445419	
				Significantly lower OD than plate average/negative logFC
yrbL	-3.46436912	*dnaJ*	1.943807542	
ompA	-3.370508115	*argR*	2.10636033	
ydgH	-3.270032231	*dsbB*	2.188350739	
iscR	-3.204867193	*ftsN*	2.213102993	
dedA	-3.159649916	*fabR*	2.37776147	
fbp	-3.126913274	*kdsD*	2.458790821	
cvpA	-3.088368482	*skp*	2.510734765	
ybjX	-3.040792035	*truA*	2.721837882	
waaP	-2.962906875	*dnaK*	2.981012405	
yceG (mltG)	-2.934249077	*leuV*	3.021861591	
mlaD	-2.824128901	*dsbA*	3.154766263	
rbsR	-2.741790731	*kdsC*	3.544988468	
nlpI	-2.692187914	*lptC*	3.883968253	
ygeR	-2.677389398	*ydcZ*	4.463060881	
tig	-2.547921418	*lpxM*	6.304596163	
fepB	-2.371796342			
yafD	-2.334268969			
yfcD	-2.249695085			
ompC	-2.113010415			
lrp	-2.111372737			
galE	-2.072225453			
sanA	-2.042948636			
relA	-1.913792193			
elyC	-1.839529726			
mltA	-1.791409174			
fes	-1.783776862			
minC	-1.769927978			
yfiH	-1.764334883			
dsdA	-1.762039446			
gcvP	-1.727409141			

Gene	logFC
yajG	-1.695713511
rlpA	-1.687315928
pitA	-1.620857596
dcm	-1.608714021
gcvT	-1.573287896
ftsP	-1.530552824
lpoA	-1.519435943
bepA	-1.505296613
ygcO	-1.468248473
ampG	-1.3638939
dsrA	-1.348764807
proQ	-1.29201986
metL	-1.268421934
waaZ	-1.242886553
uup	-1.167099607
waaR	-1.164278643
flhE	-1.104199313
yggN	-1.062743664
crl	-1.059985162
alpA	-1.023915082
ydeS	-1.02006805
waaU	-1.01767509
yieP	-0.98741311
selB	-0.981505802
yoaH	-0.980819156
galR	-0.978401643
uvrD	-0.978270544
aroA	-0.966731593
dacB	-0.924773893
osmB	-0.901281405
amiB	-0.870078577

Key
Significantly greater OD than plate average/positive logFC
Significantly lower OD than plate average/negative logFC

Gene	logFC
pqiB	-0.866118046
fadD	-0.854875141
prlF	-0.852319926
ptsP	-0.82713541
ydcI	-0.804020447
dusB	-0.799394108
waaL	-0.776520848
ydiK	-0.775073608
rfbD	-0.676910575
arpA	-0.668534076
mppA	-0.640538419
rob	-0.639000143
ydeN	-0.625264807
rlmI	-0.555243132

Key
Significantly greater OD than plate average/positive logFC
Significantly lower OD than plate average/negative logFC

Table A10. Comparative analysis of bile tolerance in *E. coli* and *S.* Typhi using TraDIS. Genes under negative selection uniquely in *E. coli*

alpA
amiA
amiB
amiC
aroA
arpA
bepA
crl
dacB
dcm
dsrA
dusB
elyC
emtA
envC
fadD
fbp
fepB
fes
fimE
flhE
galE
gcvH
gcvP
gcvT
iscR
lapA
lpoA
lpp
lpxT
lrp
mepM
metL
minC
mipA
mlaB
mlaC
mlaD
mlaE
mlaF
mppA
nlpD
nlpI
nuoA
nuoB
nuoC
nuoF
nuoG
nuoH
nuoI
nuoJ
nuoL
nuoM
nuoN
opgG
opgH
osmB
pitA
pqiB
prlF
proQ
qseC
rbsR
rfaH
rlmI
rlpA
rnhA
rpoS
rpoZ
selB
tatA
tatC
tolA
uup
waaZ
wecD
wecE
wzzE
yafD
yajG
yceG (mltG)
ydcI
ydeN
ydeS
ydgH
ydiK
yfcD
yfiH (pgeF)
ygcO
ygeR
yggN
yieP
yoaH
yraP
yrbL

Table A11. Comparative analysis of bile tolerance in *E. coli* and *S.* Typhi using TraDIS.
Genes under negative selection uniquely in *S*. Typhi

Gene	Function
adh	Central metabolism
frsA	Central metabolism
gltA	Central metabolism
pfkA	Central metabolism
pflA	Central metabolism
pflB	Central metabolism
ppc	Central metabolism
pagP	LPS
pgi	O antigen
rfbA	O antigen
rfbB	O antigen
rfbC	O antigen
rfbE	O antigen
rfbH	O antigen
rfbI	O antigen
rfbK	O antigen
rfbP	O antigen
rfbU	O antigen
wbbH	O antigen
rfbM	O antigen
wzzB	O antigen
oxyR	Redox/protein folding
ahpC	Redox/protein folding
djlA	Redox/protein folding
dsbD	Redox/protein folding
fkpA	Redox/protein folding
hslU	Redox/protein folding
ldtB	Peptidogycan/cell division
mltC	Peptidogycan/cell division
mltD	Peptidogycan/cell division
mrcA	Peptidogycan/cell division
mrcB	Peptidogycan/cell division
slt	Peptidogycan/cell division
zapB	Peptidogycan/cell division
lepA	Amino acid metabolism/translation
tyrR	Amino acid metabolism/translation
mnmE	Amino acid metabolism/translation
sdaA	Amino acid metabolism/translation
smpB	Amino acid metabolism/translation
tgt	Amino acid metabolism/translation
gcvA	Amino acid metabolism/translation
hfq	DNA/RNA
seqA	DNA/RNA
yejK	DNA/RNA
hupA	DNA/RNA
sbcB	DNA/RNA
xseA	DNA/RNA
pcnB	DNA/RNA
topB	DNA/RNA
yejH	DNA/RNA
pnp	DNA/RNA
rph	DNA/RNA
radA	DNA/RNA
rapA	DNA/RNA

ridA	Redox/protein folding
sodA	Redox/protein folding
ytfK	Redox/protein folding
aphA	Nucleotide metabolism
deoB	Nucleotide metabolism
guaA	Nucleotide metabolism
guaB	Nucleotide metabolism
nadR	Nucleotide metabolism
pncB	Nucleotide metabolism
purC	Nucleotide metabolism
purH	Nucleotide metabolism
sthA	Nucleotide metabolism
rcsC	Cell envelope
rcsD	Cell envelope
clsA	Cell envelope
cpxA	Cell envelope
fabR	Cell envelope
ompX	Cell envelope
rseB	Cell envelope
tamA	Cell envelope
tamB	Cell envelope
rffM	Cell envelope
yrfF	Cell envelope
cysW	Transporters
cysA	Transporters
trkH	Transporters
proP	Transporters
mgtA	Transporters
glpT	Transporters
treB	Carbon metabolism
rmuC	DNA/RNA
recD	DNA/RNA
recG	DNA/RNA
recJ	DNA/RNA
virK	S. Typhi-specific
t2932	S. Typhi-specific
t0641	S. Typhi-specific
t2899	S. Typhi-specific
t3184	S. Typhi-specific
t1165	S. Typhi-specific
t1847	S. Typhi-specific
t0533	S. Typhi-specific
t3230	S. Typhi-specific
t1442	S. Typhi-specific
t1077	S. Typhi-specific
t1344	S. Typhi-specific
t1707	S. Typhi-specific
t0642	S. Typhi-specific
t0412	S. Typhi-specific
t2603	S. Typhi-specific
t1040	S. Typhi-specific
t2964	S. Typhi-specific
t1398	S. Typhi-specific
t3099	S. Typhi-specific
t1460	S. Typhi-specific
t0335	S. Typhi-specific
t3015	S. Typhi-specific
t1486	S. Typhi-specific
t3199	S. Typhi-specific
t2206	S. Typhi-specific

barA	Carbon metabolism
uvrY	Carbon metabolism
kdgR	Carbon metabolism
cpdA	Carbon metabolism
glpD	Carbon metabolism
rbsK	Carbon metabolism
ulaR	Carbon metabolism
manA	Carbon metabolism
gntR	Carbon metabolism

t4122	S. Typhi-specific
cspC	Misc
cspE	Misc

Table A12. Essential genes following growth under anaerobic conditions

Essential genes in glucose	Essential genes in glucose nitrate		Essential genes in glycerol nitrate
aceE	*aceE*	*rpmG*	*apaH*
ackA	*ackA*	*rpoZ*	*arcA*
acnB	*acnB*	*rppH*	*argA*
adhE	*apaG*	*rseX*	*argB*
apaH	*apaH*	*rsmA*	*argC*
argA	*appY*	*rsmD*	*argE*
argB	*argA*	*rsmH*	*argG*
argC	*argB*	*rttR*	*aroD*
argE	*argC*	*ruvA*	*asnT*
argG	*argD*	*ruvB*	*aspC*
argH	*argE*	*ruvC*	*atpB*
argR	*argG*	*rydB*	*atpD*
argW	*argH*	*sapC*	*clpP*
aroA	*argQ*	*sapD*	*clpX*
aroB	*argR*	*sapF*	*crr*
aroD	*argW*	*secM*	*cvpA*
aroE	*ariR*	*serW*	*cyaA*
aroK	*aroA*	*sgrT*	*cysA*
arrS	*aroB*	*shoB*	*cysD*
aspC	*aroD*	*sixA*	*cysG*
atpA	*aroE*	*smpB*	*cysN*
atpB	*aroK*	*sokA*	*cysP*
atpC	*aroL*	*spoT*	*cysQ*
atpD	*arrS*	*ssrA*	*cysU*
bioB	*asnU*	*tatC*	*cysW*
carA	*asnV*	*thrA*	*dsbB*
carB	*aspC*	*thrB*	*eda*
clpB	*atpA*	*thrC*	*epmA*
clpP	*atpB*	*tktA*	*fbp*
crr	*atpC*	*tolA*	*fepB*
cvpA	*atpD*	*tolB*	*fepC*

Essential genes in glucose	Essential genes in glucose nitrate		Essential genes in glycerol nitrate
cysA	*bhsA*	*tolC*	*fepD*
cysD	*bioA*	*tolQ*	*fepG*
cysG	*bioB*	*tolR*	*fes*
cysH	*bioC*	*tonB*	*fliQ*
cysI	*bioD*	*tpr*	*fnr*
cysK	*bioF*	*trkA*	*fur*
cysN	*carA*	*trkH*	*galU*
cysP	*carB*	*trpA*	*gcvR*
cysQ	*cfa*	*trpC*	*glpF*
cysU	*citX*	*trpD*	*glpK*
cysW	*clpP*	*trpE*	*gltA*
dcd	*clpX*	*trpL*	*hemN*
dsbA	*copA*	*truA*	*higA*
dsbB	*corA*	*trxA*	*hisG*
eda	*cpxA*	*trxB*	*hisI*
envC	*crr*	*ttcC*	*hisL*
fepD	*cspI*	*tusC*	*ibsD*
ftsB	*cvpA*	*tyrA*	*ihfA*
ftsE	*cysA*	*tyrU*	*ihfB*
fur	*cysD*	*uof*	*ilvA*
galU	*cysG*	*usg*	*ilvB*
gcvR	*cysH*	*wzb*	*ilvC*
glgA	*cysI*	*xerC*	*ilvD*
glnV	*cysK*	*xerD*	*ilvE*
glnX	*cysM*	*xseB*	*iscA*
gltA	*cysN*	*yafN*	*ivbL*
gpmA	*cysP*	*ybaM*	*lepA*
hdfR	*cysQ*	*ybbC*	*leuA*
higA	*cysU*	*ybcK*	*leuB*
hisA	*cysW*	*ybcV*	*leuC*
hisD	*dcd*	*ybfB*	*leuD*
hisG	*dgkA*	*ybfC*	*leuT*
hisI	*dsbA*	*ybfQ*	*leuV*

Essential genes in glucose	Essential genes in glucose nitrate		Essential genes in glycerol nitrate
hns	*dsbB*	*yccE*	*leuX*
ibsD	*eda*	*yceD*	*lpp*
ilvA	*elaD*	*yceO*	*lrp*
ilvC	*envC*	*yceQ*	*lysA*
ilvD	*exbD*	*ycgX*	*lysR*
ilvE	*fepB*	*ychF*	*lysZ*
ilvX	*fepC*	*yciB*	*metA*
ilvY	*fepD*	*yciE*	*metB*
iscA	*fepG*	*yciG*	*metC*
ldhA	*fes*	*ydeO*	*metF*
leuA	*fnr*	*ydfO*	*metR*
leuB	*fpr*	*ydiE*	*metY*
leuC	*ftsB*	*yedN*	*mgtS*
leuD	*ftsE*	*yehA*	*minE*
lipA	*fur*	*yehC*	*moaA*
lipB	*galU*	*yehD*	*moaC*
lpp	*gcvR*	*yejL*	*moaD*
lrp	*gdhA*	*yejM*	*moaE*
lysA	*glnV*	*yfbN*	*mobA*
lysR	*glnX*	*yfdF*	*modA*
metA	*gltA*	*yfjI*	*modB*
metB	*glyY*	*yfjW*	*moeA*
metC	*gtrS*	*ygeH*	*moeB*
metF	*hflD*	*ygeI*	*mog*
metJ	*hfq*	*ygeN*	*nadA*
metR	*higA*	*yhaC*	*nadC*
metY	*hinT*	*yhbJ*	*narG*
minE	*hisA*	*yheO*	*narH*
nadA	*hisD*	*yhhH*	*narI*
nadB	*hisG*	*yhhK*	*nuoB*
nadC	*hisI*	*yhhZ*	*nuoC*
ndh	*hisL*	*yhiL*	*nuoE*
nudB	*hns*	*yhiS*	*nuoF*

Essential genes in glucose	Essential genes in glucose nitrate		Essential genes in glycerol nitrate
nuoB	*ibsD*	*yibG*	*nuoG*
nuoE	*ilvA*	*yibV*	*nuoH*
nuoF	*ilvB*	*yibW*	*nuoI*
nuoG	*ilvC*	*yigG*	*nuoJ*
ompA	*ilvD*	*yjbM*	*nuoK*
ompR	*ilvE*	*yjcF*	*nuoL*
pal	*ilvX*	*yjdK*	*nuoM*
pdxA	*ilvY*	*yjeA*	*nuoN*
pdxB	*iscA*	*yjgL*	*ompA*
pgi	*ivbL*	*yjhB*	*ompR*
pgl	*kilR*	*yjiC*	*pal*
pgm	*lepA*	*ykgH*	*pdxB*
pheA	*leuA*	*ykgP*	*pgi*
pheL	*leuB*	*ykgR*	*pheA*
ppc	*leuC*	*ymfD*	*pheL*
ppk	*leuD*	*ymfE*	*pmrR*
prc	*leuP*	*ymgC*	*ppc*
proA	*leuT*	*yncI*	*prc*
proC	*leuX*	*yneG*	*proA*
psrO	*lipA*	*yneM*	*proC*
pstB	*lipB*	*ynfK*	*proK*
pstS	*lpp*	*ynfN*	*psaA*
pta	*lrp*	*yniD*	*psrO*
ptsG	*lysA*	*ypjJ*	*pstC*
ptsN	*lysR*	*yqcG*	*ptsN*
purE	*lysZ*	*yqeH*	*purN*
purK	*metA*	*yqeJ*	*rapZ*
pyrB	*metB*	*yqeL*	*rdlC*
pyrC	*metC*	*ytfI*	*relA*
pyrD	*metF*	*znuB*	*relB*
pyrE	*metR*	*zwf*	*rfaH*
pyrF	*mgrB*		*rfaP*
rdlA	*miaA*		*rng*

Essential genes in glucose	Essential genes in glucose nitrate	Essential genes in glycerol nitrate
recA	*micA*	*rnhA*
recC	*minE*	*rnlB*
relA	*mnmG*	*rpoZ*
relB	*moaA*	*rseA*
rfaC	*moaC*	*ruvB*
rfaG	*moaE*	*ruvC*
rfaH	*mobA*	*sapB*
rfaP	*modC*	*sapC*
rng	*moeA*	*sapD*
rnhA	*moeB*	*sapF*
rnlB	*mog*	*serU*
rpe	*mrcB*	*serW*
rpmG	*nadA*	*smpB*
rppH	*nadB*	*spoT*
ruvB	*nadC*	*sspA*
ruvC	*narI*	*ssrA*
sapC	*ndh*	*tatC*
sapD	*nhaA*	*thrA*
sapF	*nohQ*	*thrB*
secM	*nudB*	*thrL*
sixA	*nuoJ*	*tktA*
smpB	*ompA*	*tolA*
speE	*ompR*	*tolB*
sspA	*pabA*	*tolC*
ssrA	*pal*	*tolR*
sucC	*panB*	*tpr*
sucD	*panC*	*trkA*
thrA	*panD*	*trpC*
thrB	*pdxA*	*trpD*
thrC	*pdxB*	*trpE*
tktA	*pgi*	*truA*
tolA	*pgl*	*trxA*
tolB	*pgm*	*tyrA*

Essential genes in glucose	Essential genes in glucose nitrate	Essential genes in glycerol nitrate
tolC	*pheA*	*tyrU*
tolQ	*pheL*	*ulaR*
tolR	*pitA*	*valW*
trkA	*ppc*	*xerC*
trpA	*proA*	*yafN*
trpC	*proC*	*ybaM*
trpD	*pstB*	*yceD*
trpE	*pstC*	*yceQ*
trpL	*pstS*	*yecJ*
truA	*pta*	*ygeI*
trxA	*ptsG*	*yhcB*
trxB	*ptsN*	*yheO*
ttcC	*purE*	*yhfG*
tusC	*purK*	*ykfN*
tyrA	*purN*	*yohO*
uof	*pyrB*	*ypaA*
xerC	*pyrC*	*ypdK*
xerD	*pyrD*	*yqcG*
ybbC	*pyrE*	*yraP*
ybfQ	*pyrF*	
yccE	*racC*	
yceQ	*rbfA*	
yfcE	*rdlA*	
yffS	*rdlC*	
ygeI	*recA*	
ygeN	*recC*	
yhbJ	*recG*	
yifE	*relA*	
yjeA	*relB*	
yjeF	*rfaC*	
ymfE	*rfaG*	
ymgC	*rfaP*	
yqcG	*rimP*	

Essential genes in glucose	Essential genes in glucose nitrate
yqeJ	*rnb*
yqeL	*rng*
znuB	*rnhA*
zwf	*rpe*

Table A13. Genes with significant logFC values following growth under anaerobic conditions Glucose

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b3008	*metC*	cystathionine beta-lyase, PLP-dependent	-9.877686805	5.960575564	6.79E-23	3.17E-20
b1262	*trpC*	fused indole-3-glycerolphosphate synthetase/N-(5-phosphoribosyl)anthranilate isomerase	-9.570003527	5.662675708	6.31E-20	2.21E-17
b3939	*metB*	cystathionine gamma-synthase, PLP-dependent	-9.112287133	5.225864686	2.88E-14	4.03E-12
b0750	*nadA*	quinolinate synthase, subunit A	-9.004585791	5.123251498	4.34E-16	9.10E-14
b0071	*leuD*	3-isopropylmalate dehydratase small subunit	-8.992720357	5.112257624	4.64E-16	9.27E-14
b0073	*leuB*	3-isopropylmalate dehydrogenase, NAD(+)-dependent	-8.934201588	5.057164605	3.12E-15	5.24E-13
b0889	*lrp*	DNA-binding transcriptional dual regulator, leucine-binding	-8.702971266	4.840150774	3.99E-14	5.08E-12
b0720	*gltA*	citrate synthase	-8.688344067	4.826801145	9.52E-14	1.11E-11
b0740	*tolB*	periplasmic protein	-8.658008392	4.799547754	4.62E-09	2.69E-07
b2838	*lysA*	diaminopimelate decarboxylase, PLP-binding	-8.3221699	4.488213521	1.65E-10	1.24E-08
b3771	*ilvD*	dihydroxyacid dehydratase	-8.284034367	4.454113845	1.56E-11	1.34E-09
b3172	*argG*	argininosuccinate synthetase	-8.2641262	4.436609082	1.05E-09	7.09E-08
b1241	*adhE*	fused acetaldehyde-CoA dehydrogenase/iron-dependent alcohol dehydrogenase/pyruvate-formate lyase deactivase	-8.12711882	7.872053704	7.91E-85	3.32E-81
b1677	*lpp*	murein lipoprotein	-7.900902967	4.109992488	2.60E-09	1.63E-07
b1236	*galU*	glucose-1-phosphate uridylyltransferase	-7.878700631	4.088975404	1.21E-07	5.58E-06
b3774	*ilvC*	ketol-acid reductoisomerase, NAD(P)-binding	-7.876510967	4.088309685	2.42E-09	1.56E-07
b0243	*proA*	gamma-glutamylphosphate reductase	-7.82690662	4.044089608	3.51E-09	2.08E-07
b3829	*metE*	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	-7.82644191	4.043940389	2.74E-09	1.69E-07
b0003	*thrB*	homoserine kinase	-7.826208391	4.043865783	3.52E-09	2.08E-07
b3770	*ilvE*	branched-chain amino-acid aminotransferase	-7.774685514	3.998199548	5.97E-09	3.43E-07
b3734	*atpA*	F1 sector of membrane-bound ATP synthase, alpha subunit	-7.665041285	3.902036552	1.16E-07	5.40E-06
b3082	*higA*	antitoxinof the HigB-HigA toxin-antitoxin system	-7.636458714	3.876997395	4.98E-07	2.13E-05
b2557	*purL*	phosphoribosylformyl-glycineamide synthetase	-7.574807889	3.824873226	2.77E-08	1.47E-06
b0738	*tolR*	membrane spanning protein in TolA-TolQ-TolR complex	-7.572592923	3.82432976	7.04E-06	0.000226398
b3941	*metF*	5,10-methylenetetrahydrofolate reductase	-7.511670273	3.771142666	1.62E-07	7.41E-06

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b4214	*cysQ*	PAPS (adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase	-7.482096727	3.744301418	1.71E-06	6.63E-05
b0052	*pdxA*	4-hydroxy-L-threonine phosphate dehydrogenase, NAD-dependent	-7.415258488	3.687584547	1.90E-06	7.20E-05
b0688	*pgm*	phosphoglucomutase	-7.415073771	3.68754482	1.18E-06	4.87E-05
b0741	*pal*	peptidoglycan-associated outer membrane lipoprotein	-7.377321087	3.657572462	3.29E-06	0.000120244
b1101	*ptsG*	fused glucose-specific PTS enzymes: IIB component/IIC component	-7.297952271	5.706412079	1.01E-16	2.50E-14
b0386	*proC*	pyrroline-5-carboxylate reductase, NAD(P)-binding	-7.230981929	3.534708156	1.37E-06	5.42E-05
b3738	*atpB*	F0 sector of membrane-bound ATP synthase, subunit a	-7.193334958	3.502621757	1.25E-06	5.09E-05
b2026	*hisI*	fused phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase	-7.191195924	3.502131514	5.26E-06	0.000178101
b0908	*aroA*	5-enolpyruvylshikimate-3-phosphate synthetase	-7.15347799	3.469505817	4.03E-06	0.000141899
b2414	*cysK*	cysteine synthase A, O-acetylserine sulfhydrolase A subunit	-7.112094744	3.43552738	6.83E-06	0.000222436
b0004	*thrC*	threonine synthase	-7.068107284	3.400440397	1.89E-06	7.20E-05
b3732	*atpD*	F1 sector of membrane-bound ATP synthase, beta subunit	-7.023007051	3.364533461	5.53E-06	0.000184365
b2320	*pdxB*	erythronate-4-phosphate dehydrogenase	-6.979605952	3.328391368	9.38E-06	0.000287369
b2574	*nadB*	quinolinate synthase, L-aspartate oxidase (B protein) subunit	-6.960080683	5.387086801	2.51E-16	5.55E-14
b0242	*proB*	gamma-glutamate kinase	-6.883184995	3.252122544	6.38E-06	0.000209413
b4025	*pgi*	glucosephosphate isomerase	-6.874187571	5.303674869	9.79E-16	1.87E-13
b2762	*cysH*	3'-phosphoadenosine 5'-phosphosulfate reductase	-6.834755334	3.212856519	7.49E-05	0.001930143
b2296	*ackA*	acetate kinase A and propionate kinase 2	-6.672237763	3.087050515	7.13E-05	0.001848065
b0032	*carA*	carbamoyl phosphate synthetase small subunit, glutamine amidotransferase	-6.669598307	3.086576331	7.01E-05	0.001828404
b3916	*pfkA*	6-phosphofructokinase I	-6.551968573	2.99651939	4.85E-05	0.001296019
b3860	*dsbA*	periplasmic protein disulfide isomerase I	-6.352981601	7.057432395	1.79E-34	1.25E-31
b0957	*ompA*	outer membrane protein A (3a;II*;G;d)	-6.130712955	4.615455605	6.58E-10	4.61E-08
b0074	*leuA*	2-isopropylmalate synthase	-5.838044041	5.647205694	1.13E-16	2.63E-14
b0729	*sucD*	succinyl-CoA synthetase, NAD(P)-binding, alpha subunit	-5.776216715	6.714728724	1.33E-35	1.12E-32
b3828	*metR*	DNA-binding transcriptional activator	-5.51646693	4.066183682	3.08E-08	1.61E-06

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b3035	*tolC*	transport channel	-5.353209404	5.185658504	9.63E-13	9.63E-11
b0109	*nadC*	quinolinate phosphoribosyltransferase	-5.317714739	4.659802707	7.56E-11	5.99E-09
b0002	*thrA*	fused aspartokinase I and homoserine dehydrogenase I	-5.284121069	5.498559803	1.25E-14	1.87E-12
b2839	*lysR*	DNA-binding transcriptional dual regulator	-5.260991633	5.760905306	3.54E-15	5.72E-13
b0728	*sucC*	succinyl-CoA synthetase, beta subunit	-5.258319888	6.560345317	2.59E-27	1.36E-24
b3772	*ilvA*	threonine deaminase	-5.249993753	5.090254374	4.65E-12	4.44E-10
b2752	*cysD*	sulfate adenylyltransferase, subunit 2	-5.24209839	5.09123256	4.89E-10	3.54E-08
b3960	*argH*	argininosuccinate lyase	-5.197876611	3.798178878	1.48E-06	5.80E-05
b1274	*topA*	DNA topoisomerase I, omega subunit	-5.030603652	3.658171887	3.09E-06	0.000113892
b2313	*cvpA*	membrane protein required for colicin V production	-4.996645249	3.627671696	8.47E-06	0.000263298
b1263	*trpD*	fused glutamine amidotransferase (component II) of anthranilate synthase/anthranilate phosphoribosyl transferase	-4.967863743	4.84040049	5.09E-10	3.62E-08
b2526	*hscA*	DnaK-like molecular chaperone specific for IscU	-4.956951853	3.597911834	4.64E-06	0.000159784
b2019	*hisG*	ATP phosphoribosyltransferase	-4.950454069	4.329668247	2.17E-08	1.17E-06
b3956	*ppc*	phosphoenolpyruvate carboxylase	-4.863018325	4.745887213	4.33E-10	3.19E-08
b1264	*trpE*	component I of anthranilate synthase	-4.811190017	4.214310622	5.50E-08	2.78E-06
b1693	*aroD*	3-dehydroquinate dehydratase	-4.80050788	3.468888593	1.36E-05	0.000405783
b2020	*hisD*	bifunctional histidinal dehydrogenase/ histidinol dehydrogenase	-4.797627706	3.469424595	1.98E-05	0.000575367
b0072	*leuC*	3-isopropylmalate dehydratase large subunit	-4.788054198	5.034248852	1.06E-11	9.25E-10
b2751	*cysN*	sulfate adenylyltransferase, subunit 1	-4.662832897	5.216618836	5.36E-11	4.33E-09
b3631	*rfaG*	glucosyltransferase I	-4.647271826	6.592416267	2.74E-28	1.64E-25
b0739	*tolA*	membrane anchored protein in TolA-TolQ-TolR complex	-4.542624532	4.453088306	4.80E-07	2.08E-05
b1830	*prc*	carboxy-terminal protease for penicillin-binding protein 3	-4.511852143	5.861861158	2.66E-14	3.85E-12
b2340	*sixA*	phosphohistidine phosphatase	-4.459348753	4.746300899	3.50E-08	1.81E-06
b2763	*cysI*	sulfite reductase, beta subunit, NAD(P)-binding, heme-binding	-4.41734534	3.876612531	5.35E-06	0.000179601
b0049	*apaH*	diadenosine tetraphosphatase	-4.270757477	5.089450706	8.01E-10	5.51E-08
b2592	*clpB*	protein disaggregation chaperone	-4.236488634	7.913801793	4.30E-53	6.01E-50
b2501	*ppk*	polyphosphate kinase, component of RNA degradosome	-4.225773438	7.792660693	7.19E-14	8.62E-12

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b3368	*cysG*	fused siroheme synthase 1,3-dimethyluroporphyriongen III dehydrogenase and siroheme ferrochelatase/uroporphyrinogen methyltransferase	-4.157578102	5.000667134	9.40E-08	4.43E-06
b1185	*dsbB*	oxidoreductase that catalyzes reoxidation of DsbA protein disulfide isomerase I	-4.14564326	5.522307594	3.68E-12	3.59E-10
b4155	*yjeA*	EF-P-lysine34-lysine ligase	-4.066552616	4.902934609	2.08E-07	9.29E-06
b0033	*carB*	carbamoyl-phosphate synthase large subunit	-4.049365814	4.903087638	3.78E-07	1.65E-05
b0118	*acnB*	bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase	-3.928822854	4.801304773	4.16E-05	0.001118572
b2423	*cysW*	sulfate/thiosulfate ABC transporter subunit	-3.906285273	4.976681305	1.18E-08	6.49E-07
b3247	*rng*	ribonuclease G	-3.904650104	6.093117989	8.34E-13	8.54E-11
b0239	*frsA*	fermentation-respiration switch protein; PTS Enzyme IIA(Glc)-binding protein; pNP-butyrate esterase activity	-3.859524415	7.675827373	1.45E-47	1.52E-44
b1850	*eda*	multifunctional 2-keto-3-deoxygluconate 6-phosphate aldolase and 2-keto-4-hydroxyglutarate aldolase and oxaloacetate decarboxylase	-3.823018333	3.851012825	2.32E-05	0.000645956
b1380	*ldhA*	fermentative D-lactate dehydrogenase, NAD-dependent	-3.768671639	6.357721025	6.05E-08	3.02E-06
b3781	*trxA*	thioredoxin 1	-3.688585478	4.772862575	1.30E-05	0.000390414
b2822	*recC*	exonuclease V (RecBCD complex), gamma chain	-3.627512593	4.731711443	9.06E-08	4.32E-06
b3613	*envC*	activator of AmiB,C murein hydrolases, septal ring factor	-3.622443434	6.718019614	3.07E-13	3.40E-11
b2300	*yfcE*	phosphodiesterase activity on bis-pNPP	-3.603994448	5.302661246	2.22E-09	1.46E-07
b1857	*znuA*	zinc transporter subunit: periplasmic-binding component of ABC superfamily	-3.558226292	6.653358793	1.67E-15	3.05E-13
b2784	*relA*	(p)ppGpp synthetase I/GTP pyrophosphokinase	-3.492474935	7.333602488	9.01E-12	8.22E-10
b3205	*yhbJ*	glmZ(sRNA)-inactivating NTPase, glucosamine-6-phosphate regulated	-3.482728587	5.430107511	1.04E-08	5.85E-07
b3839	*tatC*	TatABCE protein translocation system subunit	-3.449558009	6.594080216	8.53E-17	2.24E-14
b3290	*trkA*	NAD-binding component of TrK potassium transporter	-3.21585534	4.38397544	3.72E-06	0.000133509
b3429	*glgA*	glycogen synthase	-3.129446551	6.44525508	3.39E-13	3.65E-11
b0755	*gpmA*	phosphoglyceromutase 1	-3.127498311	6.396306195	3.08E-15	5.24E-13

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b0890	*ftsK*	DNA translocase at septal ring sorting daughter chromsomes	-3.06398268	4.716820648	8.23E-07	3.45E-05
b3021	*mqsA*	antitoxin for MqsR toxin; predicted transcriptional regulator	-3.021254016	5.537314633	6.33E-09	3.59E-07
b2297	*pta*	phosphate acetyltransferase	-2.879456965	4.567723118	8.43E-06	0.000263298
b0902	*pflA*	pyruvate formate lyase activating enzyme 1	-2.818384087	7.0526443	1.55E-10	1.18E-08
b0767	*pgl*	6-phosphogluconolactonase	-2.788918323	5.668575642	5.39E-08	2.76E-06
b0903	*pflB*	pyruvate formate lyase I	-2.760815255	8.522955142	5.49E-17	1.54E-14
b1859	*znuB*	zinc transporter subunit: membrane component of ABC superfamily	-2.746876857	5.312908221	3.08E-06	0.000113892
b3405	*ompR*	DNA-binding response regulator in two-component regulatory system with EnvZ	-2.669690831	5.732452948	4.09E-06	0.000141899
b1232	*purU*	formyltetrahydrofolate hydrolase	-2.65584935	6.426578955	6.21E-13	6.52E-11
b2425	*cysP*	thiosulfate-binding protein	-2.549598844	6.036650093	3.94E-06	0.000140363
b1858	*znuC*	zinc transporter subunit: ATP-binding component of ABC superfamily	-2.527866456	6.079349803	2.50E-09	1.59E-07
b3630	*rfaP*	kinase that phosphorylates core heptose of lipopolysaccharide	-2.503892757	5.110698147	3.32E-06	0.000120244
b3911	*cpxA*	sensory histidine kinase in two-component regulatory system with CpxR	-2.485264902	8.343866482	9.14E-20	2.95E-17
b0912	*ihfB*	integration host factor (IHF), DNA-binding protein, beta subunit	-2.462999374	5.340060329	2.30E-06	8.62E-05
b0077	*ilvI*	acetolactate synthase III, large subunit	-2.432712107	6.99191795	3.99E-14	5.08E-12
b2741	*rpoS*	RNA polymerase, sigma S (sigma 38) factor	-2.28832011	6.14884197	8.97E-08	4.32E-06
b3930	*menA*	1,4-dihydroxy-2-naphthoate octaprenyltransferase	-2.283979715	6.834113827	7.63E-06	0.000240993
b3163	*nlpI*	lipoprotein involved in osmotic sensitivity and filamentation	-2.247934928	5.6594604	7.30E-06	0.000232259
b3495	*uspA*	universal stress global response regulator	-2.239740795	7.41444384	3.39E-05	0.000922421
b1176	*minC*	cell division inhibitor	-2.213189634	6.941272965	7.64E-12	7.13E-10
b2188	*yejM*	predicted hydrolase, inner membrane	-2.135126353	6.848540511	5.22E-11	4.30E-09
b2290	*alaA*	valine-pyruvate aminotransferase 2	-2.068767117	7.942808659	2.28E-22	9.57E-20
b1109	*ndh*	respiratory NADH dehydrogenase 2/cupric reductase	-2.026132989	5.885073262	7.06E-06	0.000226398
b1175	*minD*	membrane ATPase of the MinC-MinD-MinE system	-1.950413117	7.425924333	5.27E-15	8.20E-13
b0121	*speE*	spermidine synthase (putrescine aminopropyltransferase)	-1.947111079	6.309573973	1.17E-05	0.000352366

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b0585	*fes*	enterobactin/ferric enterobactin esterase	-1.906095516	5.884663404	9.60E-06	0.000292238
b0120	*speD*	S-adenosylmethionine decarboxylase	-1.884043785	6.511625557	2.65E-05	0.00073231
b0080	*cra*	DNA-binding transcriptional repressor-activator for carbon metabolism	-1.870649324	6.472646678	6.43E-07	2.73E-05
b1593	*ynfK*	predicted dethiobiotin synthetase	-1.823466815	6.256317143	1.46E-05	0.000432505
b3933	*ftsN*	essential cell division protein	-1.741387412	6.10020507	2.96E-05	0.000810986
b0075	*leuL*	leu operon leader peptide	-1.690484345	6.42286185	2.25E-05	0.000628548
b0484	*copA*	copper transporter	-1.659125299	8.170393789	7.87E-18	2.36E-15
b1812	*pabB*	aminodeoxychorismate synthase, subunit I	-1.606313856	7.4868262	2.73E-07	1.21E-05
b4218	*ytfL*	inner membrane protein, UPF0053 family	-1.581293423	7.594920672	1.04E-11	9.25E-10
b2330	*prmB*	N5-glutamine methyltransferase	-1.537275109	6.875445146	1.30E-06	5.18E-05
b3432	*glgB*	1,4-alpha-glucan branching enzyme	-1.500788882	7.966667667	2.12E-05	0.000600143
b3252	*csrD*	targeting factor for csrBC sRNA degradation	-1.475911332	8.06721401	3.56E-14	4.82E-12
b1096	*pabC*	4-amino-4-deoxychorismate lyase component of para-aminobenzoate synthase multienzyme complex	-1.422266459	6.843289563	1.04E-06	4.31E-05
b3409	*feoB*	fused ferrous iron transporter, protein B: GTP-binding protein/membrane protein	-1.414812691	8.693810847	4.19E-14	5.17E-12
b0406	*tgt*	tRNA-guanine transglycosylase	-1.383602401	7.719513079	7.57E-08	3.74E-06
b1249	*cls*	cardiolipin synthase 1	-1.340272139	7.273139359	1.83E-06	7.04E-05
b3408	*feoA*	ferrous iron transporter, protein A	-1.276612527	7.432111792	1.94E-07	8.78E-06
b0819	*ybiS*	L,D-transpeptidase linking Lpp to murein	-1.112947598	7.89184208	1.28E-06	5.15E-05
b1661	*cfa*	cyclopropane fatty acyl phospholipid synthase (unsaturated-phospholipid methyltransferase)	-1.108982403	7.335002453	9.24E-06	0.000285359
b1676	*pykF*	pyruvate kinase I	-1.065559994	7.171812847	1.99E-05	0.000575367
b0919	*ycbJ*	conserved protein	-0.939535991	7.538559992	5.27E-05	0.001401194
b2144	*sanA*	vancomycin high temperature exclusion protein; mutants have a defective envelope more permeable to vancomycin at 42 degrees	-0.755495375	9.48761452	4.07E-06	0.000141899
b4141	*yjeH*	predicted transporter	0.667347796	8.847998102	2.07E-05	0.000595574
b3940	*metL*	fused aspartokinase II/homoserine dehydrogenase II	0.668284054	9.197728295	1.74E-05	0.000511619
b0830	*gsiB*	glutathione periplasmic binding protein, ABC superfamily transporter	0.736473597	9.167451285	3.40E-05	0.000922421

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff = 0.002188183)
b1243	*oppA*	oligopeptide transporter subunit	0.76357489	8.724166605	5.09E-06	0.00017372
b3780	*rhlB*	ATP-dependent RNA helicase	0.820019648	8.899124762	2.21E-05	0.000621677
b0381	*ddlA*	D-alanine-D-alanine ligase A	0.917493042	9.06825475	8.69E-08	4.24E-06
b1824	*yobF*	predicted protein	0.922173491	7.64787107	6.38E-05	0.001675392
b2786	*barA*	hybrid sensory histidine kinase, in two-component regulatory system with UvrY	1.014427146	11.38237041	3.78E-11	3.18E-09
b1914	*uvrY*	DNA-binding response regulator in two-component regulatory system with BarA	1.027145787	9.271608784	2.79E-09	1.70E-07
b2829	*ptsP*	fused PTS enzyme: PEP-protein phosphotransferase (enzyme I)/GAF domain containing protein	1.036221077	9.675391387	9.76E-11	7.59E-09
b2903	*gcvP*	glycine decarboxylase, PLP-dependent, subunit (protein P) of glycine cleavage complex	1.095457083	9.826901999	1.44E-13	1.63E-11
b0059	*rapA*	RNA polymerase-associated helicase protein (ATPase and RNA polymerase recycling factor)	1.292103347	10.42964037	1.38E-09	9.20E-08
b3181	*greA*	transcript cleavage factor	1.314013256	6.792704055	5.69E-05	0.001501875
b4260	*pepA*	multifunctional aminopeptidase A: a cyteinylglycinase, transcription regulator and site-specific recombination factor	1.382370758	8.826883534	1.95E-08	1.06E-06
b2947	*gshB*	glutathione synthetase	1.544847185	6.566575326	6.18E-06	0.00020435
b2688	*gshA*	glutamate-cysteine ligase	2.089686175	8.831550986	6.16E-20	2.21E-17
b3783	*rho*	transcription termination factor	2.456384738	5.474325994	2.11E-05	0.000600143
b1849	*purT*	phosphoribosylglycinamide formyltransferase 2	4.173653583	10.73466199	5.91E-54	1.24E-50

Table A13. -
Glucose nitrate

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b2784	*relA*	(p)ppGpp synthetase I/GTP pyrophosphokinase	-11.18475759	7.254643588	1.14E-25	2.25E-23
b3924	*fpr*	ferredoxin-NADP reductase	-10.45654815	6.535842049	5.83E-28	1.44E-25
b3008	*metC*	cystathionine beta-lyase, PLP-dependent	-9.912646016	6.003259678	6.76E-17	5.20E-15
b2839	*lysR*	DNA-binding transcriptional dual regulator	-9.670635283	5.76811679	7.07E-15	4.17E-13
b1101	*ptsG*	fused glucose-specific PTS enzymes: IIB component/IIC component	-9.63963832	5.738481548	1.72E-13	8.12E-12
b1262	*trpC*	fused indole-3-glycerolphosphate synthetase/N-(5-phosphoribosyl)anthranilate isomerase	-9.602195677	5.702138445	2.54E-14	1.37E-12
b0585	*fes*	enterobactin/ferric enterobactin esterase	-9.491201028	5.594780301	4.44E-14	2.33E-12
b0592	*fepB*	iron-enterobactin transporter subunit	-9.422965642	5.529153904	3.75E-13	1.64E-11
b0767	*pgl*	6-phosphogluconolactonase	-9.414776235	5.521117115	1.60E-13	7.60E-12
b0002	*thrA*	fused aspartokinase I and homoserine dehydrogenase I	-9.396992632	5.504069722	2.34E-13	1.07E-11
b4025	*pgi*	glucosephosphate isomerase	-9.21904103	5.333172597	5.84E-12	2.17E-10
b0438	*clpX*	ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease	-9.208207816	5.323071825	6.02E-12	2.21E-10
b0783	*moaC*	molybdopterin biosynthesis, protein C	-9.148150436	5.265126756	1.18E-09	2.76E-08
b3939	*metB*	cystathionine gamma-synthase, PLP-dependent	-9.147793484	5.264915496	1.56E-10	4.51E-09
b0588	*fepC*	iron-enterobactin transporter subunit	-9.083067014	5.203299897	8.76E-11	2.62E-09
b2751	*cysN*	sulfate adenylyltransferase, subunit 1	-9.082216561	5.202785392	1.58E-11	5.39E-10
b3035	*tolC*	transport channel	-9.071023435	5.192187237	2.53E-11	8.34E-10
b0928	*aspC*	aspartate aminotransferase, PLP-dependent	-9.049933273	5.171690199	3.68E-10	9.94E-09
b0750	*nadA*	quinolinate synthase, subunit A	-9.037511437	5.16028208	3.09E-11	1.01E-09
b0071	*leuD*	3-isopropylmalate dehydratase small subunit	-9.026519225	5.149701062	3.38E-11	1.09E-09
b0590	*fepD*	iron-enterobactin transporter subunit	-8.980120685	5.105644308	6.10E-11	1.89E-09
b0073	*leuB*	3-isopropylmalate dehydrogenase, NAD(+)-dependent	-8.968392597	5.09448148	9.33E-11	2.77E-09
b3772	*ilvA*	threonine deaminase	-8.967720159	5.094085644	8.19E-11	2.47E-09
b2752	*cysD*	sulfate adenylyltransferase, subunit 2	-8.967082977	5.093718068	9.25E-10	2.23E-08
b1062	*pyrC*	dihydro-orotase	-8.93426283	5.061353831	2.73E-06	3.17E-05
b0072	*leuC*	3-isopropylmalate dehydratase large subunit	-8.894784201	5.024961165	1.35E-10	3.93E-09

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3368	*cysG*	fused siroheme synthase 1,3-dimethyluroporphyriongen III dehydrogenase and siroheme ferrochelatase/uroporphyrinogen methyltransferase	-8.83018226	4.964236353	8.32E-10	2.03E-08
b4013	*metA*	homoserine O-transsuccinylase	-8.792083736	4.927708012	7.79E-09	1.49E-07
b2422	*cysA*	sulfate/thiosulfate transporter subunit	-8.79103979	4.927135512	3.99E-10	1.07E-08
b2423	*cysW*	sulfate/thiosulfate ABC transporter subunit	-8.790878123	4.927045116	4.26E-10	1.11E-08
b0630	*lipB*	octanoyltransferase; octanoyl-[ACP]:protein N-octanoyltransferase	-8.736200802	4.8756199	7.30E-10	1.82E-08
b0889	*lrp*	DNA-binding transcriptional dual regulator, leucine-binding	-8.736033114	4.875527776	7.87E-10	1.94E-08
b0033	*carB*	carbamoyl-phosphate synthase large subunit	-8.72322437	4.863021278	1.65E-08	2.94E-07
b2424	*cysU*	sulfate/thiosulfate ABC transporter permease	-8.707979304	4.849160581	1.11E-09	2.62E-08
b0740	*tolB*	periplasmic protein	-8.695361314	4.83663592	3.66E-07	5.16E-06
b1263	*trpD*	fused glutamine amidotransferase (component II) of anthranilate synthase/anthranilate phosphoribosyl transferase	-8.693109466	4.835452072	2.68E-09	5.67E-08
b0945	*pyrD*	dihydro-orotate oxidase, FMN-linked	-8.679823805	7.024976372	9.07E-12	3.22E-10
b3958	*argC*	N-acetyl-gamma-glutamylphosphate reductase, NAD(P)-binding	-8.62070059	4.767210325	1.60E-08	2.87E-07
b0589	*fepG*	iron-enterobactin transporter subunit	-8.604279694	4.752388176	3.50E-09	7.28E-08
b3956	*ppc*	phosphoenolpyruvate carboxylase	-8.589287214	4.738258151	3.21E-09	6.72E-08
b3781	*trxA*	thioredoxin 1	-8.556589594	4.708433021	1.33E-07	2.05E-06
b0109	*nadC*	quinolinate phosphoribosyltransferase	-8.509615651	4.664281448	6.61E-09	1.30E-07
b0957	*ompA*	outer membrane protein A (3a;II*;G;d)	-8.475732163	4.633237475	4.34E-08	7.33E-07
b3237	*argR*	DNA-binding transcriptional dual regulator, L-arginine-binding	-8.373960072	4.53832441	5.26E-07	7.25E-06
b0437	*clpP*	proteolytic subunit of ClpA-ClpP and ClpX-ClpP ATP-dependent serine proteases	-8.31884375	4.48776082	4.25E-07	5.96E-06
b3771	*ilvD*	dihydroxyacid dehydratase	-8.317784289	4.487270104	3.27E-08	5.63E-07
b3172	*argG*	argininosuccinate synthetase	-8.299862944	4.470450064	2.65E-07	3.88E-06
b0739	*tolA*	membrane anchored protein in TolA-TolQ-TolR complex	-8.261459988	4.435344212	2.86E-07	4.12E-06
b2297	*pta*	phosphate acetyltransferase	-8.260756119	4.435023535	5.87E-08	9.76E-07

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3959	*argB*	acetylglutamate kinase	-8.180388911	4.362014242	8.62E-08	1.39E-06
b3290	*trkA*	NAD-binding component of TrK potassium transporter	-8.095792398	4.28538605	1.57E-07	2.40E-06
b2500	*purN*	phosphoribosylglycinamide formyltransferase 1	-8.050847494	4.245202963	7.12E-07	9.54E-06
b1264	*trpE*	component I of anthranilate synthase	-8.006014314	4.204538244	2.98E-07	4.27E-06
b1677	*lpp*	murein lipoprotein	-7.935354972	4.141035257	6.98E-07	9.38E-06
b3774	*ilvC*	ketol-acid reductoisomerase, NAD(P)-binding	-7.910658041	4.119090786	6.95E-07	9.37E-06
b0003	*thrB*	homoserine kinase	-7.860159357	4.074266865	8.62E-07	1.12E-05
b3828	*metR*	DNA-binding transcriptional activator, homocysteine-binding	-7.859765268	4.074108698	7.76E-07	1.03E-05
b3204	*ptsN*	sugar-specific enzyme IIA component of PTS	-7.859666746	4.074069157	8.06E-07	1.06E-05
b0243	*proA*	gamma-glutamylphosphate reductase	-7.859567856	4.074029616	8.80E-07	1.14E-05
b3770	*ilvE*	branched-chain amino-acid aminotransferase	-7.807320408	4.027821279	1.38E-06	1.70E-05
b0775	*bioB*	biotin synthase	-7.66310283	6.037447966	6.82E-17	5.20E-15
b0738	*tolR*	membrane spanning protein in TolA-TolQ-TolR complex	-7.610179423	3.854248181	8.83E-05	0.000701023
b2557	*purL*	phosphoribosylformyl-glycineamide synthetase	-7.608458977	3.853662571	3.81E-06	4.25E-05
b4048	*yjbM*	predicted protein	-7.578503923	3.827278731	1.52E-05	0.000144069
b3941	*metF*	5,10-methylenetetrahydrofolate reductase	-7.546668849	3.799977038	1.05E-05	0.00010453
b3960	*argH*	argininosuccinate lyase	-7.546668645	3.799977038	1.03E-05	0.000104046
b2699	*recA*	DNA strand exchange and recombination protein with protease and nuclease activity	-7.545699344	3.799638078	5.49E-06	5.89E-05
b1850	*eda*	multifunctional 2-keto-3-deoxygluconate 6-phosphate aldolase and 2-keto-4-hydroxyglutarate aldolase and oxaloacetate decarboxylase	-7.545089129	3.799426241	1.04E-05	0.00010421
b4214	*cysQ*	PAPS (adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase	-7.512234945	3.771517165	3.91E-05	0.00034303
b0688	*pgm*	phosphoglucomutase	-7.445712578	3.714533784	3.38E-05	0.000298502
b0052	*pdxA*	4-hydroxy-L-threonine phosphate dehydrogenase, NAD-dependent	-7.445573244	3.71449058	4.45E-05	0.000384048
b0741	*pal*	peptidoglycan-associated outer membrane lipoprotein	-7.41375577	3.685987463	6.52E-05	0.000539595
b2313	*cvpA*	membrane protein required for colicin V production	-7.341517844	3.625193627	3.52E-05	0.000310005
b2894	*xerD*	site-specific tyrosine recombinase	-7.340956705	3.625017292	1.89E-05	0.000176525
b4687	*shoB*	toxic membrane protein	-7.304080064	3.593844842	3.18E-05	0.000281989

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b0386	*proC*	pyrroline-5-carboxylate reductase, NAD(P)-binding	-7.265944304	3.561898506	4.24E-05	0.000369247
b3738	*atpB*	F0 sector of membrane-bound ATP synthase, subunit a	-7.225289444	3.528785816	4.13E-05	0.000360204
b0908	*aroA*	5-enolpyruvylshikimate-3-phosphate synthetase	-7.184514582	3.4952347	8.19E-05	0.000659255
b0564	*appY*	DNA-binding global transcriptional activator; DLP12 prophage	-7.144448132	3.461444497	5.48E-05	0.00046385
b1693	*aroD*	3-dehydroquinate dehydratase	-7.144292547	3.461398882	5.00E-05	0.000428945
b2020	*hisD*	bifunctional histidinal dehydrogenase/ histidinol dehydrogenase	-7.143362548	3.461125207	6.34E-05	0.000526384
b0004	*thrC*	threonine synthase	-7.101145513	3.426240651	5.88E-05	0.000492458
b3613	*envC*	activator of AmiB,C murein hydrolases, septal ring factor	-6.929164836	6.662600734	1.18E-25	2.25E-23
b0785	*moaE*	molybdopterin synthase, large subunit	-6.868864849	5.294768666	2.19E-10	6.15E-09
b4245	*pyrB*	aspartate carbamoyltransferase, catalytic subunit	-6.310573528	4.782123339	9.65E-06	9.80E-05
b4171	*miaA*	delta(2)-isopentenylpyrophosphate tRNA-adenosine transferase	-6.191999724	4.66455824	6.83E-08	1.11E-06
b2425	*cysP*	thiosulfate-binding protein	-6.075749121	5.877368636	4.16E-13	1.78E-11
b0049	*apaH*	diadenosine tetraphosphatase	-5.776059662	5.083765539	1.39E-08	2.54E-07
b0131	*panD*	aspartate 1-decarboxylase	-5.725710117	4.245781276	2.40E-06	2.82E-05
b3911	*cpxA*	sensory histidine kinase in two-component regulatory system with CpxR	-5.679446541	8.185044155	7.31E-43	3.83E-40
b0544	*ybcK*	DLP12 prophage; predicted recombinase	-5.384593648	5.612138489	7.05E-09	1.38E-07
b3957	*argE*	acetylornithine deacetylase	-5.33844691	5.1937161	1.77E-07	2.67E-06
b1252	*tonB*	membrane spanning protein in TonB-ExbB-ExbD transport complex	-5.249205015	4.603002151	8.82E-07	1.14E-05
b3631	*rfaG*	glucosyltransferase I	-5.207662804	6.617215812	1.32E-20	1.63E-18
b1661	*cfa*	cyclopropane fatty acyl phospholipid synthase (unsaturated-phospholipid methyltransferase)	-5.120147891	6.873769264	1.20E-20	1.52E-18
b1552	*cspI*	Qin prophage; cold shock protein	-5.106824157	3.715194939	8.53E-05	0.000683949
b2846	*yqeH*	conserved protein with bipartite regulator domain	-5.046910752	5.200029254	9.45E-06	9.64E-05
b4042	*dgkA*	diacylglycerol kinase	-5.03050752	5.19864747	7.52E-08	1.22E-06
b0074	*leuA*	2-isopropylmalate synthase	-5.008302466	5.705556051	2.11E-09	4.59E-08
b1254	*yciB*	predicted inner membrane protein	-4.93613468	5.203317049	1.94E-06	2.31E-05
b3671	*ilvB*	acetolactate synthase I, large subunit	-4.918726966	7.807636977	7.73E-49	9.83E-46

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3829	*metE*	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	-4.755211033	4.116101939	2.73E-05	0.000244664
b4660	*yhiL*	pseudogene	-4.70815705	6.21143162	1.05E-08	1.96E-07
b1334	*fnr*	DNA-binding transcriptional dual regulator, global regulator of anaerobic growth	-4.689498906	6.338091013	9.14E-12	3.22E-10
b0484	*copA*	copper transporter	-4.659683484	7.878410981	1.59E-53	3.34E-50
b4325	*yjiC*	predicted protein	-4.600451439	5.358780314	1.31E-07	2.02E-06
b3818	*yigG*	conserved inner membrane protein	-4.549768091	5.269219778	2.61E-06	3.05E-05
b2818	*argA*	fused acetylglutamate kinase homolog (inactive)/amino acid N-acetyltransferase	-4.517672851	5.23896739	4.01E-06	4.44E-05
b2273	*yfbN*	conserved protein	-4.460534429	5.556780736	1.36E-05	0.000132159
b4612	*yrhD*	hypothetical protein	-4.459228684	5.616089477	2.66E-07	3.88E-06
b0009	*mog*	molybdochelatase incorporating molybdenum into molybdopterin	-4.296341295	5.990815484	1.14E-09	2.67E-08
b1058	*yceO*	predicted protein	-4.296018359	4.532586757	1.26E-05	0.000123066
b1166	*ariR*	connector protein for RcsB regulation of biofilm and acid-resistance	-4.295642531	5.280072789	4.33E-08	7.33E-07
b3773	*ilvY*	DNA-binding transcriptional dual regulator	-4.241291777	5.363478053	4.33E-06	4.73E-05
b3860	*dsbA*	periplasmic protein disulfide isomerase I	-4.179913579	7.163161338	9.70E-18	9.24E-16
b0774	*bioA*	7,8-diaminopelargonic acid synthase, PLP-dependent	-4.17172153	5.966646137	1.16E-08	2.13E-07
b2318	*truA*	pseudouridylate synthase I	-4.113506807	5.467185264	1.47E-06	1.79E-05
b0826	*moeB*	molybdopterin synthase sulfurylase	-4.050137384	5.723106101	8.38E-06	8.64E-05
b0720	*gltA*	citrate synthase	-4.048902424	4.941678373	5.52E-05	0.000465796
b2188	*yejM*	predicted hydrolase, inner membrane	-4.019626729	6.686681156	1.29E-17	1.13E-15
b1593	*ynfK*	predicted dethiobiotin synthetase	-4.016846444	6.032348595	9.77E-09	1.84E-07
b3839	*tatC*	TatABCE protein translocation system subunit	-3.905256958	6.60814834	4.79E-07	6.65E-06
b3459	*yhhK*	pantothenate synthesis protein, predicted acetyltransferase	-3.827324457	6.036080054	7.43E-09	1.44E-07
b0827	*moeA*	molybdopterin biosynthesis protein	-3.780127596	6.046797539	4.17E-10	1.09E-08
b4279	*yjhB*	predicted transporter	-3.690031273	5.873137474	8.55E-05	0.000683949
b0777	*bioC*	malonyl-CoA methyltransferase, SAM-dependent	-3.648782492	5.924274414	3.04E-05	0.000270638
b2108	*yehA*	predicted fimbrial-like adhesin protein	-3.564876557	5.824385689	3.65E-06	4.11E-05

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b0051	*rsmA*	16S rRNA dimethyladenosine transferase, SAM-dependent	-3.446132983	5.824748133	1.04E-05	0.00010421
b0217	*yafT*	lipoprotein	-3.390244719	6.645618851	1.57E-09	3.47E-08
b0764	*modB*	molybdate transporter subunit	-3.231715142	6.684980547	4.10E-13	1.77E-11
b3442	*yhhZ*	conserved protein	-3.178486436	6.359213251	3.29E-08	5.65E-07
b1505	*ydeT*	pseudogene	-3.171650493	8.014320539	3.75E-33	1.12E-30
b4495	*yedN*	pseudogene	-3.158735598	5.194675892	5.11E-05	0.000437318
b0781	*moaA*	molybdopterin biosynthesis protein A	-3.132984223	6.718503272	1.89E-13	8.82E-12
b2852	*ygeH*	predictedtranscriptional regulator	-3.122891167	5.73991846	1.21E-05	0.000118541
b0134	*panB*	3-methyl-2-oxobutanoate hydroxymethyltransferase	-3.115749023	6.186132502	2.54E-07	3.76E-06
b0279	*yagM*	CP4-6 prophage; predicted protein	-3.11002534	6.989086844	1.58E-10	4.53E-09
b2144	*sanA*	vancomycin high temperature exclusion protein; mutants have a defective envelope more permeable to vancomycin at 42 degrees	-3.108893411	9.022742707	2.60E-43	1.56E-40
b2352	*gtrS*	serotype-specific glucosyl transferase, CPS-53 (KpLE1) prophage	-3.096695416	6.103931847	1.07E-07	1.68E-06
b1499	*ydeO*	transcriptional activator for mdtEF	-3.090450058	5.938958689	1.27E-06	1.57E-05
b2642	*yfjW*	CP4-57 prophage; predicted inner membrane protein	-2.949161514	6.029586683	2.68E-05	0.000241362
b1761	*gdhA*	glutamate dehydrogenase, NADP-specific	-2.939411151	6.881156667	1.23E-09	2.83E-08
b1812	*pabB*	aminodeoxychorismate synthase, subunit I	-2.899488645	7.308124752	2.11E-16	1.38E-14
b1182	*hlyE*	hemolysin E	-2.898316474	6.83724966	1.02E-09	2.45E-08
b3493	*pitA*	phosphate transporter, low-affinity	-2.847202379	7.118895954	4.02E-10	1.07E-08
b1096	*pabC*	4-amino-4-deoxychorismate lyase component of para-aminobenzoate synthase multienzyme complex	-2.823669013	6.625884328	2.42E-10	6.75E-09
b1497	*ydeM*	conserved protein	-2.792415885	7.620829213	5.11E-09	1.03E-07
b3867	*hemN*	coproporphyrinogen III oxidase, SAM and NAD(P)H dependent, oxygen-independent	-2.772649622	7.320490407	7.79E-17	5.83E-15
b2965	*speC*	ornithine decarboxylase, constitutive	-2.764549426	7.835123759	3.65E-26	8.05E-24
b1139	*lit*	e14 prophage; cell death peptidase, inhibitor of T4 late gene expression	-2.747029306	7.288666605	1.15E-08	2.12E-07
b4650	*yibS*	pseudogene	-2.740379415	8.141389656	5.59E-23	8.69E-21
b4590	*ybfK*	hypothetical protein	-2.738954742	6.021522415	1.85E-05	0.000172885

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3047	*yqiH*	predicted periplasmic pilin chaperone	-2.721203095	7.51427333	1.48E-17	1.27E-15
b0310	*ykgH*	predicted inner membrane protein	-2.710012905	5.660833035	7.62E-05	0.000618991
b4280	*yjhC*	predicted oxidoreductase	-2.697726427	7.825930096	1.21E-17	1.08E-15
b2367	*emrY*	predicted multidrug efflux system	-2.695545885	8.451580136	1.14E-23	1.91E-21
b0234	*yafP*	predicted acyltransferase with acyl-CoA N-acyltransferase domain	-2.692376721	6.884492973	9.04E-12	3.22E-10
b2071	*yegJ*	predicted protein	-2.655591343	5.585906562	2.43E-05	0.000222659
b3889	*yiiE*	predicted transcriptional regulator	-2.632115743	6.365017614	1.14E-06	1.42E-05
b0603	*ybdO*	predicted DNA-binding transcriptional regulator	-2.609921165	7.433615581	4.50E-14	2.33E-12
b3359	*argD*	bifunctional acetylornithine aminotransferase/ succinyldiaminopimelate aminotransferase	-2.593379632	7.414874656	2.17E-07	3.22E-06
b4017	*arpA*	ankyrin repeat protein	-2.504068446	8.175078784	6.27E-23	9.39E-21
b3629	*rfaS*	lipopolysaccharide core biosynthesis protein	-2.457290703	6.701600902	9.20E-08	1.48E-06
b0531	*sfmC*	pilin chaperone, periplasmic	-2.419189101	7.563582604	8.29E-09	1.58E-07
b1235	*rssB*	response regulator binding RpoS to initiate proteolysis by ClpXP; required for the PcnB-degradosome interaction during stationary phase	-2.345036958	6.917576351	1.38E-09	3.13E-08
b1983	*yeeN*	conserved protein, UPF0082 family	-2.344281309	8.098086088	1.43E-16	9.80E-15
b4462	*ygaQ*	pseudogene	-2.341587209	7.759435043	1.11E-05	0.000110036
b4345	*mcrC*	5-methylcytosine-specific restriction enzyme McrBC, subunit McrC	-2.27317908	6.959039301	1.80E-09	3.96E-08
b2625	*yfjI*	CP4-57 prophage; predicted protein	-2.271681672	6.457212816	8.05E-06	8.33E-05
b3912	*cpxR*	DNA-binding response regulator in two-component regulatory system with CpxA	-2.255416612	7.468130251	4.18E-12	1.58E-10
b0278	*yagL*	CP4-6 prophage; DNA-binding protein	-2.246747698	8.04820258	1.07E-17	9.72E-16
b3623	*waaU*	lipopolysaccharide core biosynthesis	-2.241404962	7.496107025	6.18E-12	2.25E-10
b4299	*yjhI*	KpLE2 phage-like element; predicted DNA-binding transcriptional regulator	-2.228221523	8.532349322	8.35E-19	8.75E-17
b4346	*mcrB*	5-methylcytosine-specific restriction enzyme McrBC, subunit McrB	-2.220762931	9.0885033	4.13E-21	5.41E-19
b3247	*rng*	ribonuclease G	-2.214975896	6.321363805	1.30E-05	0.000126649
b3048	*yqiI*	conserved protein	-2.213540138	8.741132135	7.27E-24	1.27E-21

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3264	*envR*	DNA-binding transcriptional regulator	-2.210659836	7.715588186	1.82E-17	1.49E-15
b0136	*yadK*	predicted fimbrial-like adhesin protein	-2.203870606	7.360231932	1.85E-07	2.78E-06
b1951	*rcsA*	DNA-binding transcriptional activator, co-regulator with RcsB	-2.199719577	7.171825659	2.24E-08	3.96E-07
b0765	*modC*	molybdate transporter subunit	-2.185426971	6.131699784	7.83E-05	0.000633643
b2669	*stpA*	DNA binding protein, nucleoid-associated	-2.180592355	8.530910514	4.31E-18	4.30E-16
b1409	*ynbB*	predicted CDP-diglyceride synthase	-2.154851088	6.525066347	2.90E-06	3.34E-05
b1690	*ydiM*	inner membrane protein, predicted transporter	-2.153402782	9.209800553	3.29E-13	1.45E-11
b2290	*alaA*	valine-pyruvate aminotransferase 2	-2.131180278	7.977959437	1.71E-19	1.94E-17
b1502	*ydeQ*	predicted fimbrial-like adhesin protein	-2.126425831	7.100403334	4.34E-07	6.07E-06
b1258	*yciF*	predicted rubrerythrin/ferritin-like metal-binding protein	-2.124843139	6.722438526	4.24E-06	4.64E-05
b1168	*ycgG*	predicted cyclic-di-GMP phosphodiesterase	-2.10127729	8.499801494	5.68E-20	6.62E-18
b1221	*narL*	DNA-binding response regulator in two-component regulatory system with NarX (or NarQ)	-2.100854126	6.699001692	7.27E-07	9.67E-06
b1121	*ycfZ*	inner membrane protein	-2.054295804	7.688044496	5.85E-10	1.49E-08
b2774	*ygcW*	predicted dehydrogenase	-2.047109875	8.01204064	2.56E-15	1.56E-13
b2271	*yfbL*	predicted peptidase	-2.044149113	8.363171081	5.11E-20	6.12E-18
b1557	*cspB*	Qin prophage; cold shock protein	-2.038929572	6.848244295	6.16E-06	6.54E-05
b3504	*yhiS*	pseudogene	-2.026175613	7.262925171	3.22E-08	5.59E-07
b1223	*narK*	nitrate/nitrite transporter	-2.014875034	7.220459205	1.04E-07	1.64E-06
b1040	*csgD*	DNA-binding transcriptional activator for csgBA	-1.997938555	7.076901267	1.98E-07	2.94E-06
b1496	*yddA*	fused predicted multidrug transporter subunits of ABC superfamily: membrane component/ATP-binding component	-1.997279562	9.591210745	4.66E-19	5.14E-17
b0364	*yaiS*	conserved protein	-1.993152744	8.009759369	2.15E-15	1.32E-13
b0326	*yahL*	predicted protein	-1.986788104	7.3591027	1.19E-07	1.85E-06
b1041	*csgB*	curlin nucleator protein, minor subunit in curli complex	-1.965426362	6.890815571	9.47E-08	1.51E-06
b3723	*bglG*	transcriptional antiterminator of the bgl operon	-1.961329397	8.919459564	1.72E-16	1.16E-14
b0728	*sucC*	succinyl-CoA synthetase, beta subunit	-1.957211603	6.896538472	5.51E-06	5.89E-05
b1175	*minD*	membrane ATPase of the MinC-MinD-MinE system	-1.956291511	7.47339284	2.34E-08	4.12E-07
b0508	*hyi*	hydroxypyruvate isomerase	-1.952523701	7.294702604	4.93E-09	1.00E-07
b3429	*glgA*	glycogen synthase	-1.950481928	6.662585999	2.26E-05	0.000208334

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3412	*bioH*	pimeloyl-ACP carboxylesterase	-1.935785636	6.36861946	2.55E-05	0.000232023
b1735	*chbR*	rRepressor, chb operon for N,N'-diacetylchitobiose utilization	-1.93193687	7.732413499	7.28E-13	2.99E-11
b0121	*speE*	spermidine synthase (putrescine aminopropyltransferase)	-1.922663981	6.359179846	3.92E-05	0.00034303
b3120	*yhaB*	predicted protein	-1.917875206	6.626792799	1.93E-05	0.000179744
b0080	*cra*	DNA-binding transcriptional repressor-activator for carbon metabolism	-1.914291307	6.508000228	4.51E-05	0.00038844
b2778	*ygcG*	predicted protein	-1.883576494	7.870089074	8.31E-09	1.58E-07
b1576	*ydfD*	Qin prophage; predicted protein	-1.875526842	8.403793266	5.73E-12	2.14E-10
b1320	*ycjW*	predicted DNA-binding transcriptional regulator	-1.874752286	8.282623682	9.88E-15	5.67E-13
b0763	*modA*	molybdate transporter subunit	-1.846539266	7.226477005	3.46E-07	4.91E-06
b4455	*hokA*	toxic polypeptide, small	-1.841232879	6.957769124	2.95E-05	0.000263868
b0546	*ybcM*	DLP12 prophage; predicted DNA-binding transcriptional regulator	-1.834317234	7.877714702	7.19E-13	2.98E-11
b4494	*arpB*	pseudogene	-1.826393002	9.558497085	5.36E-23	8.64E-21
b0209	*yafD*	conserved protein	-1.816067621	6.784423935	1.85E-06	2.23E-05
b0729	*sucD*	succinyl-CoA synthetase, NAD(P)-binding, alpha subunit	-1.79385656	7.098269464	5.77E-07	7.91E-06
b2372	*yfdV*	predicted transporter	-1.770218165	8.310869789	2.33E-14	1.27E-12
b0779	*uvrB*	excinulease of nucleotide excision repair, DNA damage recognition component	-1.765984847	7.695881102	3.49E-11	1.12E-09
b4640	*yoeG*	pseudogene	-1.762283723	7.618428693	1.44E-07	2.20E-06
b2623	*yfjH*	CP4-57 prophage; predicted protein	-1.751624894	7.821489932	7.59E-09	1.46E-07
b2055	*wcaE*	predicted glycosyl transferase	-1.744489276	7.306465959	6.15E-08	1.02E-06
b0833	*yliE*	predicted cyclic-di-GMP phosphodiesterase, inner membrane protein	-1.741857018	9.470829396	6.49E-19	6.98E-17
b2371	*yfdE*	predicted CoA-transferase, NAD(P)-binding	-1.737370786	8.430965533	7.94E-17	5.84E-15
b0938	*elfA*	predicted fimbrial-like adhesin protein	-1.716988505	8.247639677	6.45E-11	1.99E-09
b2847	*yqeI*	predicted transcriptional regulator	-1.707265921	6.947211892	8.98E-07	1.15E-05
b1159	*mcrA*	e14 prophage; 5-methylcytosine-specific restriction endonuclease B	-1.697114537	7.456191535	8.63E-09	1.63E-07
b1968	*yedV*	predicted sensory kinase in two-component regulatory system with YedW	-1.678737769	8.927240451	1.88E-15	1.18E-13

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b0077	*ilvI*	acetolactate synthase III, large subunit	-1.678263332	7.183238436	3.80E-06	4.25E-05
b1498	*ydeN*	conserved protein	-1.676482799	9.292484297	1.04E-16	7.41E-15
b1234	*rssA*	predicted phospholipase, patatin-like family	-1.665977923	7.92702817	1.69E-12	6.67E-11
b3043	*ygiL*	predicted fimbrial-like adhesin protein	-1.661903295	8.416691665	1.16E-13	5.72E-12
b0691	*ybfG*	pseudogene	-1.659413614	8.733273849	5.26E-15	3.15E-13
b4311	*nanC*	N-acetylnuraminic acid outer membrane channel protein	-1.656151719	9.015417078	1.46E-13	7.06E-12
b4647	*mokA*	pseudogene	-1.655038892	7.126256859	8.24E-07	1.08E-05
b1773	*ydjI*	predicted aldolase	-1.626928334	7.541369096	2.67E-07	3.89E-06
b1566	*flxA*	Qin prophage; predicted protein	-1.616586966	7.068379019	5.46E-05	0.000463116
b0535	*fimZ*	predicted DNA-binding transcriptional regulator	-1.615376255	7.37437355	9.42E-08	1.51E-06
b0138	*yadM*	predicted fimbrial-like adhesin protein	-1.609503165	8.403383412	2.72E-13	1.22E-11
b2624	*alpA*	CP4-57 prophage; DNA-binding transcriptional activator	-1.606152573	6.938208183	7.45E-06	7.79E-05
b0406	*tgt*	tRNA-guanine transglycosylase	-1.593833141	7.711098361	2.65E-09	5.63E-08
b1504	*ydeS*	predicted fimbrial-like adhesin protein	-1.570732236	7.288674273	1.66E-06	2.00E-05
b4181	*yjfI*	conserved protein, DUF2170 family	-1.567506102	6.843860283	5.73E-05	0.000480348
b1694	*ydiF*	fused predicted acetyl-CoA:acetoacetyl-CoA transferase: alpha subunit/beta subunit	-1.554920615	9.142611051	1.23E-16	8.59E-15
b2064	*asmA*	predicted assembly protein	-1.553949009	7.634299666	2.60E-07	3.83E-06
b4559	*yjdO*	predicted protein	-1.549364993	6.977823246	1.01E-05	0.000101396
b0137	*yadL*	predicted fimbrial-like adhesin protein	-1.54762863	9.427734718	6.13E-17	4.85E-15
b4497	*yeeL*	pseudogene	-1.53431686	8.120517286	3.38E-09	7.06E-08
b3146	*rsmI*	16S rRNA C1402 ribose 2'-O-methyltransferase, SAM-dependent	-1.506741368	6.979559232	1.48E-05	0.000141319
b3215	*yhcA*	predicted periplasmic chaperone protein	-1.502290477	8.183209634	2.70E-10	7.44E-09
b2163	*yeiL*	DNA-binding transcriptional activator of stationary phase nitrogen survival	-1.497111743	8.493101262	7.53E-12	2.72E-10
b1022	*pgaC*	predicted glycosyl transferase	-1.49594091	9.648478673	1.11E-13	5.56E-12
b3624	*rfaZ*	lipopolysaccharide core biosynthesis protein	-1.490084635	8.058133844	4.43E-05	0.000384048
b0769	*ybhH*	conserved protein	-1.48764128	8.634543684	1.94E-11	6.50E-10
b3875	*ompL*	outer membrane porin L	-1.481691363	8.440403289	3.73E-10	1.00E-08
b2760	*casA*	CRISP RNA (crRNA) containing Cascade antiviral complex protein	-1.479923653	9.899564229	9.80E-17	7.08E-15

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b4238	*nrdD*	anaerobic ribonucleoside-triphosphate reductase	-1.467997195	7.890540221	8.16E-10	2.00E-08
b2626	*yfjJ*	CP4-57 prophage; predicted protein	-1.456309029	9.386531603	1.27E-15	8.05E-14
b3622	*rfaL*	O-antigen ligase	-1.448909429	8.821701205	2.45E-10	6.81E-09
b3408	*feoA*	ferrous iron transporter, protein A	-1.436996811	7.435714081	3.73E-06	4.19E-05
b0499	*ylbH*	pseudogene	-1.435545112	8.175045202	5.59E-09	1.12E-07
b1691	*ydiN*	Inner membrane protein, predicted MFS superfamily transporter	-1.416461155	9.213372232	1.11E-09	2.62E-08
b0239	*frsA*	fermentation-respiration switch protein; PTS Enzyme IIA(Glc)-binding protein; pNP-butyrate esterase activity	-1.405042162	8.086652354	1.37E-09	3.11E-08
b1545	*pinQ*	Qin prophage; predicted site-specific recombinase	-1.397147706	6.954156417	8.56E-05	0.00068404
b0141	*yadN*	predicted fimbrial-like adhesin protein	-1.389994835	8.302601558	6.04E-10	1.52E-08
b4348	*hsdS*	specificity determinant for hsdM and hsdR	-1.374709081	9.623595133	1.65E-12	6.61E-11
b1023	*pgaB*	biofilm adhesin polysaccharide PGA export lipoprotein with a polysaccharide deacetylase activity needed for export	-1.365542858	9.582914773	1.21E-11	4.18E-10
b4116	*adiY*	DNA-binding transcriptional activator	-1.359157609	9.085407446	4.41E-13	1.87E-11
b1029	*ycdU*	predicted inner membrane protein	-1.354848137	8.825995501	2.87E-06	3.32E-05
b1615	*uidC*	predicted outer membrane porin protein	-1.351428822	8.301405653	5.56E-10	1.44E-08
b0990	*cspG*	cold shock protein homolog, cold-inducible	-1.344737259	7.158282952	3.03E-05	0.000270391
b2368	*emrK*	EmrKY-TolC multidrug resistance efflux pump, membrane fusion protein component	-1.329349235	8.758211831	2.20E-09	4.76E-08
b0511	*ybbW*	predicted allantoin transporter	-1.327670694	9.04118319	1.90E-10	5.38E-09
b3138	*agaB*	N-acetylgalactosamine-specific enzyme IIB component of PTS	-1.312271841	7.740233063	1.06E-06	1.34E-05
b1044	*ymdA*	predicted protein	-1.303592628	8.217456401	4.22E-09	8.68E-08
b2037	*rfbX*	predicted polisoprenol-linked O-antigen transporter	-1.303255088	8.691135111	7.02E-06	7.37E-05
b0241	*phoE*	outer membrane phosphoporin protein E	-1.299962478	9.509018164	1.69E-11	5.70E-10
b4636	*ybfI*	pseudogene	-1.299607232	7.372273815	2.61E-05	0.000236031
b1772	*ydjH*	predicted kinase	-1.288685389	8.048429911	1.08E-07	1.68E-06
b2360	*yfdQ*	CPS-53 (KpLE1) prophage; predicted protein	-1.276423754	8.771522386	2.90E-12	1.13E-10
b1365	*ynaK*	Rac prophage; conserved protein	-1.274312857	9.006253453	1.07E-07	1.68E-06
b3563	*yiaB*	inner membrane protein, YiaAB family	-1.269983615	7.454389581	9.63E-06	9.80E-05
b4313	*fimE*	tyrosine recombinase/inversion of on/off regulator of fimA	-1.264425156	9.02544183	1.09E-11	3.80E-10

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b0304	*ykgC*	predicted pyridine nucleotide-disulfide oxidoreductase	-1.253071424	10.00427182	3.27E-12	1.26E-10
b0545	*ybcL*	DLP12 prophage; secreted protein, UPF0098 family	-1.252785124	8.567049977	5.57E-08	9.31E-07
b3142	*yraH*	predicted fimbrial-like adhesin protein	-1.25109417	8.432294271	1.02E-08	1.91E-07
b1407	*ydbD*	predicted PF10971 family periplasmic methylglyoxal resistance protein	-1.247603987	9.779154932	5.58E-10	1.44E-08
b3252	*csrD*	targeting factor for csrBC sRNA degradation	-1.241348193	8.180780772	5.98E-09	1.19E-07
b3659	*setC*	predicted sugar efflux system	-1.230042901	9.412666653	3.19E-11	1.04E-09
b4184	*yjfL*	inner membrane protein, UPF0719 family	-1.213114281	8.360725023	2.63E-09	5.62E-08
b1785	*yeaI*	predicted diguanylate cyclase	-1.209623953	7.769312419	1.24E-05	0.000121215
b1915	*yecF*	conserved protein, DUF2594 family	-1.205478396	7.640074176	2.93E-06	3.36E-05
b4312	*fimB*	tyrosine recombinase/inversion of on/off regulator of fimA	-1.203531495	8.253963012	3.75E-07	5.27E-06
b1696	*ydiP*	predicted DNA-binding transcriptional regulator	-1.203224619	8.148570176	7.25E-07	9.67E-06
b2339	*yfcV*	predicted fimbrial-like adhesin protein	-1.199458205	7.832259086	1.36E-05	0.000132159
b2370	*evgS*	hybrid sensory histidine kinase in two-component regulatory system with EvgA	-1.198206081	10.28384016	1.62E-10	4.63E-09
b4257	*yjgN*	inner membrane protein, DUF898 family	-1.195643017	7.807392278	1.55E-06	1.88E-05
b2109	*yehB*	predicted outer membrane protein	-1.194292266	10.83511631	3.48E-12	1.33E-10
b2548	*yphF*	predicted sugar transporter subunit: periplasmic-binding component of ABC superfamily	-1.192629595	9.462481198	5.10E-09	1.03E-07
b0301	*ykgB*	inner membrane protein, DUF417 family	-1.190460499	9.224151594	7.85E-10	1.94E-08
b4145	*yjeJ*	predicted protein	-1.184053058	9.32237182	3.39E-06	3.86E-05
b3587	*yiaW*	inner membrane protein, DUF3302 family	-1.180233388	7.603613332	4.21E-06	4.62E-05
b2354	*yfdK*	CPS-53 (KpLE1) prophage; conserved protein	-1.178240729	8.329689265	7.95E-07	1.05E-05
b2149	*mglA*	fused methyl-galactoside transporter subunits of ABC superfamily: ATP-binding components	-1.175244753	9.823508414	1.43E-12	5.77E-11
b2374	*frc*	formyl-CoA transferase, NAD(P)-binding	-1.17509232	9.023050437	3.54E-09	7.32E-08
b3626	*rfaJ*	UDP-D-glucose:(galactosyl)lipopolysaccharide glucosyltransferase	-1.171343923	8.614695214	1.94E-07	2.90E-06
b2777	*queE*	7-carboxy-7-deazaguanine synthase; queosine biosynthesis	-1.169619838	8.389229627	1.58E-08	2.84E-07
b4365	*yjjQ*	DNA-binding transcriptional regulator	-1.168267972	9.934923621	1.29E-10	3.77E-09
b3557	*insJ*	IS150 transposase A	-1.167087619	8.511742205	1.54E-08	2.80E-07

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b1877	*yecT*	predicted protein	-1.161305301	8.34858634	6.73E-06	7.09E-05
b1734	*chbF*	phospho-chitobiase; general 6-phospho-beta-glucosidase activity	-1.161277364	9.10667351	2.22E-05	0.000205151
b0461	*tomB*	Hha toxicity attenuator; conjugation-related protein	-1.159696091	9.00462877	5.57E-09	1.12E-07
b3861	*yihF*	conserved protein, DUF945 family	-1.159351242	9.551026411	2.43E-08	4.26E-07
b2522	*sseB*	rhodanase-like enzyme, sulfur transfer from thiosulfate	-1.15490054	7.850334032	1.66E-06	2.00E-05
b1163	*ycgF*	anti-repressor for YcgE, blue light-responsive; FAD-binding; has c-di-GMP phosphodiesterase-like EAL domain, but does not degrade c-di-GMP	-1.152874608	9.415209341	1.70E-09	3.75E-08
b3046	*yqiG*	pseudogene	-1.150245971	10.23982291	2.30E-11	7.67E-10
b0719	*ybgD*	predicted fimbrial-like adhesin protein	-1.149083982	9.121402892	4.07E-10	1.07E-08
b2252	*ais*	predicted LPS core heptose(II)-phosphate phosphatase	-1.144634752	8.7201984	3.06E-09	6.45E-08
b1330	*ynaI*	MscS family inner membrane protein	-1.142410365	8.631749688	8.92E-07	1.15E-05
b0901	*ycaK*	conserved protein	-1.137177728	9.056830125	1.06E-09	2.53E-08
b0902	*pflA*	pyruvate formate lyase activating enzyme 1	-1.136994547	7.445509477	7.46E-05	0.000611959
b1503	*ydeR*	predicted fimbrial-like adhesin protein	-1.136973348	7.706230988	5.94E-06	6.32E-05
b1363	*trkG*	Rac prophage; potassium transporter subunit	-1.134791129	8.40863797	6.29E-07	8.51E-06
b1762	*ynjI*	inner membrane protein	-1.1335447	9.382515228	1.48E-09	3.32E-08
b0533	*sfmH*	predicted fimbrial-like adhesin protein	-1.132556869	8.277990975	9.42E-07	1.20E-05
b1090	*plsX*	probable phosphate acyltransferase	-1.131175648	7.546904397	5.12E-05	0.000437407
b3896	*yiiG*	conserved lipoprotein	-1.119750808	10.10943649	5.74E-10	1.47E-08
b0300	*ykgA*	pseudogene	-1.118572119	7.458341308	5.49E-05	0.000464494
b3117	*tdcB*	catabolic threonine dehydratase, PLP-dependent	-1.115825894	9.240655614	9.70E-08	1.54E-06
b0647	*ybeT*	conserved protein, Sel1 family	-1.103064522	8.873245332	8.56E-07	1.12E-05
b2305	*yfcI*	conserved protein	-1.101293874	9.145529698	1.48E-05	0.000141541
b0532	*sfmD*	predicted outer membrane export usher protein	-1.097172728	9.900530234	3.41E-10	9.29E-09
b0587	*fepE*	regulator of length of O-antigen component of lipopolysaccharide chains	-1.093866067	9.049372104	5.99E-10	1.51E-08
b1450	*mcbR*	colanic acid and biofilm gene transcriptional regulator, MqsR-controlled	-1.092887117	8.71106595	2.65E-08	4.61E-07
b0462	*acrB*	multidrug efflux system protein	-1.08788792	8.743144273	8.09E-08	1.31E-06
b2032	*wbbK*	lipopolysaccharide biosynthesis protein	-1.072042705	9.083870957	6.75E-08	1.11E-06

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b1862	*yebB*	conserved protein, DUF830 family	-1.071989838	8.262282603	2.27E-05	0.000209315
b3104	*yhaI*	inner membrane protein, DUF805 family	-1.070614854	7.593778417	2.53E-05	0.000230938
b2272	*yfbM*	conserved protein, DUF1877 family	-1.049949122	8.119942133	4.21E-06	4.62E-05
b0463	*acrA*	multidrug efflux system	-1.049755086	8.071824756	2.72E-05	0.000244571
b2877	*mocA*	CTP:molybdopterin cytidylyltransferase	-1.045980103	8.944150757	4.48E-09	9.16E-08
b3680	*yidL*	predicted transcriptional regulator, AraC family	-1.039687723	9.451881017	4.34E-08	7.33E-07
b0787	*ybhM*	inner membrane protein, UPF0005 family	-1.035324624	9.436042147	1.09E-08	2.03E-07
b3720	*bglH*	carbohydrate-specific outer membrane porin, cryptic	-1.034303178	10.54424559	1.48E-09	3.32E-08
b1155	*tfaP*	e14 prophage; predicted protein	-1.030144578	8.475101878	1.39E-06	1.70E-05
b1535	*ydeH*	diguanylate cyclase, required for pgaD induction	-1.029861125	8.759946082	2.74E-07	3.98E-06
b3595	*yibJ*	pseudogene	-1.011347031	8.031431271	4.66E-06	5.06E-05
b2677	*proV*	glycine betaine transporter subunit	-1.007555473	10.37998213	8.72E-10	2.11E-08
b4133	*cadC*	DNA-binding transcriptional activator	-1.006872739	10.29547962	2.92E-07	4.20E-06
b3214	*gltF*	periplasmic protein	-0.998682777	10.03681597	1.96E-07	2.92E-06
b2863	*ygeQ*	pseudogene	-0.995910287	9.401537433	1.68E-07	2.54E-06
b1579	*intQ*	pseudogene	-0.995438739	9.420467681	1.08E-06	1.35E-05
b2373	*oxc*	oxalyl CoA decarboxylase, ThDP-dependent	-0.988465811	9.423875803	2.29E-06	2.71E-05
b0584	*fepA*	iron-enterobactin outer membrane transporter	-0.979666383	7.957473086	1.70E-05	0.000160499
b1766	*sppA*	protease IV (signal peptide peptidase)	-0.977610384	8.06972095	5.91E-05	0.00049256
b1798	*leuE*	neutral amino-acid efflux system	-0.976195088	9.087933286	4.69E-07	6.54E-06
b2355	*yfdL*	pseudogene	-0.976005081	8.098584525	8.77E-05	0.000697723
b3562	*yiaA*	inner membrane protein, YiaAB family	-0.974672533	8.969366946	6.19E-07	8.41E-06
b4579	*yaiX*	pseudogene	-0.971977321	8.078354099	7.97E-05	0.000644241
b3874	*yihN*	inner membrane protein, predicted transporter	-0.970199332	10.16320322	7.12E-09	1.38E-07
b1495	*yddB*	predicted porin protein	-0.945272475	10.44600106	1.56E-08	2.82E-07
b0562	*ybcY*	pseudogene	-0.933848785	9.3307018	3.78E-06	4.23E-05
b4310	*nanM*	N-acetylneuraminic acid mutarotase	-0.932110673	10.14120244	2.61E-08	4.57E-07
b0551	*quuD*	DLP12 prophage; predicted antitermination protein	-0.931324873	8.205501911	7.79E-06	8.12E-05
b3625	*rfaY*	lipopolysaccharide core biosynthesis protein	-0.927136467	7.973416343	8.93E-05	0.000708095
b1932	*yedL*	predicted acyltransferase	-0.922789269	8.658063311	1.01E-06	1.29E-05
b2845	*yqeG*	predicted transporter	-0.922180688	9.647045084	1.23E-06	1.52E-05
b2274	*yfbO*	conserved protein	-0.921099619	8.016976055	2.73E-05	0.000244664

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b1309	*ycjM*	predicted glucosyltransferase	-0.916416195	9.082833557	6.06E-07	8.27E-06
b3508	*yhiD*	predicted Mg(2+) transport ATPase, inner membrane protein	-0.905877307	9.624406262	4.35E-08	7.33E-07
b0139	*htrE*	predicted outer membrane usher protein	-0.90570193	11.17622988	1.10E-07	1.72E-06
b2734	*pphB*	serine/threonine-specific protein phosphatase 2	-0.900935822	8.920122778	4.61E-06	5.02E-05
b4088	*alsB*	D-allose transporter subunit	-0.900326103	9.449144566	7.81E-07	1.03E-05
b1695	*ydiO*	predicted acyl-CoA dehydrogenase	-0.894495714	7.979168137	8.66E-05	0.00069027
b1319	*ompG*	outer membrane porin G	-0.891981329	9.198407592	8.03E-06	8.33E-05
b1769	*ydjE*	predicted transporter	-0.891465374	9.431380262	6.65E-05	0.000547701
b2351	*gtrB*	CPS-53 (KpLE1) prophage; bactoprenol glucosyl transferase	-0.888322646	9.997551625	8.85E-06	9.07E-05
b0834	*yliF*	predicted diguanylate cyclase	-0.885639832	9.755485038	1.71E-05	0.000160738
b1494	*pqqL*	predicted peptidase	-0.880779347	9.883791154	6.45E-08	1.06E-06
b2028	*ugd*	UDP-glucose 6-dehydrogenase	-0.879922671	9.231397934	1.12E-06	1.41E-05
b0703	*ybfO*	pseudogene	-0.877983038	9.780967594	1.43E-07	2.20E-06
b3876	*yihO*	predicted transporter	-0.871924504	10.29539353	3.42E-07	4.87E-06
b4524	*ycjV*	pseudogene	-0.862298172	8.07733218	7.60E-05	0.000618411
b1969	*yedW*	predicted DNA-binding response regulator in two-component system with YedV	-0.855763785	9.38082216	1.03E-06	1.31E-05
b1238	*tdk*	thymidine kinase/deoxyuridine kinase	-0.847970321	9.010128033	3.41E-06	3.86E-05
b0135	*yadC*	predicted fimbrial-like adhesin protein	-0.846776492	10.31512954	5.73E-06	6.12E-05
b0343	*lacY*	lactose permease	-0.84254226	9.043489083	2.34E-05	0.000214854
b1967	*hchA*	Glyoxalase III and Hsp31 molecular chaperone	-0.83901273	9.300435312	5.20E-06	5.61E-05
b0645	*ybeR*	predicted protein	-0.832220083	8.905084811	8.75E-06	9.00E-05
b3324	*gspC*	general secretory pathway component, cryptic	-0.809281154	9.657420425	4.80E-06	5.20E-05
b1490	*dosC*	diguanylate cyclase; cold-and stationary phase-induced oxygen-dependent biofilm regulator; positively regulates csgBAC and pgaABCD	-0.808140044	9.472360155	1.65E-05	0.000156212
b2035	*rfc*	O-antigen polymerase	-0.80495099	9.794978569	3.46E-06	3.91E-05
b4317	*fimD*	outer membrane usher protein, type 1 fimbrial synthesis	-0.799748364	11.33819063	4.17E-06	4.60E-05
b0319	*yahE*	predicted protein	-0.798534151	8.80498941	7.81E-05	0.000633099

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b1024	*pgaA*	biofilm adhesin polysaccharide PGA export, predicted OM protein	-0.796333055	10.87637205	1.49E-05	0.000141541
b1250	*kch*	voltage-gated potassium channel	-0.787502757	9.296065846	1.56E-05	0.000148348
b4571	*wbbL*	pseudogene	-0.782071724	9.037149272	8.93E-06	9.13E-05
b2761	*ygcB*	Cas3 predicted helicase needed for Cascade anti-viral activity	-0.778511293	10.72104568	3.37E-06	3.84E-05
b0342	*lacA*	thiogalactoside acetyltransferase	-0.775347982	9.172445734	2.66E-05	0.000240392
b0513	*ybbY*	predicted uracil/xanthine transporter	-0.773370601	9.003360704	2.60E-05	0.000235914
b4316	*fimC*	chaperone, periplasmic	-0.766348305	9.086305593	7.20E-05	0.000592035
b3078	*ygjI*	predicted transporter	-0.764235471	9.647310898	1.20E-05	0.00011841
b4600	*ydfJ*	pseudogene	-0.758557687	12.82441042	3.03E-06	3.47E-05
b3561	*wecH*	O-acetyltransferase for enterobacterial common antigen (ECA)	-0.748888965	10.65888916	2.39E-06	2.82E-05
b3558	*insK*	IS150 transposase B	-0.747344304	9.509502969	1.38E-05	0.000133094
b4580	*yaiT*	pseudogene	-0.744090084	10.89649032	8.03E-06	8.33E-05
b1905	*ftnA*	ferritin iron storage protein (cytoplasmic)	-0.740152145	9.321156153	5.56E-05	0.000467007
b4061	*yjcC*	predicted cyclic-di-GMP phosphodiesterase	-0.740007277	10.35250906	1.83E-05	0.000171237
b3115	*tdcD*	propionate kinase/acetate kinase C, anaerobic	-0.739452329	9.469281515	1.10E-05	0.000109909
b0491	*ybbM*	inner membrane protein, UPF0014 family	-0.737588312	9.42112118	1.21E-05	0.000118541
b0328	*yahN*	amino acid exporter for proline, lysine, glutamate, homoserine	-0.729062653	9.906510561	6.64E-06	7.01E-05
b2888	*ygfU*	predicted transporter	-0.72692791	9.604597613	1.17E-05	0.000116136
b0076	*leuO*	DNA-binding transcriptional activator	-0.714901212	9.698672019	4.06E-05	0.00035496
b3491	*yhiM*	inner membrane protein, DUF1323 family	-0.709411531	10.31944483	8.09E-06	8.36E-05
b1786	*yeaJ*	predicted diguanylate cyclase	-0.708487114	9.768956848	4.81E-05	0.000413633
b1481	*bdm*	biofilm-dependent modulation protein	-0.695949086	9.552434043	5.33E-05	0.000454165
b4267	*idnD*	L-idonate 5-dehydrogenase, NAD-binding	-0.695454502	9.925470099	5.55E-05	0.000467007
b3173	*yhbX*	predicted hydrolase, inner membrane	-0.685830031	10.85080988	4.45E-05	0.000384048
b4031	*xylE*	D-xylose transporter	-0.668337796	10.47793716	6.56E-05	0.000541534
b2647	*ypjA*	adhesin-like autotransporter	-0.644866045	11.54410855	7.52E-05	0.000614885
b4308	*yjhR*	pseudogene	-0.641502971	10.57628901	7.55E-05	0.000615634
b3632	*rfaQ*	lipopolysaccharide core biosynthesis protein	0.694286991	11.07487251	6.02E-05	0.00050091

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3599	*mtlA*	fused mannitol-specific PTS enzymes: IIA components/IIB components/IIC components	0.715571516	9.246265304	4.44E-05	0.000384048
b3407	*yhgF*	predicted transcriptional accessory protein	0.733537643	9.428379208	2.31E-05	0.000212445
b4392	*slt*	lytic murein transglycosylase, soluble	0.734365726	8.901721873	3.76E-05	0.000330508
b0679	*nagE*	fused N-acetyl glucosamine specific PTS enzyme: IIC, IIB, and IIA components	0.742226493	8.97204689	3.19E-05	0.000282054
b4242	*mgtA*	magnesium transporter	0.749203753	8.994869232	5.90E-05	0.00049256
b3196	*yrbG*	predicted calcium/sodium:proton antiporter	0.761757199	8.886263696	5.34E-05	0.000454165
b0473	*htpG*	protein refolding molecular co-chaperone Hsp90, Hsp70-dependent; heat-shock protein; ATPase	0.795829746	9.182295924	7.25E-06	7.60E-05
b3780	*rhlB*	ATP-dependent RNA helicase	0.812988106	8.942263028	8.19E-05	0.000659255
b3416	*malQ*	4-alpha-glucanotransferase (amylomaltase)	0.833069967	8.826753437	5.49E-06	5.89E-05
b2552	*hmp*	fused nitric oxide dioxygenase/dihydropteridine reductase 2	0.845105714	8.728029963	1.44E-05	0.000138086
b3207	*yrbL*	predicted protein	0.845967038	8.201077483	7.48E-05	0.00061249
b0153	*fhuB*	fused iron-hydroxamate transporter subunits of ABC superfamily: membrane components	0.848944659	8.483643109	1.37E-05	0.000132754
b3785	*wzzE*	Entobacterial Common Antigen (ECA) polysaccharide chain length modulation protein	0.861135242	9.014273264	1.63E-06	1.97E-05
b2905	*gcvT*	aminomethyltransferase, tetrahydrofolate-dependent, subunit (T protein) of glycine cleavage complex	0.861389718	9.058284688	1.04E-06	1.31E-05
b0284	*paoC*	PaoABC aldehyde oxidoreductase, Moco-containing subunit	0.88026726	8.797956361	2.73E-06	3.17E-05
b3084	*rlmG*	23S rRNA mG1835 methyltransferase, SAM-dependent	0.88249578	8.449008862	2.15E-05	0.000199458
b3469	*zntA*	zinc, cobalt and lead efflux system	0.885831466	9.067405664	5.77E-07	7.91E-06
b4221	*ytfN*	large conserved protein, DUF490 family	0.90906686	10.38681983	4.82E-06	5.21E-05
b3159	*yhbV*	predicted protease	0.90928033	8.310925829	8.29E-05	0.000665683
b4015	*aceA*	isocitrate lyase	0.929864755	8.306277094	4.01E-06	4.44E-05
b3092	*uxaC*	uronate isomerase	0.950857974	8.648850532	2.41E-06	2.82E-05
b3260	*dusB*	tRNA-dihydrouridine synthase B	0.958074386	8.951286484	1.91E-06	2.29E-05
b2837	*galR*	DNA-binding transcriptional repressor	0.964197677	8.266420316	1.87E-05	0.000174586
b0699	*ybfA*	predicted protein	0.974684343	8.381652672	3.14E-06	3.59E-05

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b2158	*yeiH*	inner membrane protein, UPF0324 family	0.9885938	8.424883109	2.19E-06	2.60E-05
b4014	*aceB*	malate synthase A	0.991564554	8.43951499	4.89E-07	6.77E-06
b4102	*phnF*	predicted DNA-binding transcriptional regulator of phosphonate uptake and biodegradation	0.991614453	8.12889453	1.00E-05	0.000101321
b0852	*rimK*	ribosomal protein S6 modification protein	0.991860666	7.698005844	6.48E-05	0.000536733
b3510	*hdeA*	stress response protein acid-resistance protein	1.010529878	10.86154642	2.91E-10	7.99E-09
b3806	*cyaA*	adenylate cyclase	1.012503491	8.003740129	6.21E-06	6.58E-05
b2938	*speA*	biosynthetic arginine decarboxylase, PLP-binding	1.025830984	8.925381009	6.36E-08	1.05E-06
b0676	*nagC*	DNA-binding transcriptional dual regulator, repressor of N-acetylglucosamine	1.027890873	8.711887335	2.67E-06	3.11E-05
b3635	*mutM*	formamidopyrimidine/5-formyluracil/ 5-hydroxymethyluracil DNA glycosylase	1.041047421	8.2840559	1.22E-06	1.51E-05
b0195	*yaeB*	conserved protein, UPF0066 family	1.048015354	7.768166043	2.45E-05	0.000223675
b1129	*phoQ*	sensory histidine kinase in two-component regulatory system with PhoP	1.052989651	8.306831816	8.73E-07	1.13E-05
b0877	*ybjX*	conserved protein	1.06567693	9.362589767	2.37E-09	5.09E-08
b3940	*metL*	fused aspartokinase II/homoserine dehydrogenase II	1.066410645	9.50307197	1.51E-09	3.37E-08
b2667	*ygaV*	tributyltin-inducible repressor of ygaVP	1.113122349	8.783471812	3.64E-07	5.16E-06
b0605	*ahpC*	alkyl hydroperoxide reductase, C22 subunit	1.132967881	7.920194506	3.07E-05	0.000272705
b4111	*proP*	proline/glycine betaine transporter	1.140957816	9.351877395	1.13E-10	3.34E-09
b2709	*norR*	Anaerobic nitric oxide reductase DNA-binding transcriptional activator	1.14822518	8.452620606	1.94E-06	2.31E-05
b3961	*oxyR*	DNA-binding transcriptional dual regulator	1.154961893	8.649424927	6.54E-09	1.29E-07
b4251	*yjgJ*	predicted transcriptional regulator	1.175296669	8.530257614	1.23E-09	2.83E-08
b0198	*metI*	DL-methionine transporter subunit	1.198079013	7.568719218	1.25E-05	0.000121842
b4391	*yjjK*	fused predicted transporter subunits of ABC superfamily: ATP-binding components	1.239490171	8.905911815	7.19E-11	2.19E-09
b1823	*cspC*	stress protein, member of the CspA-family	1.257384831	9.372692153	1.39E-12	5.66E-11
b4457	*csrC*		1.259187624	7.845955896	1.59E-07	2.42E-06
b3753	*rbsR*	DNA-binding transcriptional repressor of ribose metabolism	1.290988827	8.336823302	3.41E-08	5.83E-07
b4371	*rsmC*	16S rRNA m(2)G1207 methyltransferase	1.332765305	7.521554415	1.06E-06	1.34E-05
b0197	*metQ*	DL-methionine transporter subunit	1.334155247	7.657893703	1.62E-05	0.000153646

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3963	*fabR*	DNA-binding transcriptional repressor	1.33581746	8.583121957	1.41E-11	4.86E-10
b2903	*gcvP*	glycine decarboxylase, PLP-dependent, subunit (protein P) of glycine cleavage complex	1.35161472	10.05436586	3.20E-16	2.06E-14
b2579	*yfiD*	autonomous glycyl radical cofactor	1.424867312	7.097096855	1.20E-05	0.000118535
b1049	*opgH*	membrane glycosyltransferase	1.430159352	8.714192918	1.30E-14	7.28E-13
b2477	*bamC*	lipoprotein required for OM biogenesis, in BamABCDE complex	1.434393834	8.621491933	3.66E-11	1.16E-09
b0677	*nagA*	N-acetylglucosamine-6-phosphate deacetylase	1.445788816	8.295694849	7.01E-11	2.14E-09
b4375	*prfC*	peptide chain release factor RF-3	1.446165855	7.990548622	7.13E-09	1.38E-07
b0436	*tig*	peptidyl-prolyl cis/trans isomerase (trigger factor)	1.448694457	8.858078821	1.92E-14	1.06E-12
b3676	*yidH*	inner membrane protein, DUF202 family	1.535974682	7.12622966	9.80E-06	9.93E-05
b3871	*typA*	GTP-binding protein	1.5602021	8.924178184	1.87E-16	1.25E-14
b4260	*pepA*	multifunctional aminopeptidase A: a cyteinylglycinase, transcription regulator and site-specific recombination factor	1.567148895	9.012174849	7.27E-15	4.23E-13
b4377	*yjjU*	predicted phospholipase, patatin-like family	1.598833457	8.911410952	1.74E-17	1.46E-15
b4408	*csrB*		1.621428872	8.114593358	2.33E-13	1.07E-11
b0814	*ompX*	outer membrane protein X	1.628397378	8.245466294	4.29E-14	2.28E-12
b0196	*rcsF*	predicted outer membrane protein, signal	1.637082356	7.474830797	2.23E-08	3.96E-07
b2682	*ygaZ*	probable L-valine exporter, norvaline resistance	1.638270041	7.476135854	1.21E-08	2.21E-07
b1048	*opgG*	osmoregulated periplasmic glucan (OPG) biosynthesis periplasmic protein	1.661877057	8.465671151	3.28E-13	1.45E-11
b2217	*rcsB*	DNA-binding response regulator in two-component regulatory system with RcsC and YojN	1.680189754	8.601556572	6.46E-14	3.26E-12
b1304	*pspA*	regulatory protein for phage-shock-protein operon	1.734264495	7.4750051	2.80E-07	4.05E-06
b3687	*ibpA*	heat shock chaperone	1.74321486	7.882240361	1.46E-13	7.06E-12
b0059	*rapA*	RNA polymerase-associated helicase protein (ATPase and RNA polymerase recycling factor)	1.775236446	10.83567125	1.08E-18	1.11E-16
b2513	*yfgM*	conserved protein, UPF0070 family	1.788714507	7.055356501	6.15E-07	8.38E-06
b2168	*fruK*	fructose-1-phosphate kinase	1.789724841	7.591052313	4.35E-11	1.36E-09
b3556	*cspA*	RNA chaperone and anti-terminator, cold-inducible	1.794562115	8.284949421	1.02E-14	5.77E-13

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b4390	*nadR*	bifunctional DNA-binding transcriptional repressor/ NMN adenylyltransferase	1.825438812	9.915638455	1.86E-25	3.39E-23
b1914	*uvrY*	DNA-binding response regulator in two-component regulatory system with BarA	1.826116835	9.89980296	2.15E-21	3.10E-19
b2215	*ompC*	outer membrane porin protein C	1.890763507	9.622987328	2.63E-12	1.03E-10
b1831	*proQ*	RNA chaperone, probable regulator of ProP translation	1.890863014	7.498213787	1.34E-09	3.07E-08
b2786	*barA*	hybrid sensory histidine kinase, in two-component regulatory system with UvrY	1.892488051	12.0721507	7.22E-26	1.51E-23
b3191	*mlaB*	ABC transporter maintaining OM lipid asymmetry, cytoplasmic STAS component	1.9012909	9.149911519	3.39E-21	4.59E-19
b2516	*rodZ*	cytoskeletal protein required for MreB assembly	1.902914388	6.871360278	1.13E-06	1.41E-05
b3181	*greA*	transcript cleavage factor	1.932895225	7.301730688	5.06E-08	8.49E-07
b0759	*galE*	UDP-galactose-4-epimerase	1.959739202	7.236735029	2.11E-09	4.59E-08
b3206	*npr*	phosphohistidinoprotein-hexose phosphotransferase component of N-regulated PTS system (Npr)	1.99618964	6.68390844	3.93E-06	4.38E-05
b3192	*mlaC*	ABC transporter maintaining OM lipid asymmetry, periplasmic binding protein	2.003098266	8.379494313	3.78E-11	1.19E-09
b3194	*mlaE*	ABC transporter maintaining OM lipid asymmetry, inner membrane permease protein	2.018592104	7.961915861	5.39E-14	2.76E-12
b2346	*mlaA*	ABC transporter maintaining OM lipid asymmetry, OM lipoprotein component	2.055634878	9.137104056	7.31E-18	7.13E-16
b2815	*metW*		2.110796386	6.365243483	1.41E-05	0.000135824
b2829	*ptsP*	fused PTS enzyme: PEP-protein phosphotransferase (enzyme I)/GAF domain containing protein	2.15792221	10.56294265	3.16E-32	8.83E-30
b0611	*rna*	ribonuclease I	2.256328189	9.389455407	1.02E-17	9.51E-16
b3195	*mlaF*	ABC transporter maintaining OM lipid asymmetry, ATP-binding protein	2.256809534	10.12463383	4.65E-35	1.50E-32
b3071	*yqjI*	predicted transcriptional regulator, PadR family	2.29796337	7.606402996	5.58E-17	4.50E-15
b3229	*sspA*	stringent starvation protein A	2.572029868	6.833663536	1.39E-06	1.70E-05
b4368	*leuV*		2.680726236	6.298480809	2.14E-06	2.54E-05
b4203	*rplI*	50S ribosomal subunit protein L9	2.711454721	8.665483891	5.18E-39	1.81E-36
b1095	*fabF*	3-oxoacyl-[acyl-carrier-protein] synthase II	2.73202337	9.066532399	8.50E-43	3.88E-40

locus_tag	gene_name	function	logFC	logCPM	Pvalue	q.value (cutoff 0.000716332)
b3193	*mlaD*	ABC transporter maintaining OM lipid asymmetry, anchored periplasmic binding protein	2.883968103	8.362859797	5.92E-30	1.55E-27
b1849	*purT*	phosphoribosylglycinamide formyltransferase 2	3.114932089	9.805819837	1.08E-55	4.51E-52
b0623	*cspE*	DNA-binding transcriptional repressor	3.157081834	8.88961655	2.13E-48	1.78E-45
b2572	*rseA*	anti-sigma factor	3.177657285	7.562183118	2.93E-21	4.10E-19
b2684	*mprA*	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux	3.500495683	8.607309689	9.24E-43	3.88E-40
b3783	*rho*	transcription termination factor	3.636104597	6.545837709	4.48E-13	1.88E-11
b2617	*bamE*	lipoprotein component of BamABCDE OM biogenesis complex	3.845467948	9.162334554	9.38E-49	9.83E-46
b2571	*rseB*	anti-sigma E factor, binds RseA	4.174457428	9.843858702	6.51E-42	2.48E-39
b0178	*skp*	periplasmic chaperone	4.488457614	9.020232339	2.90E-27	6.76E-25
b2512	*bamB*	lipoprotein required for OM biogenesis, in BamABCDE complex	4.575863157	9.087828104	5.48E-47	3.83E-44

Table A13 -
Glycerol nitrate

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b1334	*fnr*	DNA-binding transcriptional dual regulator, global regulator of anaerobic growth	-10.24626081	6.211411398	1.19E-27	5.57E-25
b3008	*metC*	cystathionine beta-lyase, PLP-dependent	-9.955337888	5.925089284	2.33E-20	6.99E-18
b0009	*mog*	molybdochelatase incorporating molybdenum into molybdopterin	-9.872761617	5.843997623	3.10E-20	8.67E-18
b0438	*clpX*	ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease	-9.252985917	5.240499335	1.86E-14	2.52E-12
b0783	*moaC*	molybdopterin biosynthesis, protein C	-9.189591971	5.180111975	3.12E-11	2.85E-09
b2752	*cysD*	sulfate adenylyltransferase, subunit 2	-9.012963279	5.009761906	2.01E-11	1.92E-09
b3772	*ilvA*	threonine deaminase	-9.012331387	5.009438246	5.60E-13	6.35E-11
b4013	*metA*	homoserine O-transsuccinylase	-8.833938691	4.840219087	2.39E-10	1.80E-08
b3956	*ppc*	phosphoenolpyruvate carboxylase	-8.633073414	4.650250024	2.49E-11	2.32E-09
b3781	*trxA*	thioredoxin 1	-8.603182574	4.621527682	5.85E-09	3.23E-07
b3771	*ilvD*	dihydroxyacid dehydratase	-8.361368855	4.397260819	4.41E-10	3.04E-08
b0764	*modB*	molybdate transporter subunit	-8.250530276	6.471077013	4.44E-34	3.73E-31
b2019	*hisG*	ATP phosphoribosyltransferase	-8.183196809	4.233408478	2.97E-09	1.73E-07
b3290	*trkA*	NAD-binding component of TrK potassium transporter	-8.140044505	4.194227506	2.81E-09	1.66E-07
b2500	*purN*	phosphoribosylglycinamide formyltransferase 1	-8.096504323	4.154325851	2.98E-08	1.44E-06
b1264	*trpE*	component I of anthranilate synthase	-8.050100586	4.112759867	6.14E-09	3.35E-07
b1677	*lpp*	murein lipoprotein	-7.978243272	4.048371416	2.59E-08	1.31E-06
b0243	*proA*	gamma-glutamylphosphate reductase	-7.904037618	3.9815428	3.01E-08	1.44E-06
b3204	*ptsN*	sugar-specific enzyme IIA component of PTS	-7.903943688	3.981509959	2.70E-08	1.35E-06
b0003	*thrB*	homoserine kinase	-7.903476508	3.981345756	3.05E-08	1.44E-06
b3770	*ilvE*	branched-chain amino-acid aminotransferase	-7.851802584	3.935044621	4.73E-08	2.16E-06
b1088	*yceD*	conserved protein, DUF177 family	-7.825312741	3.911349436	1.26E-07	5.44E-06
b3734	*atpA*	F1 sector of membrane-bound ATP synthase, alpha subunit	-7.741944546	3.837597725	5.17E-07	2.01E-05
b3082	*higA*	antitoxinof the HigB-HigA toxin-antitoxin system	-7.71325284	3.812224266	1.72E-06	5.73E-05
b2557	*purL*	phosphoribosylformyl-glycineamide synthetase	-7.651981655	3.759501439	1.65E-07	6.81E-06
b0738	*tolR*	membrane spanning protein in TolA-TolQ-TolR complex	-7.65028289	3.75902092	1.66E-05	0.000427829

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b3941	*metF*	5,10-methylenetetrahydrofolate reductase	-7.588975674	3.705121332	7.16E-07	2.68E-05
b4214	*cysQ*	PAPS (adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase	-7.558830313	3.677859366	5.02E-06	0.000149545
b0052	*pdxA*	4-hydroxy-L-threonine phosphate dehydrogenase, NAD-dependent	-7.492002003	3.620451362	6.48E-06	0.00018905
b0741	*pal*	peptidoglycan-associated outer membrane lipoprotein	-7.45477837	3.59017803	1.01E-05	0.000281163
b1262	*trpC*	fused indole-3-glycerolphosphate synthetase/N-(5-phosphoribosyl)anthranilate isomerase	-7.39254793	5.630035951	3.35E-16	5.41E-14
b2313	*cvpA*	membrane protein required for colicin V production	-7.383571895	3.529375458	3.75E-06	0.000115062
b2526	*hscA*	DnaK-like molecular chaperone specific for IscU	-7.347572605	3.498271265	2.06E-06	6.69E-05
b0826	*moeB*	molybdopterin synthase sulfurylase	-7.328838687	5.568594659	1.12E-15	1.68E-13
b0386	*proC*	pyrroline-5-carboxylate reductase, NAD(P)-binding	-7.308195285	3.46583354	4.41E-06	0.000133172
b3927	*glpF*	glycerol facilitator	-7.280668408	5.521697188	5.16E-16	8.03E-14
b2026	*hisI*	fused phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase	-7.268517487	3.432892934	1.33E-05	0.000356923
b0908	*aroA*	5-enolpyruvylshikimate-3-phosphate synthetase	-7.230240502	3.399821552	1.10E-05	0.00030275
b2414	*cysK*	cysteine synthase A, O-acetylserine sulfhydrolase A subunit	-7.188810722	3.365456378	1.73E-05	0.000443403
b2020	*hisD*	bifunctional histidinal dehydrogenase/ histidinol dehydrogenase	-7.188338366	3.365346269	7.12E-06	0.000204728
b1693	*aroD*	3-dehydroquinate dehydratase	-7.187471462	3.365126066	4.98E-06	0.000149458
b0004	*thrC*	threonine synthase	-7.145061556	3.330009843	5.94E-06	0.000175646
b2320	*pdxB*	erythronate-4-phosphate dehydrogenase	-7.05633063	3.257145443	2.41E-05	0.000599298
b0242	*proB*	gamma-glutamate kinase	-6.960061593	3.180087149	1.64E-05	0.000424326
b3926	*glpK*	glycerol kinase	-6.959755009	6.555033024	1.71E-29	1.03E-26
b2784	*relA*	(p)ppGpp synthetase I/GTP pyrophosphokinase	-6.919090338	7.200899084	1.16E-22	4.07E-20
b0726	*sucA*	2-oxoglutarate decarboxylase, thiamin-requiring	-6.910831659	3.140297288	6.55E-05	0.001529395
b3839	*tatC*	TatABCE protein translocation system subunit	-6.856261099	6.452927687	2.92E-26	1.23E-23
b0912	*ihfB*	integration host factor (IHF), DNA-binding protein, beta subunit	-6.815496877	5.075314054	1.47E-12	1.59E-10
b0071	*leuD*	3-isopropylmalate dehydratase small subunit	-6.815356435	5.075241016	1.78E-12	1.87E-10
b0439	*lon*	DNA-binding ATP-dependent protease La	-6.803234598	3.056535283	6.29E-05	0.001482418

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b3170	*rimP*	ribosome maturation factor for 30S subunits	-6.748171863	3.013190136	7.76E-05	0.001757259
b0720	*gltA*	citrate synthase	-6.510930037	4.787931669	1.21E-10	9.78E-09
b2424	*cysU*	sulfate/thiosulfate ABC transporter permease	-6.497066524	4.774798545	8.92E-11	7.35E-09
b1263	*trpD*	fused glutamine amidotransferase (component II) of anthranilate synthase/anthranilate phosphoribosyl transferase	-6.483230722	4.76165311	3.50E-10	2.53E-08
b0073	*leuB*	3-isopropylmalate dehydrogenase, NAD(+)-dependent	-5.916094831	5.030371165	6.75E-11	5.67E-09
b0002	*thrA*	fused aspartokinase I and homoserine dehydrogenase I	-5.817988964	5.446649728	4.03E-13	4.84E-11
b3368	*cysG*	fused siroheme synthase 1,3-dimethyluroporphyriongen III dehydrogenase and siroheme ferrochelatase/uroporphyrinogen methyltransferase	-5.779727265	4.902131423	1.16E-09	7.15E-08
b0033	*carB*	carbamoyl-phosphate synthase large subunit	-5.669121254	4.800049428	3.03E-08	1.44E-06
b3828	*metR*	DNA-binding transcriptional activator, homocysteine-binding	-5.648938996	4.003315945	2.31E-07	9.41E-06
b0740	*tolB*	periplasmic protein	-5.639549478	4.773330429	1.27E-06	4.41E-05
b3671	*ilvB*	acetolactate synthase I, large subunit	-5.5558189	7.726688674	2.76E-48	1.16E-44
b1712	*ihfA*	integration host factor (IHF), DNA-binding protein, alpha subunit	-5.516217699	3.886095238	1.01E-06	3.66E-05
b2600	*tyrA*	fused chorismate mutase T/prephenate dehydrogenase	-5.45793055	3.83617999	1.17E-06	4.13E-05
b0957	*ompA*	outer membrane protein A (3a;II*;G;d)	-5.425575237	4.574686527	7.27E-08	3.28E-06
b0784	*moaD*	molybdopterin synthase, small subunit	-5.408369137	4.559079612	1.14E-08	6.07E-07
b3867	*hemN*	coproporphyrinogen III oxidase, SAM and NAD(P)H dependent, oxygen-independent	-5.346959759	7.089604661	5.14E-43	1.08E-39
b0072	*leuC*	3-isopropylmalate dehydratase large subunit	-5.318104353	4.972596262	8.33E-10	5.38E-08
b2838	*lysA*	diaminopimelate decarboxylase, PLP-binding	-5.303404375	4.463559745	1.41E-07	6.05E-06
b0437	*clpP*	proteolytic subunit of ClpA-ClpP and ClpX-ClpP ATP-dependent serine proteases	-5.264264492	4.42939017	1.05E-06	3.77E-05
b3172	*argG*	argininosuccinate synthetase	-5.245744124	4.412501594	6.02E-07	2.32E-05
b0688	*pgm*	phosphoglucomutase	-5.236701228	3.648443237	2.92E-05	0.000708968
b2276	*nuoN*	NADH:ubiquinone oxidoreductase, membrane subunit N	-5.217122141	5.788746056	6.07E-15	8.50E-13
b2423	*cysW*	sulfate/thiosulfate ABC transporter subunit	-5.211525643	4.876536812	3.21E-09	1.82E-07
b0214	*rnhA*	ribonuclease HI, degrades RNA of DNA-RNA hybrids	-5.201371288	3.619207637	6.75E-06	0.0001956

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b2278	*nuoL*	NADH:ubiquinone oxidoreductase, membrane subunit L	-5.157596518	6.125592815	6.68E-19	1.48E-16
b2284	*nuoF*	NADH:ubiquinone oxidoreductase, chain F	-5.133465408	6.104163287	1.99E-16	3.35E-14
b4232	*fbp*	fructose-1,6-bisphosphatase I	-5.105293392	7.131913147	1.10E-28	5.80E-26
b0763	*modA*	molybdate transporter subunit	-5.096871071	6.847275953	4.17E-33	2.92E-30
b2281	*nuoI*	NADH:ubiquinone oxidoreductase, chain I	-4.945331449	4.63316728	1.60E-07	6.72E-06
b1224	*narG*	nitrate reductase 1, alpha subunit	-4.548347239	8.243603751	4.22E-40	5.90E-37
b2114	*metG*	methionyl-tRNA synthetase	-4.461424064	3.730783089	1.62E-05	0.000423375
b3957	*argE*	acetylornithine deacetylase	-4.301249935	5.145341843	1.48E-06	5.02E-05
b3939	*metB*	cystathionine gamma-synthase, PLP-dependent	-4.223981442	5.25434012	2.54E-06	8.01E-05
b3857	*mobA*	molybdopterin-guanine dinucleotide synthase	-4.193947399	5.2262159	3.39E-06	0.000105416
b3035	*tolC*	transport channel	-4.150020189	5.186175164	8.10E-07	2.96E-05
b2287	*nuoB*	NADH:ubiquinone oxidoreductase, chain B	-4.134807793	5.475694681	1.20E-06	4.21E-05
b2422	*cysA*	sulfate/thiosulfate transporter subunit	-4.061011652	4.922327991	1.49E-07	6.30E-06
b2318	*truA*	pseudouridylate synthase I	-4.040781518	5.390448467	2.15E-05	0.000540553
b4191	*ulaR*	transcriptional tepressor for the L-ascorbate utilization (ula) divergon	-3.904376924	5.39006463	3.78E-09	2.12E-07
b0588	*fepC*	iron-enterobactin transporter subunit	-3.847805648	5.214278359	2.07E-06	6.69E-05
b0827	*moeA*	molybdopterin biosynthesis protein	-3.797284326	5.970521434	1.04E-07	4.60E-06
b3205	*yhbJ*	glmZ(sRNA)-inactivating NTPase, glucosamine-6-phosphate regulated	-3.745723743	5.372166384	2.69E-05	0.000661581
b2283	*nuoG*	NADH:ubiquinone oxidoreductase, chain G	-3.718937619	6.88547825	1.41E-10	1.12E-08
b3958	*argC*	N-acetyl-gamma-glutamylphosphate reductase, NAD(P)-binding	-3.703615569	4.782312666	1.30E-05	0.000353404
b0781	*moaA*	molybdopterin biosynthesis protein A	-3.696847251	6.600487931	2.13E-18	4.26E-16
b3860	*dsbA*	periplasmic protein disulfide isomerase I	-3.611431123	7.13187855	3.81E-20	1.00E-17
b3630	*rfaP*	kinase that phosphorylates core heptose of lipopolysaccharide	-3.560400509	4.956162421	1.65E-07	6.81E-06
b0750	*nadA*	quinolinate synthase, subunit A	-3.550983836	5.193302118	4.63E-08	2.13E-06
b1225	*narH*	nitrate reductase 1, beta (Fe-S) subunit	-3.454662035	5.945121349	5.82E-11	4.99E-09
b2277	*nuoM*	NADH:ubiquinone oxidoreductase, membrane subunit M	-3.43537714	5.79545253	2.25E-10	1.72E-08
b0785	*moaE*	molybdopterin synthase, large subunit	-3.362008959	5.33424555	1.11E-06	3.95E-05
b2839	*lysR*	DNA-binding transcriptional dual regulator	-3.322288045	5.826769629	1.17E-09	7.15E-08

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b1223	*narK*	nitrate/nitrite transporter	-3.275999661	6.97486453	1.79E-18	3.76E-16
b2285	*nuoE*	NADH:ubiquinone oxidoreductase, chain E	-3.263644323	4.407593662	2.54E-05	0.000628067
b2425	*cysP*	thiosulfate-binding protein	-3.176685059	5.931632682	3.67E-10	2.61E-08
b2288	*nuoA*	NADH:ubiquinone oxidoreductase, membrane subunit A	-3.152713148	5.403574565	2.00E-05	0.000508942
b2286	*nuoC*	NADH:ubiquinone oxidoreductase, fused CD subunit	-3.113493188	6.078868704	7.18E-10	4.78E-08
b2282	*nuoH*	NADH:ubiquinone oxidoreductase, membrane subunit H	-3.064345773	5.076689129	8.58E-06	0.0002401
b4025	*pgi*	glucosephosphate isomerase	-3.007177885	5.419284534	6.32E-07	2.41E-05
b3426	*glpD*	sn-glycerol-3-phosphate dehydrogenase, aerobic, FAD/NAD(P)-binding	-2.996922432	6.934383293	1.56E-17	2.86E-15
b0590	*fepD*	iron-enterobactin transporter subunit	-2.91962325	5.199068219	1.21E-05	0.000330471
b0049	*apaH*	diadenosine tetraphosphatase	-2.86671944	5.159145591	6.32E-05	0.001482418
b3924	*fpr*	ferredoxin-NADP reductase	-2.772378824	6.662823721	6.52E-14	8.30E-12
b1235	*rssB*	response regulator binding RpoS to initiate proteolysis by ClpXP; required for the PcnB-degradosome interaction during stationary phase	-2.728416111	6.792665745	1.21E-13	1.49E-11
b3773	*ilvY*	DNA-binding transcriptional dual regulator	-2.705926894	5.417351207	1.55E-06	5.22E-05
b0143	*pcnB*	poly(A) polymerase I	-2.703467195	5.980480773	1.99E-08	1.02E-06
b3408	*feoA*	ferrous iron transporter, protein A	-2.642682127	7.131109436	5.13E-18	9.78E-16
b3613	*envC*	activator of AmiB,C murein hydrolases, septal ring factor	-2.629008568	6.798613789	8.59E-10	5.47E-08
b3387	*dam*	DNA adenine methyltransferase	-2.621230655	5.834235517	5.30E-05	0.001257015
b0585	*fes*	enterobactin/ferric enterobactin esterase	-2.60682122	5.735472752	1.89E-06	6.19E-05
b0928	*aspC*	aspartate aminotransferase, PLP-dependent	-2.575002564	5.310339859	7.61E-05	0.001737326
b3163	*nlpI*	lipoprotein involved in osmotic sensitivity and filamentation	-2.488073273	5.59035028	2.36E-05	0.000588894
b3236	*mdh*	malate dehydrogenase, NAD(P)-binding	-2.302785308	6.081978066	2.26E-06	7.25E-05
b3806	*cyaA*	adenylate cyclase	-2.294315909	6.620723115	1.85E-09	1.11E-07
b1185	*dsbB*	oxidoreductase that catalyzes reoxidation of DsbA protein disulfide isomerase I	-2.290185069	5.67585859	5.17E-05	0.001239575
b1221	*narL*	DNA-binding response regulator in two-component regulatory system with NarX (or NarQ)	-2.234719205	6.608429924	3.08E-09	1.77E-07
b0765	*modC*	molybdate transporter subunit	-2.177202322	6.063797148	1.43E-06	4.92E-05
b0051	*rsmA*	16S rRNA dimethyladenosine transferase, SAM-dependent	-2.14077022	5.921282882	2.71E-05	0.000662756

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b0080	*cra*	DNA-binding transcriptional repressor-activator for carbon metabolism	-2.062286024	6.411638626	1.46E-06	4.99E-05
b3409	*feoB*	fused ferrous iron transporter, protein B: GTP-binding protein/membrane protein	-2.008126611	8.542288277	4.72E-21	1.52E-18
b4383	*deoB*	phosphopentomutase	-1.972698094	7.113903977	2.03E-10	1.58E-08
b3631	*rfaG*	glucosyltransferase I	-1.916720297	6.852068873	1.84E-06	6.09E-05
b3026	*qseC*	quorum sensing sensory histidine kinase in two-component regulatory system with QseB	-1.882706064	7.016851866	3.08E-08	1.44E-06
b3574	*yiaJ*	DNA-binding transcriptional repressor of yiaK-S operon	-1.845313759	7.279072193	6.40E-10	4.33E-08
b0484	*copA*	copper transporter	-1.735698387	8.139142271	1.65E-17	2.88E-15
b0121	*speE*	spermidine synthase (putrescine aminopropyltransferase)	-1.686434832	6.345043762	1.47E-05	0.000394198
b2264	*menD*	bifunctional 2-oxoglutarate decarboxylase/ SHCHC synthase	-1.603688248	7.347690167	2.61E-07	1.05E-05
b1232	*purU*	formyltetrahydrofolate hydrolase	-1.601162905	6.602740737	6.03E-06	0.000177
b1176	*minC*	cell division inhibitor	-1.588083248	7.055586906	6.60E-07	2.50E-05
b1770	*ydjF*	predicted DNA-binding transcriptional regulator	-1.521942776	8.464514427	2.43E-15	3.52E-13
b1603	*pntA*	pyridine nucleotide transhydrogenase, alpha subunit	-1.506285611	7.121298073	1.56E-05	0.000415036
b1175	*minD*	membrane ATPase of the MinC-MinD-MinE system	-1.46356281	7.524913471	9.98E-09	5.37E-07
b2290	*alaA*	valine-pyruvate aminotransferase 2	-1.44457321	8.071996059	7.22E-12	7.40E-10
b1761	*gdhA*	glutamate dehydrogenase, NADP-specific	-1.425645416	7.095182256	2.47E-06	7.84E-05
b3616	*tdh*	threonine 3-dehydrogenase, NAD(P)-binding	-1.360979882	6.988192623	2.03E-05	0.000513675
b3166	*truB*	tRNA U55 pseudouridine synthase	-1.323030242	6.697151224	7.00E-05	0.001607462
b2965	*speC*	ornithine decarboxylase, constitutive	-1.306733364	8.066436442	1.16E-08	6.09E-07
b2501	*ppk*	polyphosphate kinase, component of RNA degradosome	-1.280099597	8.200549855	3.76E-10	2.63E-08
b4488	*ilvG*	pseudogene	-1.244649564	7.226809132	8.19E-06	0.000232301
b3930	*menA*	1,4-dihydroxy-2-naphthoate octaprenyltransferase	-1.242805799	7.055041582	5.19E-05	0.001239575
b0676	*nagC*	DNA-binding transcriptional dual regulator, repressor of N-acetylglucosamine	-1.235500786	7.561393271	1.17E-05	0.000321694
b3252	*csrD*	targeting factor for csrBC sRNA degradation	-1.229031342	8.123824136	2.94E-07	1.17E-05
b2707	*srlR*	DNA-bindng transcriptional repressor	-1.204260533	7.206828145	7.51E-06	0.000214665
b1222	*narX*	sensory histidine kinase in two-component regulatory system with NarL	-1.162663029	7.581364564	7.66E-07	2.82E-05

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b1812	*pabB*	aminodeoxychorismate synthase, subunit I	-1.155004069	7.597660336	4.11E-07	1.61E-05
b2144	*sanA*	vancomycin high temperature exclusion protein; mutants have a defective envelope more permeable to vancomycin at 42 degrees	-1.098747531	9.359237045	9.46E-10	5.93E-08
b0584	*fepA*	iron-enterobactin outer membrane transporter	-0.892292631	7.927569063	1.62E-05	0.000423375
b4390	*nadR*	bifunctional DNA-binding transcriptional repressor/ NMN adenylyltransferase	0.729386652	9.087215147	1.60E-05	0.000421446
b0401	*brnQ*	Branched-chain amino acid transport system 2 carrier protein; LIV-II transport system for Ile, Leu, and Val	0.773879073	8.423668391	6.94E-05	0.001601407
b2158	*yeiH*	inner membrane protein, UPF0324 family	0.782708693	8.234698098	4.60E-05	0.001110182
b0059	*rapA*	RNA polymerase-associated helicase protein (ATPase and RNA polymerase recycling factor)	0.816609275	10.10051801	8.43E-06	0.000237597
b4141	*yjeH*	predicted transporter	0.818915265	8.934355898	7.24E-07	2.69E-05
b1823	*cspC*	stress protein, member of the CspA-family	0.849725964	9.042005844	3.62E-06	0.00011166
b3529	*yhjK*	predicted diguanylate cyclase	0.855478799	9.244665229	8.48E-08	3.79E-06
b3253	*yhdH*	predicted oxidoreductase, Zn-dependent and NAD(P)-binding	0.863854519	8.430353495	2.67E-06	8.37E-05
b2903	*gcvP*	glycine decarboxylase, PLP-dependent, subunit (protein P) of glycine cleavage complex	0.882809182	9.677320818	1.27E-08	6.60E-07
b1824	*yobF*	predicted protein	1.046005136	7.722692538	4.18E-06	0.000127328
b3780	*rhlB*	ATP-dependent RNA helicase	1.079850628	9.06160068	2.83E-08	1.40E-06
b2786	*barA*	hybrid sensory histidine kinase, in two-component regulatory system with UvrY	1.098021572	11.42990721	3.49E-11	3.12E-09
b1914	*uvrY*	DNA-binding response regulator in two-component regulatory system with BarA	1.10569565	9.316454154	1.13E-07	4.95E-06
b1198	*dhaM*	fused predicted dihydroxyacetone-specific PTS enzymes: HPr component/EI component	1.118807089	8.795159616	2.67E-07	1.07E-05
b2521	*sseA*	3-mercaptopyruvate sulfurtransferase	1.197426679	8.397586029	2.77E-10	2.04E-08
b1201	*dhaR*	DNA-binding transcription activator of the dhaKLM operon	1.267317206	9.201449493	1.77E-11	1.73E-09
b1200	*dhaK*	dihydroxyacetone kinase, PTS-dependent, dihydroxyacetone-binding subunit	1.267413298	8.06625522	7.93E-10	5.20E-08
b2571	*rseB*	anti-sigma E factor, binds RseA	1.350615189	7.379953847	6.73E-05	0.001560814

locus_tag	gene_name	function	logFC	logCPM	Pvalue (cutoff 0.05)	q.value (cutoff 0.001831501)
b3791	*rffA*	TDP-4-oxo-6-deoxy-D-glucose transaminase	1.411669084	8.752361927	1.32E-11	1.32E-09
b1199	*dhaL*	dihydroxyacetone kinase, C-terminal domain	1.469106005	6.655344072	7.78E-05	0.001757259
b1084	*rne*	fused ribonucleaseE: endoribonuclease/RNA-binding protein/RNA degradosome binding protein	1.50880981	7.992764854	4.81E-13	5.61E-11
b2168	*fruK*	fructose-1-phosphate kinase	1.675351141	7.449603936	4.40E-11	3.85E-09
b1284	*yciT*	global regulator of transcription; DeoR family	1.734680115	8.961422792	1.40E-19	3.45E-17
b1849	*purT*	phosphoribosylglycinamide formyltransferase 2	1.749533854	8.607558539	3.48E-19	8.12E-17
b3790	*rffC*	TDP-fucosamine acetyltransferase	1.8097286	7.857258347	3.55E-14	4.65E-12
b2912	*fau*	conserved protein, 5-formyltetrahydrofolate cyclo-ligase family	1.865522658	7.521691234	9.16E-13	1.01E-10
b3934	*cytR*	DNA-binding transcriptional dual regulator	2.02436647	8.793370599	1.95E-25	7.43E-23
b4260	*pepA*	multifunctional aminopeptidase A: a cyteinylglycinase, transcription regulator and site-specific recombination factor	2.405265436	9.621110782	3.17E-38	3.33E-35

Table A14. Comparative analysis of a mutant library grown under three different anaerobic conditions. Genes shared under all conditions

Gene	Function
aroA	Essential for growth in M9 medium (Joyce *et al*., 2006)
aroD	Essential for growth in M9 medium (Joyce *et al*., 2006)
carB	Essential for growth in M9 medium (Joyce *et al*., 2006)
cysD	Essential for growth in M9 medium (Joyce *et al*., 2006)
cysP	Essential for growth in M9 medium (Joyce *et al*., 2006)
cysQ	Essential for growth in M9 medium (Joyce *et al*., 2006)
fes	Essential for growth in M9 medium (Joyce *et al*., 2006)
gltA	Essential for growth in M9 medium (Joyce *et al*., 2006)
hisD	Essential for growth in M9 medium (Joyce *et al*., 2006)
ilvA	Essential for growth in M9 medium (Joyce *et al*., 2006)
ilvD	Essential for growth in M9 medium (Joyce *et al*., 2006)
leuB	Essential for growth in M9 medium (Joyce *et al*., 2006)
leuC	Essential for growth in M9 medium (Joyce *et al*., 2006)
leuD	Essential for growth in M9 medium (Joyce *et al*., 2006)
metB	Essential for growth in M9 medium (Joyce *et al*., 2006)
metC	Essential for growth in M9 medium (Joyce *et al*., 2006)
metF	Essential for growth in M9 medium (Joyce *et al*., 2006)
metR	Essential for growth in M9 medium (Joyce *et al*., 2006)
nadA	Essential for growth in M9 medium (Joyce *et al*., 2006)
pabB	Essential for growth in M9 medium (Joyce *et al*., 2006)
pdxA	Essential for growth in M9 medium (Joyce *et al*., 2006)
ppc	Essential for growth in M9 medium (Joyce *et al*., 2006)
proA	Essential for growth in M9 medium (Joyce *et al*., 2006)
proC	Essential for growth in M9 medium (Joyce *et al*., 2006)
purL	Essential for growth in M9 medium (Joyce *et al*., 2006)
thrA	Essential for growth in M9 medium (Joyce *et al*., 2006)
thrB	Essential for growth in M9 medium (Joyce *et al*., 2006)
thrC	Essential for growth in M9 medium (Joyce *et al*., 2006)
trpC	Essential for growth in M9 medium (Joyce *et al*., 2006)
trpD	Essential for growth in M9 medium (Joyce *et al*., 2006)
trpE	Essential for growth in M9 medium (Joyce *et al*., 2006)

Gene	Function
ompA	Cell envelope
rfaG	Cell envelope
sanA	Cell envelope
pepA	Central metabolism
pgi	Central metabolism
pgm	Central metabolism
copA	Metal and vitamin metabolism
cysG	Metal and vitamin metabolism
feoA	Metal and vitamin metabolism
trkA	Metal and vitamin metabolism
cvpA	Miscellaneous
cysW	Miscellaneous
purT	Miscellaneous
rapA	Miscellaneous
rhlB	Miscellaneous
speE	Miscellaneous
tolC	Miscellaneous
envC	Peptidoglycan synthesis and cell division
lpp	Peptidoglycan synthesis and cell division
minD	Peptidoglycan synthesis and cell division
pal	Peptidoglycan synthesis and cell division
tolB	Peptidoglycan synthesis and cell division
tolR	Peptidoglycan synthesis and cell division
dsbA	Protein processing
tatC	Protein processing
trxA	Protein processing
apaH	Regulation
barA	Regulation
cra	Regulation
csrD	Regulation
lysR	Regulation

Gene	Function
alaA	Amino acid metabolism
argG	Amino acid metabolism
gcvP	Amino acid metabolism
ilvE	Amino acid metabolism

Gene	Function
relA	Regulation
uvrY	Regulation

Table A15. Comparative analysis of a mutant library grown under three different anaerobic conditions.
Genes unique to glucose

Gene	Function
carA	Amino acid metabolism
cysH	Amino acid metabolism
cysI	Amino acid metabolism
glgB	Amino acid metabolism
leuL	Amino acid metabolism
cls	Cell envelope
ddlA	Cell envelope
ompR	Cell envelope
ackA	Central metabolism
acnB	Central metabolism
atpD	Central metabolism
gpmA	Central metabolism
ndh	Central metabolism
pfkA	Central metabolism
pykF	Central metabolism
recC	DNA
topA	DNA
gshA	Glutathione biosynthesis
gshB	Glutathione biosynthesis
gsiB	Glutathione biosynthesis
galU	Miscellaneous
mqsA	Miscellaneous
nadB	Miscellaneous
prc	Miscellaneous
prmB	Miscellaneous
speD	Miscellaneous
epmA	Miscellaneous
adhE	Mixed acid fermentation
ldhA	Mixed acid fermentation
pflB	Mixed acid fermentation
ftsK	Peptidoglycan synthesis and cell division
ftsN	Peptidoglycan synthesis and cell division
ldtB	Peptidoglycan synthesis and cell division
clpB	Protein processing
oppA	Protein processing
rpoS	Regulation
sixA	Regulation
uspA	Regulation
ycbJ	y genes
yfcE	y genes
ytfL	y genes
znuA	Zinc transport
znuB	Zinc transport
znuC	Zinc transport

Table A16. Comparative analysis of a mutant library grown under three different anaerobic conditions. Genes unique to glucose nitrate (not including y genes).

Gene	Function	Gene	Function
gcvT	Amino acid metabolism	*mgtA*	Miscellaneous
leuE	Amino acid metabolism	*kch*	Miscellaneous
leuO	Amino acid metabolism	*npr*	Miscellaneous
leuV	Amino acid metabolism	*insJ*	Miscellaneous
metW	Amino acid metabolism	*insK*	Miscellaneous
metQ	Amino acid metabolism	*gspC*	Miscellaneous
tdcB	Amino acid metabolism	*gltF*	Miscellaneous
tdcD	Amino acid metabolism	*mokA*	Miscellaneous
metI	Amino acid metabolism	*hokA*	Miscellaneous
argA	Amino acid metabolism	*shoB*	Miscellaneous
argB	Amino acid metabolism	*frc*	Miscellaneous
argD	Amino acid metabolism	*oxc*	Miscellaneous
argR	Amino acid metabolism	*arpA*	Miscellaneous
ompG	Cell envelope	*arpB*	Miscellaneous
asmA	Cell envelope	*casA*	Miscellaneous
mlaA	Cell envelope	*hlyE*	Miscellaneous
ompC	Cell envelope	*speA*	Miscellaneous
bamB	Cell envelope	*tfaP*	Miscellaneous
mlaB	Cell envelope	*trkG*	Miscellaneous
bamC	Cell envelope	*rna*	Miscellaneous
bamE	Cell envelope	*sseB*	Miscellaneous
mlaD	Cell envelope	*hsdS*	Miscellaneous
mlaE	Cell envelope	*tomB*	Miscellaneous
mlaF	Cell envelope	*opgG*	Osmolarity
cpxR	Cell envelope	*opgH*	Osmolarity
envR	Cell envelope	*proV*	Osmolarity
rcsF	Cell envelope	*proQ*	Osmolarity
rcsB	Cell envelope	*proP*	Osmolarity
rcsA	Cell envelope	*nagA*	Peptidoglycan synthesis and cell division
fabF	Cell envelope	*nagE*	Peptidoglycan synthesis and cell division
fabR	Cell envelope	*rodZ*	Peptidoglycan synthesis and cell division

Gene	Function
ftnA	Cell envelope
ompX	Cell envelope
plsX	Cell envelope
lipB	Cell envelope
dgkA	Cell envelope
phoE	Cell envelope
ais	Cell envelope
rfaJ	Cell envelope
rfaL	Cell envelope
rfaQ	Cell envelope
rfaS	Cell envelope
rfaY	Cell envelope
rfaZ	Cell envelope
rfbX	Cell envelope
rfc	Cell envelope
rseA	Cell envelope
ugd	Cell envelope
uidC	Cell envelope
waaU	Cell envelope
wbbK	Cell envelope
wbbL	Cell envelope
wcaE	Cell envelope
ompL	Cell envelope
wecH	Cell envelope
mlaC	Cell envelope
wzzE	Cell envelope
aceA	Central metabolism
aceB	Central metabolism
hyi	Central metabolism
cspA	Cold Shock
cspB	Cold Shock
cspE	Cold Shock
cspI	Cold Shock

Gene	Function
slt	Peptidoglycan synthesis and cell division
pinQ	Phage elements
queE	Phage elements
gtrB	Phage elements
gtrS	Phage elements
intQ	Phage elements
pspA	Phage elements
quuD	Phage elements
mcrC	Phage elements
mcrB	Phage elements
lit	Phage elements
mcrA	Phage elements
htpG	Protein processing
ibpA	Protein processing
pphB	Protein processing
prfC	Protein processing
sppA	Protein processing
skp	Protein processing
hdeA	Protein processing
tig	Protein processing
oxyR	Redox
ahpC	Redox
paoC	Redox
norR	Redox
hmp	Redox
adiY	Regulation
rssA	Regulation
alpA	Regulation
sspA	Regulation
appY	Regulation
stpA	Regulation
ariR	Regulation
cadC	Regulation

Gene	Function
cspG	Cold Shock
mutM	DNA
xerD	DNA
nrdD	DNA
tdk	DNA
pyrC	DNA
pyrD	DNA
pyrB	DNA
recA	DNA
hchA	DNA
uvrB	DNA
acrA	Drug efflux
acrB	Drug efflux
emrK	Drug efflux
emrY	Drug efflux
evgS	Drug efflux
mprA	Drug efflux
elfA	Fimbriae/biofilm
bdm	Fimbriae/biofilm
mcbR	Fimbriae/biofilm
csgB	Fimbriae/biofilm
pgaC	Fimbriae/biofilm
csgD	Fimbriae/biofilm
pgaA	Fimbriae/biofilm
pgaB	Fimbriae/biofilm
fimB	Fimbriae/biofilm
fimC	Fimbriae/biofilm
fimD	Fimbriae/biofilm
fimE	Fimbriae/biofilm
fimZ	Fimbriae/biofilm
flxA	Fimbriae/biofilm
htrE	Fimbriae/biofilm
sfmC	Fimbriae/biofilm

Gene	Function
dosC	Regulation
phnF	Regulation
phoQ	Regulation
pqqL	Translation
dusB	Translation
miaA	Translation
rimK	Translation
rlmG	Translation
rplI	Translation
rsmC	Translation
rsmI	Translation
typA	Translation
mglA	Use of alternative carbon sources
nanC	Use of alternative carbon sources
agaB	Use of alternative carbon sources
alsB	Use of alternative carbon sources
mtlA	Use of alternative carbon sources
bglG	Use of alternative carbon sources
nanM	Use of alternative carbon sources
bglH	Use of alternative carbon sources
rbsR	Use of alternative carbon sources
chbR	Use of alternative carbon sources
chbF	Use of alternative carbon sources
csrB	Use of alternative carbon sources
csrC	Use of alternative carbon sources
galE	Use of alternative carbon sources
galR	Use of alternative carbon sources
idnD	Use of alternative carbon sources
lacA	Use of alternative carbon sources
lacY	Use of alternative carbon sources
malQ	Use of alternative carbon sources
setC	Use of alternative carbon sources
uxaC	Use of alternative carbon sources

Gene	Function
sfmD	Fimbriae/biofilm
sfmH	Fimbriae/biofilm
bioA	Metal and vitamin metabolism
bioB	Metal and vitamin metabolism
bioC	Metal and vitamin metabolism
bioH	Metal and vitamin metabolism
fepB	Metal and vitamin metabolism
fepE	Metal and vitamin metabolism
fepG	Metal and vitamin metabolism
fhuB	Metal and vitamin metabolism
mocA	Metal and vitamin metabolism
panB	Metal and vitamin metabolism
panD	Metal and vitamin metabolism
pitA	Metal and vitamin metabolism
tonB	Metal and vitamin metabolism
zntA	Metal and vitamin metabolism

Gene	Function
xylE	Use of alternative carbon sources

Table A17. Comparative analysis of a mutant library grown under three different anaerobic conditions. Genes unique to glycerol nitrate.

Gene	Function
tyrA	Amino acid metabolism
metG	Amino acid metabolism
tdh	Amino acid metabolism
ilvG	Amino acid metabolism
brnQ	Amino acid metabolism
fbp	Central metabolism
mdh	Central metabolism
sucA	Central metabolism
rnhA	DNA
dam	DNA
deoB	DNA
pntA	DNA
dhaK	DNA
dhaL	DNA
glpF	Glycerol utilisation and uptake
glpK	Glycerol utilisation and uptake
glpD	Glycerol utilisation and uptake
dhaM	Glycerol utilisation and uptake
moaD	Metal and vitamin metabolism
mobA	Metal and vitamin metabolism
fau	Metal and vitamin metabolism
lon	Miscellaneous
pcnB	Miscellaneous
qseC	Miscellaneous
truB	Miscellaneous
rne	Miscellaneous
pdeK	Miscellaneous
sseA	Miscellaneous
wecE	Miscellaneous
rffC	Miscellaneous
rimP	Miscellaneous
dhaR	Regulation
srlR	Regulation
cytR	Regulation
ihfA	Regulation
ulaR	Regulation
yiaJ	Regulation
nuoA	Respiration
nuoC	Respiration
nuoH	Respiration
nuoN	Respiration
nuoL	Respiration
menD	Respiration
nuoF	Respiration
nuoI	Respiration
narX	Respiration
narG	Respiration
nuoB	Respiration
nuoG	Respiration
narH	Respiration
nuoM	Respiration
nuoE	Respiration
yciT	y genes
yhdH	y genes
yceD	y genes
ydjF	y genes

Table A18. Comparative analysis of a mutant library grown under three different anaerobic conditions. Genes shared by cultures grown in nitrate.

Gene	Function
argC	Amino acid metabolism
argE	Amino acid metabolism
aspC	Amino acid metabolism
gdhA	Amino acid metabolism
ilvB	Amino acid metabolism
ilvY	Amino acid metabolism
metA	Amino acid metabolism
fepA	Metal and vitamin metabolism
fepC	Metal and vitamin metabolism
fepD	Metal and vitamin metabolism
hemN	Metal and vitamin metabolism
moaA	Metal and vitamin metabolism
moaC	Metal and vitamin metabolism
moaE	Metal and vitamin metabolism
modA	Metal and vitamin metabolism
modB	Metal and vitamin metabolism
modC	Metal and vitamin metabolism
moeA	Metal and vitamin metabolism
moeB	Metal and vitamin metabolism
mog	Metal and vitamin metabolism
fnr	Regulation
nadR	Regulation
nagC	Regulation
narL	Regulation
ptsN	Regulation
rseB	Regulation
rssB	Regulation
cysA	Transporters
cysU	Transporters
narK	Transporters
purN	Miscellaneous
cspC	Miscellaneous
cyaA	Miscellaneous
fruK	Miscellaneous
rsmA	Miscellaneous
truA	Miscellaneous
yeiH	Miscellaneous
speC	Miscellaneous
clpX	Protein processing
clpP	Protein processing
fpr	Protein processing

Table A19. Comparative analysis of a mutant library grown under three different anaerobic conditions.

Y genes.

Putative y gene functional groups in glucose

Gene	Function
ycbJ	Putative phosphotransferase
yfcE	Phosphodiesterase
ytfL	Putative inner membrane protein

Putative y gene functional groups in glucose nitrate

Gene	Function	Gene	Function
yehA	Fimbriae/biofilm formation	*yahE*	Unknown/pseudogenes
yehB	Fimbriae/biofilm formation	*yahL*	Unknown/pseudogenes
yfcV	Fimbriae/biofilm formation	*ycdU*	Unknown/pseudogenes
yadC	Fimbriae/biofilm formation	*yecT*	Unknown/pseudogenes
yqiG	Fimbriae/biofilm formation	*ybeR*	Unknown/pseudogenes
ygiL	Fimbriae/biofilm formation	*ybeT*	Unknown/pseudogenes
yadK	Fimbriae/biofilm formation	*ybfA*	Unknown/pseudogenes
yqiH	Fimbriae/biofilm formation	*ybfK*	Unknown/pseudogenes
yadL	Fimbriae/biofilm formation	*ybfI*	Unknown/pseudogenes
yadM	Fimbriae/biofilm formation	*yedN*	Unknown/pseudogenes
yqiI	Fimbriae/biofilm formation	*yeeL*	Unknown/pseudogenes
yadN	Fimbriae/biofilm formation	*yiaW*	Unknown/pseudogenes
yqiK	Fimbriae/biofilm formation	*ygaQ*	Unknown/pseudogenes
yhcA	Fimbriae/biofilm formation	*ybfO*	Unknown/pseudogenes
ypjA	Fimbriae/biofilm formation	*yrhD*	Unknown/pseudogenes
ybfG	Fimbriae/biofilm formation	*ybhM*	Unknown/pseudogenes
ybgD	Fimbriae/biofilm formation	*yecF*	Unknown/pseudogenes
yceO	Fimbriae/biofilm formation	*yegJ*	Unknown/pseudogenes
ydeQ	Fimbriae/biofilm formation	*ykgH*	Unknown/pseudogenes
ydeR	Fimbriae/biofilm formation	*yfbM*	Unknown/pseudogenes
ydeS	Fimbriae/biofilm formation	*yfbN*	Unknown/pseudogenes
ydeT	Fimbriae/biofilm formation	*yfbO*	Unknown/pseudogenes
yraH	Fimbriae/biofilm formation	*ygcG*	Unknown/pseudogenes

Gene	Function
ybdO	Regulators
yeiL	Regulators
ygaV	Regulators
ycjW	Regulators
ydeH	Regulators
ydeO	Regulators
ydiP	Regulators
yjcC	Regulators
yeaI	Regulators
yeaJ	Regulators
yedV	Regulators
yedW	Regulators
yeeN	Regulators
yqeH	Regulators
yidL	Regulators
yiiE	Regulators
yjjQ	Regulators
ykgA	Regulators
yliE	Regulators
yliF	Regulators
ygeH	Regulators
yqeI	Regulators
yqjI	Regulators
yafT	Cell envelope
ybjX	Cell envelope
ycdM	Cell envelope
ycfZ	Cell envelope
yciB	Cell envelope
yhaI	Cell envelope
yhiD	Cell envelope
yiaA	Cell envelope
yiaB	Cell envelope
yidH	Cell envelope

Gene	Function
ygeQ	Unknown/pseudogenes
yhiL	Unknown/pseudogenes
yhiS	Unknown/pseudogenes
ycaW	Unknown/pseudogenes
yibJ	Unknown/pseudogenes
yibS	Unknown/pseudogenes
yihF	Unknown/pseudogenes
yjfI	Unknown/pseudogenes
yjiC	Unknown/pseudogenes
ymdA	Unknown/pseudogenes
yoeG	Unknown/pseudogenes
yahD	Unknown/pseudogenes
yaiS	Miscellaneous
yaiX	Miscellaneous
ybhH	Miscellaneous
ycaK	Miscellaneous
ycgF	Miscellaneous
yciF	Miscellaneous
ycjM	Miscellaneous
ydbD	Miscellaneous
yhgF	Miscellaneous
ydeM	Miscellaneous
ydeN	Miscellaneous
ydiF	Miscellaneous
ydiO	Miscellaneous
ydjH	Miscellaneous
ydjI	Miscellaneous
yebB	Miscellaneous
yrbL	Miscellaneous
yedL	Miscellaneous
yfbL	Miscellaneous
yfdE	Miscellaneous
yfiD	Miscellaneous

Gene	Function
yigG	Cell envelope
yiiG	Cell envelope
yjbM	Cell envelope
yjfL	Cell envelope
yjgN	Cell envelope
yjhB	Cell envelope
ynbB	Cell envelope
ytfN	Cell envelope
yagL	Phage
yagM	Phage
ybcK	Phage
ybcL	Phage
ybcM	Phage
yfjH	Phage
ybcY	Phage
ydfD	Phage
yfdK	Phage
yfdL	Phage
yfdQ	Phage
yfjI	Phage
yfjJ	Phage
yfjW	Phage
yjhC	Phage
yjhI	Phage
yjhR	Phage
ynaK	Phage
yahN	Transporters
ycjV	Transporters
ybbW	Transporters
yaiT	Transporters
ybbM	Transporters
ybbY	Transporters
ycgG	Transporters

Gene	Function
ygcB	Miscellaneous
ygcW	Miscellaneous
yhaB	Miscellaneous
yhbV	Miscellaneous
yhbX	Miscellaneous
ykgF	Miscellaneous
yhhK	Miscellaneous
yhiM	Miscellaneous
yihO	Miscellaneous
yieJ	Miscellaneous
yjjU	Miscellaneous
ykgB	Miscellaneous
ylbH	Miscellaneous
ynjI	Miscellaneous
yaeB	Miscellaneous
yjjK	Miscellaneous
yafP	Miscellaneous
yjdO	Miscellaneous

Gene	Function
ycgV	Transporters
yddA	Transporters
yddB	Transporters
ydfJ	Transporters
ydiM	Transporters
ydiN	Transporters
ydjE	Transporters
yfdV	Transporters
yfgM	Transporters
ygaZ	Transporters
ygfU	Transporters
yihN	Transporters
yigJ	Transporters
ynaI	Transporters
yphF	Transporters
yqeG	Transporters
yrbG	Transporters
yafD	DNA
yfcI	DNA
yhhZ	DNA
ykgC	DNA

Putative y gene functional groups in glycerol nitrate

Gene	Function
yciT	Putative DNA binding transcriptional regulator
yhdH	acrylyl-CoA reductase
yceD	DUF177 domain-containing protein YceD
ydjF	Putative DNA binding transcriptional regulator