Population Structure and Siderophore Production in Commensal *Escherichia coli*

Laura Jane Searle

A thesis submitted to the University of East Anglia (Norwich, UK) in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

In Biology

30th September 2015

Abstract

The factors involved in the diversity, transition and persistence of commensal *Escherichia coli*, both between and within hosts, are not fully understood. The aim of this project was to develop and use comparative approaches to investigate traits associated with colonisation of the gut by taking advantage of two collections of natural isolates of *E. coli*, the ECOR (host-associated) and GMB (plant-associated) collections. A new method for uniquely tagging and monitoring individual *E. coli* strains was developed to facilitate the assessment of large numbers of strains in complex environments, such as the gut. Competition studies between 'barcoded' strains show that this technique has a high sensitivity enabling the identification of *E. coli* present in the population at low levels, which may be undetected using currently available methods. Together with genome-wide association studies, barcoding is a powerful tool for identifying adaptive traits associated with the environment.

One trait linked to *E. coli* gut colonisation is production of the siderophore enterobactin. Comparisons of siderophore production and the distribution of siderophore loci between the GMB and ECOR collections highlighted that faecal strains produce higher levels of siderophores and possess a larger number of siderophore systems at the population level. However, while our epidemiological data indicated a role for siderophore biosynthesis in the GI-tract, we did not observe significant differences in the competitiveness of siderophore biosynthesis mutants in a mouse model. Whether this is linked to the ability of mutants to cheat and use siderophores generated by the wild-type strain or to obtain iron through other ways remains to be determined.

Overall, this study provides further details on the factors determining how the environment shapes the associated *E. coli* populations. This knowledge is essential to assess the relationship between the environment, the associated *E. coli* populations and the risk they represent to human health.

Author's declaration

I hereby certify that the work contained in this thesis is entirely the result of my own work, except where reference is made to other authors. It has not been submitted in any other form to the University of East Anglia or any other University.

Results in Chapters 4 and 5 were included in the following publication:

Searle, L.J., Méric, G., Porcelli, I., Sheppard, S.K., and Lucchini, S. (2015) 'Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *Escherichia coli*', *PLoS One* 10(3): e0117906.

Laura Searle

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Sacha Lucchini for his constant guidance, patience and encouragement, without which this work would not have been possible. I am also grateful to my supervisory committee, Nathalie Juge, Ida Porcelli, Udo Wegmann and Isabelle Hautefort, for all their support and advice throughout my PhD.

My thanks also go to my examiners Stephanie Schüller and Nicola Holden for a very enjoyable discussion about my work.

I would like to thank all my friends and colleagues at the Institute of Food Research and the University of East Anglia, for the encouragement and kindness given to me throughout my time here. To all the PhD students for the support and laughs, I'm sure I wouldn't have had such a great experience without you all.

Finally, a special thanks to my family, for all their love and support through both the good and stressful times during my PhD. Words cannot express how grateful I am to my mother, father and brother for all of the sacrifices that you've made on my behalf. Thank you.

Contents

List of figures

List of tables

1. Introduction

1.1 General introduction to *E. coli*

1.1.1 Escherichia coli

Escherichia coli is a Gram-negative, non-sporulating facultative anaerobe which is able to colonise the intestines of multiple different vertebrate hosts. Its primary host environment is in the large intestine of warm-blooded animals and reptiles, where it is most commonly found (Berg, 1996; Gordon and Cowling, 2003). The mucus layer that covers the intestinal epithelium is thought to be the primary niche of *E. coli*. Within the mucus *E. coli* forms part of the gut microbiota, which can consist of more than 500 species of microorganisms that can reach densities as high as 10^{11} CFU/g in the large intestine of humans (Berg, 1996). *E. coli* is one of the first bacteria to colonise the gut after birth, where it can reach high densities of over 10^{9} CFU/g in faeces (Penders et al., 2006), before the microbiota becomes dominated by anaerobic bacteria. In adults, the proportion of facultative anaerobes drops to only 0.01-1% of the total microbiota (Berg, 1996; Eckburg et al., 2005), of which *E. coli* is the most predominant. In humans it has been shown to be almost ubiquitous, appearing in over 90% of collected samples to date (Penders et al., 2006; Tenaillon et al., 2010).

E. coli is a versatile organism, able to occupy non-host secondary reservoirs, such as soil, plants and water, as well as colonising a wide variety of hosts. Under the assumption that *E. coli* populations die off rapidly, *E. coli* detection is frequently used as an indicator of recent faecal contamination events of water sources. However, recent studies have highlighted that *E. coli* is able to persist in the non-host environment in soils, waters and on food products for long periods, questioning its suitability in this practice (Desmarais et al., 2002; Hartz et al., 2008; Brennan, Abram, et al., 2010). The total population of *E. coli* in the wild is estimated to be 10^{20} cells (Whitman et al., 1998; Tenaillon et al., 2010), half of which are thought to be associated with non-host environments (Savageau, 1983).

Similarly to several other genera (*Salmonella*, *Yersinia*, *Klebsiella* and *Citrobacter*) in the *Enterobacteriaceae* family, to which *E. coli* belongs, *E. coli* is able to act as a pathogen. It is estimated that pathogenic strains of *E. coli* cause the deaths of more than two million humans across the world per year (Kosek et al., 2003; Russo and Johnson, 2003). In contrast, commensal strains of *E. coli* are able to acquire nutrients, shelter and/or movement without causing any harm or damage to their hosts (Boucher, 1985). As *E. coli* can be both a commensal and pathogen, it is very suitable to study the differences between pathogenesis and commensalism, as well as how bacteria switch between them. *E. coli* is one of the best studied and characterised model organisms, being important in both biotechnology and microbiology research, especially in the fields of genetics, molecular biology, cell physiology and biochemistry.

1.1.2 E. coli as a pathogen

Although most *E. coli* strains are harmless commensals, many isolates can be pathogenic and cause a wide variety of infectious diseases. Pathogenic *E. coli* can be split into two main groups based on the site of infection, with those strains causing disease in the GI-tract being called intestinal pathogenic *E. coli* (IPEC) and those that cause disease in other parts of the body are called extraintestinal pathogenic *E. coli* (ExPEC). ExPEC strains that infect humans are sub-divided depending on where the infection occurs and include uropathogenic *E. coli* (UPEC), neonatal meningitis-associated *E. coli* (NMEC) and sepsis-causing *E. coli* (SEPEC). IPEC strains are also divided into pathotypes; however, this is based on the phenotypic characteristics of infection and disease. Strains within each pathotype typically share a set of virulence factors which allow for pathotype identification. The main six pathotypes are enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and enteroaggregative *E. coli* (EAEC) (reviewed in Kaper et al., 2004; Croxen and Finlay 2010). *E. coli* causes more than two million deaths due to infant diarrhoea (Kotloff, 1999; Kosek et al., 2003) and extraintestinal infections (Russo and Johnson, 2003) per year. ExPEC infections are also

responsible for approximately 150 million cases of uncomplicated cystitis (Russo and Johnson, 2003).

1.1.2.1 IPEC

The virulence factors of IPEC strains include a wide range of traits that are predominantly involved in the attachment to host epithelial cells through fimbriae or pili, as well as the secretion of toxins. The extensively-studied model EHEC O157:H7 strain EDL933, for example, possesses the locus of enterocyte effacement (LEE) Pathogenicity Island which is required for the formation of attaching and effacing (A/E) lesions on epithelial cells (McDaniel et al., 1995; Perna et al., 1998). The LEE Pathogenicity Island is also found in EPEC strains (McDaniel et al., 1995) and encodes the adhesin intimin, a type 3 secretion system (T3SS), and several effector proteins (Garmendia et al., 2005; Croxen and Finlay, 2010). The effectors secreted by EPEC and EHEC strains into host epithelial cells are slightly different, with twice as many effectors identified in EHEC compared to EPEC (Tobe et al., 2006). The distinctive feature of EHEC strains, however, is the ability to produce Shiga toxin (Stx), which is encoded on a prophage. As no secretion system is present in EHEC, the Stx is released via phage-mediated lysis as part of the SOS response (Kimmitt et al., 2000; Toshima et al., 2007). Released Stx is able to spread systemically and bind globotriaosylceramide (Gb3) on kidney endothelial and epithelial cells, which can lead to fatal haemolytic uraemic syndrome (HUS).

ETEC strains use colonisation factors (CFs) for attachment to host cells, which can be fimbrial, non-fimbrial, helical or fibrillar (Turner et al., 2006). They modulate the gut environment to facilitate access to the epithelium by secreting enterotoxins. These bind to receptors on epithelial cells on the brush border which leads to impaired uptake of Na^+ and efflux of water into the lumen which disrupts the integrity of both the epithelium and mucus (Turner et al., 2006). EAEC strains possess the pAA plasmid, which encodes aggregative adherence fimbriae (AAFs) biosynthesis genes which are required for biofilm formation and aggregation on host cells (Croxen and Finlay, 2010). For the DAEC pathovar, host cell attachment is mediated through Afa-Dr adhesins which are able to damage

the intestinal brush border by altering the cell cytoskeleton (Bétis et al., 2003; Servin, 2005).

EIEC and *Shigella* are considered to be a single pathovar rather than different species, but the *Shigella* name has been retained because of the bacteria's association with the disease shigellosis (Croxen and Finlay, 2010). These strains are intracellular pathogens that acquired the pINV plasmid, which encodes a T3SS on the Mxi-Spa locus, as well as other genes which are required for cell invasion, intracellular survival and apoptosis (Schroeder and Hilbi, 2008). As well as this, large deletions are present in the EIEC and *Shigella* chromosome which correspond to the loss of key genes involved in amino acid and carbohydrate transport and nucleotide metabolism which are important in commensal strains (Maurelli et al., 1998; Touchon et al., 2009). This suggests that EIEC and *Shigella* are adopting an obligate host-associated intracellular evolutionary pathway (van Passel et al., 2008).

1.1.2.2 ExPEC

Some virulence factors have been identified in ExPEC strains that contribute to their pathogenicity, including adhesins, invasins, iron acquisition systems, toxins and protectins (Köhler and Dobrindt, 2011). However, many ExPEC strains cannot be clearly distinguished from commensal strains of *E. coli* based on the presence or absence of specific virulence factors. Unlike IPEC strains which must use specific virulence factors to cause disease, ExPEC strains are able to cause disease using a variety of virulence factors in multiple combinations, making them difficult to classify and distinguish.

It has been suggested that ExPEC strains are opportunistic pathogens, normally being a part of the healthy intestinal microbiota and living asymptomatically as commensals (Dobrindt et al., 2010; Tenaillon et al., 2010). However, when these strains gain access to areas outside the gut, such as the urinary tract, they are capable of causing infections (Yamamoto et al., 1997; Nielsen et al., 2014). In support of this hypothesis, a study of 148 human gut samples showed that of the three most common *E. coli* strains detected, two were UPEC strains, UTI89

and CFT073 (Gao et al., 2014). Virulence factors that increase pathogenicity in ExPEC strains can possibly be considered as fitness or colonisation factors that are broadly spread throughout commensal strains of *E. coli*. These factors may be able to persist within the commensal population as they increase fitness during the normal commensal lifestyle, either in the primary host gut environment, or in the secondary external environment (Köhler and Dobrindt, 2011). Several ExPEC virulence factors have been associated with increased fitness and successful colonisation of the GI-tract (Wold et al., 1992; Aslam and Service, 2006; Nowrouzian et al., 2006; Diard et al., 2010). Commensal strains of *E. coli* have been shown to possess virulence genes associated with ExPEC infections, such as strain HS, which possesses pili and fimbriae genes required for host cell attachment, as well as a type 2 and ETT2 type 3 secretion systems (Rasko et al., 2008).

It has been suggested that ExPEC strains can be divided into two groups, the first consisting of strains that possess the typical virulence genes associated with ExPEC infection. A second group may exist, however, which contains strains that do not have these ExPEC virulence factors, but are still capable of causing disease (ö and Dobrindt, 2011). Indeed, ExPEC infections in elderly or immunocompromised patients can be caused by strains with few virulence genes, suggesting that depending on the conditions, virulence factors do not play a large role in determining whether a strain will cause an infection, but rather that the strain is in the right place at the right time (Köhler and Dobrindt, 2011).

1.2 Evolutionary biology and population genetics of *E. coli*

1.2.1 The E. coli genome

The genome of *E. coli* is approximately 4.5 to 5.5Mb in size (Bergthorsson and Ochman, 1998) with the difference of up to 1Mb reflecting the large amount of genomic variability seen in *E. coli*. As examples, the K-12 commensal *E. coli* strain has a genome of size 4.6Mb, which encodes 4,405 genes, whereas the pathogenic *E. coli* strain O157:H7 EDL933 has 5,416 genes encoded on a 5.4Mb

genome (Perna et al., 2001). Unlike with eukaryotic genomes, there is a clear correlation between the size of the chromosome and the number of genes available, with coding genes being present at similar densities between strains (Mira et al., 2001). However, the gene repertoires of similar sized genomes in *E. coli* can vary widely (Welch et al., 2002; Willenbrock et al., 2007), suggesting that strains with larger genomes are not simply adding extra genes to a large core of essential genes. Interestingly, it has been observed that pathogenic strains typically possess a larger genome than commensals, possibly due to the addition of virulence genes or Pathogenicity Islands (Croxen and Finlay, 2010).

Studies have shown that the core genome, which consists of genes conserved in all strains, is indeed much smaller than the average genome size in *E. coli*, which is approximately 4,700 genes (Touchon et al., 2009). A study of 15, mostly pathogenic, *E. coli* genomes showed that there were 2,200 core genes and 13,000 genes in the pan-genome (Rasko et al., 2008). These numbers have since been revised by another study of 21 *E. coli* genomes to 1,976 core genes and a pan-genome of 17,838 (Touchon et al., 2009). A further study of an even larger number of *E. coli* strains revised the number of core genes conserved across 61 isolates of *E. coli* down to 993 (Lukjancenko et al., 2010). This clearly shows that a large proportion of the *E. coli* genome is variable, as on average only approximately 20% of the genome of a specific strain belongs to this core set of genes and more than 90% of the pan-genome is made up of accessory genes (Touchon et al., 2009; Lukjancenko et al., 2010). This also highlights how one strain of *E. coli* does not represent the species as a whole, as a single genome will only contain about 25% of the entire pan-genome (Touchon et al., 2009).

Genes that do not make up the core genome, known as accessory genes, can vary considerably in frequency, with some genes appearing in large numbers of strains (persistent genes), whereas some genes may only be present in one or two strains (volatile genes) (Touchon et al., 2009). The core and accessory genes make up the pan-genome of *E. coli*, which has been described as “open” due to the fact that as more studies are carried out, the number of genes discovered in the *E. coli* species increases. A large, open pan-genome is

generally observed in bacteria that occupy many environments, such as *Pseudomonads*, *Streptococci*, *Salmonellae* and *Streptomyces*, or environments that have changing conditions, as not only do bacteria benefit from the variety of genes that enables them to adapt to multiple different conditions, but they also have more opportunities to share genes with other bacteria (Medini et al., 2005; Tettelin et al., 2008). The opposite can be said for bacteria that are adapted to a specific niche which is not shared with other species of bacteria. The smallest bacterial genomes have been found mainly in symbiotic bacteria, which live within specialised host cells that provide a very stable environment and also reduce the need to produce nutrients provided by the host (McCutcheon and Moran, 2012). In contrast, the largest genomes have been found in *Ktedonobacter* spp. (Cavaletti et al., 2006; Chang et al., 2011) that reside in soil, a very complex environment (Tiedje et al., 1999).

1.2.2 Gene dynamics

It is important to understand the gene dynamics that give rise to such a high level of gene diversity in *E. coli* (figure 1.1). Variation in bacterial populations at the genetic level arises through mutations, which can alter gene function, delete genes, introduce new ones, duplicate them or have no effect from a fitness perspective. Several types of mutations can occur, including point mutations, deletions, insertions, duplications and recombination. Mutations can occur spontaneously, such as through naturally occurring lesions, by transposable genetic elements, or as a consequence of inaccuracies during DNA replication or repair. However, this rate of mutation has been shown to be as low as $1x10^{-3}$ mutations per genome per generation for wild-type *E. coli* grown under conditions with minimal natural selection (Wielgoss et al., 2011; Lee et al., 2012). Spontaneous genomic rearrangements (SGRs) and intrachromosomal recombination can alter gene regulation as well as disrupting or duplicating genes and have been shown to occur in both *E. coli* (Iguchi et al., 2006) as well as *Salmonella enterica* Typhimurium (Sun et al., 2012). New genetic material which can undergo recombination can also be introduced and lost from cells via

a process called horizontal gene transfer (HGT). Together these mechanisms lead to gene acquisition and loss from the *E. coli* genome.

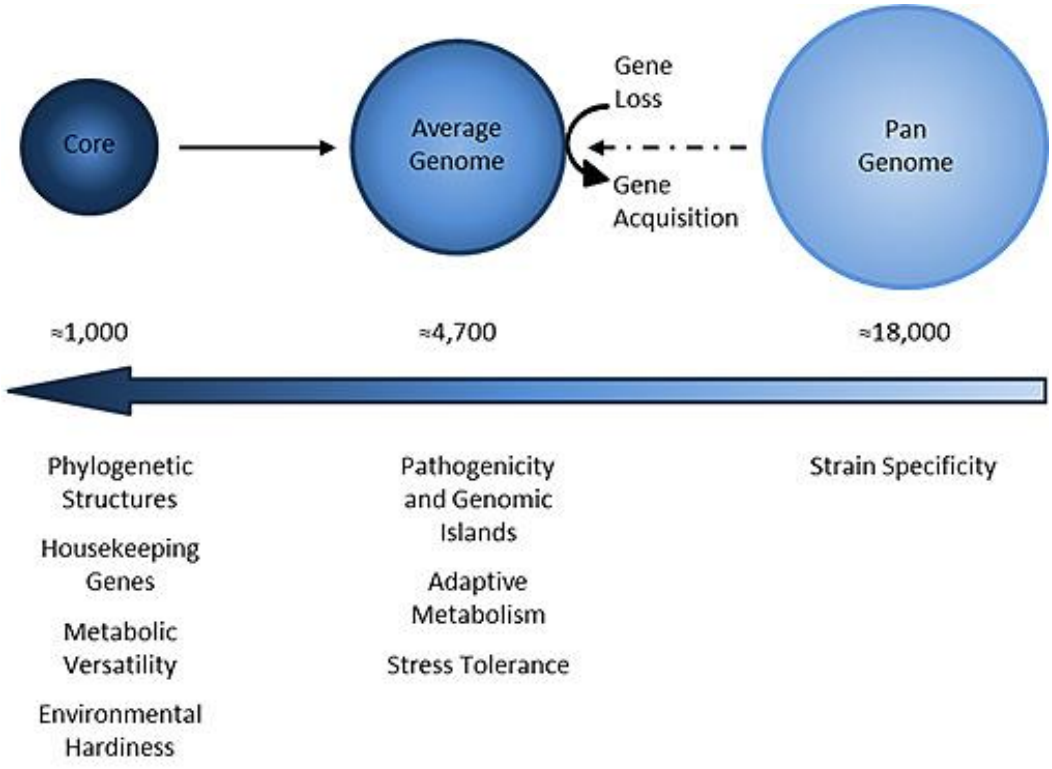


Figure 1.1: Gene dynamics in *E. coli*. The core genome of *E. coli* has been estimated to be of approximately 1,000 genes in size, which consists of key housekeeping and metabolic genes. The pan-genome consists of 18,000 gene families, of which approximately 10% are part of the core genome (Touchon et al., 2009; Lukjancenko et al., 2010). Evolutionary forces and pressures both within the gut and in the non-host environment constantly shape the genome within *E. coli* which undergoes gene acquisition, maintenance and loss (adapted from van Elsas et al., 2011).

1.2.2.1 Gene acquisition

Gene acquisition occurs through two mechanisms which are gene duplication and HGT. Gene duplication involves replicating a gene to form a homologous repeat, known as a paralog. This duplication followed by divergence occurs as a natural evolutionary response to environmental pressures and selection as it allows bacteria to adjust the amount of a specific protein being produced in a dosage type response to the environment. This can lead to increased fitness and a wider phenotype range for the species as a whole (Kondrashov, 2012). As the environment changes and the selection pressure is removed, these duplications have been shown to acquire mutations at a higher rate (Bergthorsson et al.,

2007). As such, they may develop different functions, but still belong to the same gene family (Hahn, 2009). These genes can then also undergo concerted evolution (Liao, 1999; 2000), where homologous genes within a species undergo homologous recombination and gene conversion, so they become identical for certain regions of the gene. The genes encoding ribosomal rRNA in bacteria are a good example of concerted evolution keeping diversity of paralogous genes low. Sequence divergence within the seven rRNA operons of *E. coli* is 0.195%, whereas the divergence between *E. coli* and the related Gammaproteobacterium *Haemophilus influenzae* is as high as 5.9% (Liao, 1999). However, not all genes undergo concerted evolution, with some genes being more biased towards gene conversion (Gevers et al., 2004; Serres et al., 2009). Alternatively, duplicated genes can become silenced or truncated through mutations to form pseudogenes, which can then be lost from the genome through gene erosion, as discussed later in this section.

The second process through which genes are acquired in *E. coli*, HGT, can occur through three mechanisms: conjugation, transformation and transduction. In all three of these mechanisms, genetic material is acquired from an external source. Conjugation occurs between two cells that make direct cell-to-cell contact, with a donor cell providing a mobile genetic element, often in the form of a plasmid or transposon, to a recipient cell. Transformation occurs as a result of competent cells taking up exogenous free DNA. Not all bacteria are able to carry out transformation naturally, though natural transformation has been described in *E. coli* (Baur et al., 1996; Sinha and Redfield, 2012). Competency can be induced by environmental stressors (Charpentier et al., 2012). Transduction occurs when DNA is injected into a bacterial cell by a bacteriophage. Mobile genetic elements often encode virulence genes, antibiotic resistance cassettes or colonisation-associated genes alongside the genes required for mobilisation, and as a result these traits can spread rapidly in *E. coli* populations.

The mobile genetic elements that are transferred into a cell will either remain in the cytoplasm as circular DNA or be integrated into the chromosome. The

homology between the mobile DNA and the chromosome influences the likelihood of integration into the chromosome, with DNA that does not share close homology more likely to be degraded (Shen and Huang, 1986; Thomas and Nielsen, 2005; Skippington and Ragan, 2011). This supports the observation that gene transfer is much rarer between distant taxonomic families than within or between species from the same genus (Skippington and Ragan, 2011). Homology can and does occur, however, at any of these taxonomic levels (Beiko et al., 2005; Toth et al., 2006). The physical proximity between two strains has been suggested as the most important factor for determining whether gene transfer will occur between two bacteria (Matte-Tailliez et al., 2002). It has been noted, however, that often the nearest bacteria are clones or closely related bacteria (Didelot and Maiden, 2010). Two main hypotheses have been suggested to explain how HGT evolved in bacteria. The first is that HGT is an important mechanism through which bacteria are able to repair the chromosome by using homologous external sources of DNA to replace possibly damaged areas (Vos and Didelot, 2009). The second hypothesis is that bacteria take up exogenous DNA primarily as source of nutrients, and gene transfer is a by-product (Redfield, 2001). Indeed, evidence suggests that there is a fitness benefit to metabolising DNA (Redfield, 1993; Finkel and Kolter, 2001; Palchevskiy and Finkel, 2006).

The role of HGT in *E. coli* genetics was originally thought to be quite minimal. Early multilocus enzyme electrophoresis (MLEE) studies, which characterised isolates based on differences in electrophoretic mobility of housekeeping enzymes, showed that there was a high level of linkage disequilibrium in *E. coli* populations (Selander and Levin, 1980; Whittam et al., 1983a; 1983b; Selander et al., 1986; Herzer et al., 1990). A high level of linkage disequilibrium, the non-random association of alleles at different loci, suggested that the proportions of allelic variants within the population were non-random, indicating that few recombination events occurred in *E. coli* populations. However, it has since been shown that DNA is transmitted laterally in *E. coli* through HGT, and a lot of the variation present is as a result of recombination events. The high linkage

disequilibrium could be explained by the fact that only certain regions of the chromosome are prone to recombination events, and large sections of the chromosome remain clonal (Smith et al., 1993).

HGT is important for the loss and gain of virulence factors in pathogenic *E. coli* strains, which are often transferred on Pathogenicity Islands (PAIs). PAIs are sections of a genome encoding virulence genes that are acquired through HGT. They are commonly flanked with mobile genetic elements enabling them to be readily transferred between bacteria. The large variety in virulence gene combinations arises through HGT, as seen especially in ExPEC strains, with certain combinations leading to particularly virulent strains. It was recently shown that the evolution of EHEC strains of *E. coli* was also non-linear, with several EHEC strains acquiring virulence genes independently (Ogura et al., 2009). The constant acquisition of new virulence genes is one important factor for the rapid emergence of epidemics caused by novel pathogens, as seen recently with EHEC and EIEC pathogenic strains (Wirth et al., 2006). The 2011 German outbreak strain, for example, was an EAEC pathovar that acquired the genes needed for Shiga toxin production, typically seen in EHEC and other Shiga toxin-producing *E. coli* (STEC) strains (Mellmann et al., 2011).

It is important to note that PAIs typically associated with pathogenic strains of *E. coli* can be functionally characterised as ecological, saprophytic or symbiotic as well as pathogenic (Hacker and Carniel, 2001). An example is the High-Pathogenicity Island (HPI) that contains the yersiniabactin locus. Yersiniabactin has been linked to virulence in UPEC strains (Brumbaugh et al., 2015), but has been found in many commensal strains of *E. coli*, albeit with an increased prevalence in two phylogenetically distinct groups. The recent rapid spread of the yersiniabactin locus through the *E. coli* population suggests that it plays a role in the commensal lifestyle of *E. coli* (van Elsas et al., 2011). The probiotic *E. coli* strain Nissle 1917 also raises a possibility for multiple siderophore systems as a benefit within the gut environment. Nissle 1917 is known to produce all four siderophore systems found in *E. coli* (Valdebenito et al., 2006), three of

which (aerobactin, yersiniabactin and salmochelin) are implicated in ExPEC pathogenesis (Torres et al., 2001; Schubert et al., 2002).

1.2.2.2 Gene loss

There are two mechanisms through which genes are lost from the genome, which are genetic erosion and gene deletion. Compared to eukaryotes the amount of "junk" (non-coding) DNA is very limited in bacteria. This suggests the presence of selective pressures to maintain a 'streamlined' genome. Important examples of gene loss are "black holes", or large deletions, that occur in some pathogenic bacteria that are adapting to new niches (Maurelli et al., 1998; Maurelli, 2007). For example, in *Shigella* and EIEC, a large region of DNA has been lost which includes the gene *cadA* which encodes lysine decarboxylase (LDC). LDC converts lysine to cadaverine which inhibits the activity of enterotoxins produced by *Shigella* and EIEC strains (Maurelli et al., 1998). Mutants where the *cadA* gene is reintroduced and LDC function restored have reduced virulence, supporting the hypothesis that gene loss is enabling these strains to evolve towards an intracellular pathogenic lifestyle (Maurelli et al., 1998). There is also evidence for the loss of genetic material turning a pathogenic strain into a commensal one. The *E. coli* strain 83972 was isolated from a patient with asymptomatic bacteriuria (ABU), where the bacteria colonised the urinary tract without causing symptoms (Hancock et al., 2008). This strain was shown to be phylogenetically related to other UPEC strains, but had undergone genome reduction events which possibly caused a loss of pathogenicity, and may signify continuing evolution towards host specialisation (Zdziarski et al., 2008).

To maintain the optimal gene complement as required by the environment, considering that bacteria are constantly acquiring new genes as described above, a "deletional bias" or an evolutionary selection for the loss of genes and genetic material through genetic erosion and subsequent removal of genes can be predicted (Mira et al., 2001). In gene erosion, a gene is truncated or becomes non-functional to form a pseudogene. These pseudogenes were originally considered to be neutral in terms of selection pressures and fitness (Li et al.,

1981). However, they are often removed from the genome at a rate that is greater than expected for the neutral model of stochastic loss, suggesting that they incur a cost and are actively removed from the genome (Lerat and Ochman, 2004; Kuo and Ochman, 2010). It has been suggested that these pseudogenes are under negative selection because the non-functioning proteins they encode are costly to make and may interfere with cellular processes (Kuo and Ochman, 2010). In agreement with this hypothesis, pseudogenes linked to fewer other genes, or with lower levels of protein-protein interactions have been shown to persist longer in the genome (Kuo and Ochman, 2010).

Genomic rearrangements, or recombination, can modify gene order (synteny) through inversions and translocations of large regions of DNA between two repeat sequences within the genome. Depending on the orientation of these two repeats, the type of recombination that occurs will be different. If the two repeats are facing opposite directions inversion will occur, whereas, if the repeats are facing the same direction the region of DNA between these two repeats will be excised and deleted. Both gene function as well as gene expression can be modified by recombination, as genes can become truncated or chimeric, or become relocated under a different promoter, resulting in new phenotypes. These inversions and rearrangement of the genome may be important for bacteria that live in multiple or changing environments as it may enhance diversity and provide a selective advantage (Ussery et al., 2004). In contrast, the bacterium *Buchnera aphidicola* is an obligate endosymbiont of aphids and analysis of its genome suggests that both chromosomal rearrangements and gene acquisitions are very rare events resulting in a very stable genome, probably as a reflection of its constant environment or a lack of interaction with other species of bacteria within the aphid (Tamas et al., 2002). The *E. coli* genome shows fewer large genetic rearrangements compared to other bacteria, such as *Shigella* (Tenaillon et al., 2010). The high rate observed in *Shigella* may be related to the recent adaptation to an intracellular lifestyle resulting in many genes becoming unnecessary or even detrimental to this new

lifestyle (van Passel et al., 2008). This is further highlighted by the high number of insertion sequence elements (Touchon et al., 2009).

Genome plasticity is likely to have limits, as while certain regions of the *E. coli* genome are hotspots for recombination, others display limited recombination, resulting in the genome maintaining a strong clonal structure (Touchon et al., 2009). The chromosome must maintain a certain amount of symmetry, with the origin of replication and the terminus being positioned opposite each other. Imbalances of more than 15% have been shown to affect *E. coli* cells detrimentally (Esnault et al., 2007). For this reason, recombination events do not usually occur between the left and right replichores across either the origin or the terminus (Blattner et al., 1997), as this would have a negative impact on chromosome replication (Esnault et al., 2007). Indeed, cells that possess chromosomes where the origin is adjacent to the terminus often produce anucleated daughter cells (Niki et al., 2000). The chromosome is also structured into macrodomains, which locate to specific locations within the cytoplasm during the cell cycle and are involved with chromosome localisation (Niki et al., 2000). Disruption of the synteny of certain parts of the chromosome during recombination may therefore have an impact on cell division (Niki et al., 2000).

1.2.2.3 Selection and gene maintenance

Once a gene has been added, lost or changed, new selection pressures arise for the organism based on the new phenotype they cause. This selection can be positive, negative or neutral. Beneficial mutations that arise do not always become dominant within a population and may be lost to stochastic genetic drift. In populations where strong selection pressures exist, clonal interference can be important in determining which mutations become fixed. This is where beneficial mutations are competing with one another, leading to fluctuations in which bacteria are dominant based on which mutant confers the greatest increase in fitness (Rozen et al., 2002). As a result, mutations with the largest beneficial effects are more likely to become fixed in a population, even though these mutations may be rarer than others (Fogle et al., 2008).

The effect of a mutation can be influenced by the genetic background of the bacterium in which it arises, a phenomenon known as epistasis. Positive epistasis results in an increase in fitness greater than expected and negative epistasis causes a lower fitness than expected (Phillips, 2008). High levels of positive epistasis have been observed in *E. coli*, especially in essential genes, though this may be down to essential genes causing large disturbances in non-associated pathways (He et al., 2010). Sign epistasis occurs when multiple mutations individually have one effect on fitness, which when combined together result in the opposite effect occurring (Weinreich et al., 2005). Compensatory mutations can arise, which remove deleterious effects of earlier mutations, allowing them to persist within the population (Kimura, 1990).

1.2.3 Population structure and phylogeny of E. coli

Early studies on population structure in *E. coli* used MLEE studies which demonstrated that there were clear groups of *E. coli* isolates based on their electrophoretic profiles. From these MLEE profiles, the ECOR reference collection (consisting of 72 strains of *E. coli*) was created to represent the full genetic diversity of the *E. coli* species as a whole (Ochman and Selander, 1984). Phylogenetic studies using MLEE profiles identified four major clades, or phylogenetic groups, of *E. coli*, called A, B1, B2 and D (Selander and Levin, 1980; Selander et al., 1986; Herzer et al., 1990) along with two accessory groups C and E. Multilocus sequence typing (MLST) techniques that emerged alongside DNA sequencing used the nucleotide sequences of multiple genes to determine allelic profiles and sequence types (Lecointre et al., 1998; Escobar-Páramo, Sabbagh, et al., 2004). Analysis using MLST supported the MLEE evidence for phylogroups A, B1, B2, D and E (Reid et al., 2000; Escobar-Páramo, Sabbagh, et al., 2004; Johnson et al., 2006; Wirth et al., 2006; Gordon et al., 2008). However, no evidence was found for the accessory group C, which was later split between the D and E groups (Herzer et al., 1990; Wirth et al., 2006). Group F has since been identified as a subgroup within the phylogenetic clade D (Jaureguy et al., 2008).

The order in which these phylogroups emerged is unclear, with different phylogenetic studies yielding different evolutionary trees. It was initially determined that phylogroups B2 and D were the oldest, with A and B1 diverging later (Lecointre et al., 1998; Wirth et al., 2006). Some studies have placed the B2 group as the first to emerge (Lecointre et al., 1998; Tenaillon et al., 2010; Sims and Kim, 2011), whereas a whole-genome phylogeny study has suggested that the earliest group was D, followed by B2 (Touchon et al., 2009). It was observed that using *S. enterica* as the reference group (outgroup) for determining relatedness between *E. coli* strains resulted in long-branch attraction (Felstenstein, 1978; Touchon et al., 2009). This led to distantly related lineages being incorrectly identified as closely related solely due to the amount of evolutionary changes within each lineage. However, replacement of *S. enterica* with *Escherichia furgusonii*, the closest relative of *E. coli* (Lawrence et al., 1991), reduced long-branched attraction and supported the D group as the most ancient group (Touchon et al., 2009).

Reconstruction of phylogenetic trees is also made more difficult by frequent recombination and gene acquisition or loss (figure 1.2). Phylogenetic trees for different genes can yield conflicting results, called phylogenetic incongruences, depending not only on real differences in evolutionary processes, but also on which region of DNA is used to construct the tree. This is especially problematic when DNA inside and around recombination integration sites is analysed (Dykhuizen and Green, 1991). For this reason, phylogenetic trees are often constructed from multiple genes or loci using MLEE, MLST or whole genome sequencing rather than single genes. Indeed, increasing the DNA fragment used for phylogenetic tree construction decreases the interference of recombination sites as the recombined area becomes a smaller proportion of the analysed DNA (Treangen et al., 2014). There are some exceptions, however, with some genes possessing enough sequence diversity to reflect the evolution of the whole *E. coli* population (Lescat et al., 2009; Sankar et al., 2009).

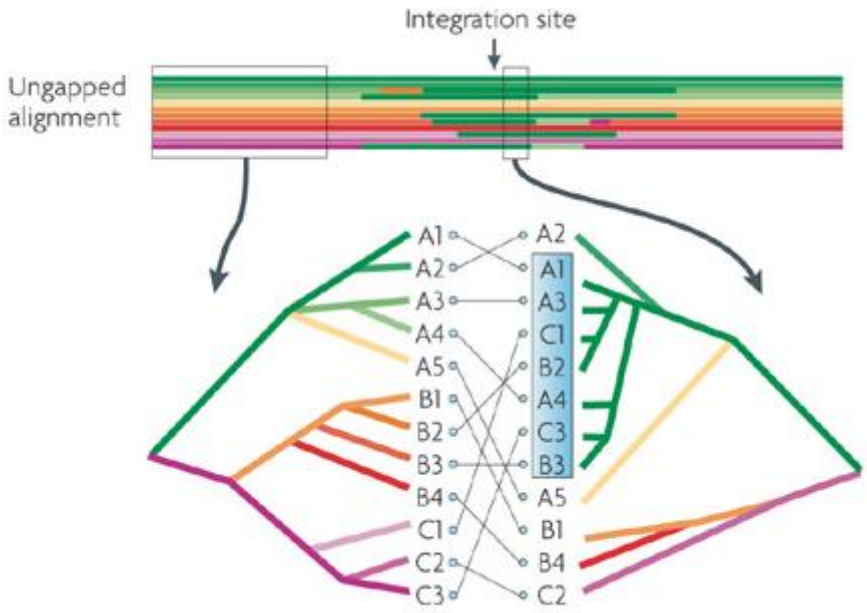


Figure 1.2: Effect of recombination on phylogenetic reconstruction. At the top, a sequence alignment of 12 *E. coli* isolates from three phylogenetic groups (A, B and C) highlights an integration site for recombination. Phylogenic trees reconstructed using the integration site and those produced using the ungapped alignment where recombination has not occurred are therefore incongruent (Tenaillon et al., 2010). This figure is copyrighted by MacMillan Publishers Ltd.

Following extensive sequence comparison it has recently been proposed that significant genetic exchange occurs, or has occurred, in *E. coli* between different phylogenetic groups (Leopold et al., 2011). Some phylogenetic groups apparently exchange DNA more than others (Leopold et al., 2011). The reason for this could either be due to different strategies with recombination for each phylogenetic group, or it could just be due to spatial or location differences (Leopold et al., 2011). Indeed, the proportion of each phylogenetic group in a population differs depending on the environment. The dominant phylogenetic group differs not only between the host and non-host environments, but also between hosts. This has largely been associated with differences in gut physiology between different hosts, as well as external environmental conditions which may limit *E. coli* survival or host exposure (Gordon and Cowling, 2003).

The fact that the phylogenetic groups A and B1 are isolated from a greater variety of host animals (Gordon and Cowling, 2003) and group B1 is also isolated

in greater numbers from the non-host environment (Gordon and Cowling, 2003; Walk et al., 2007; Méric et al., 2013) suggests that these groups are "generalists". In comparison, the B2 and D groups are generally limited to endothermic vertebrates, showing a greater level of host specialisation (Gordon and Cowling, 2003). However, it has been observed that group A strains harbour phenotypical traits that are associated with a host-specific lifestyle, including a reduction in both extracellular matrix and RpoS production (White et al., 2011; Méric et al., 2013). RpoS, the alternative sigma factor, is a central regulator of the *E. coli* General Stress Response, which includes extracellular matrix formation (Hengge-Aronis, 2002). For host generalists, a functional RpoS and stress response system has been proposed to be essential for adaptation to a variety of environmental stresses found outside the host, or in multiple different hosts (Ferenci and Spira, 2007; Peterson, 2005). This may explain the observed decreased prevalence of RpoS positive strains in host specialist phylogenetic groups A and B2 (White, 2011). An interesting observation is that strains from phylogenetic group A have some of the smallest genomes in *E. coli* (Bergthorsson and Ochman, 1998) and that their genomes have relatively few accessory genes compared to other groups (Sims and Kim, 2011). This has led to the hypothesis that group A strains are displaying "commensal minimalism", the loss of genes not associated with host commensalism.

In humans, data compiled from 1,117 individuals showed that the overall distribution of the phylogenetic groups was 40.5% group A, 25.5% B2, 17% B1 and 17% D (Tenaillon et al., 2010). However, looking at individual studies, the dominant group varies depending on geography. Samples collected from Africa (Mali and Benin), Asia (Pakistan) and South America (French Guiana, Colombia and Bolivia) all showed that group A was dominant (Duriez et al., 2001; Escobar-Páramo, Grenet, et al., 2004; Pallecchi et al., 2007; Nowrouzian et al., 2009). In contrast, samples from Europe (Sweden and France), North America (USA), Japan and Australia all showed that the B2 phylogenetic group was predominant (Obata-Yasuoka et al., 2002; Zhang et al., 2002; Escobar-Páramo, Grenet, et al., 2004; Nowrouzian et al., 2005; Gordon et al., 2015). Interestingly, samples

collected from Europe (Sweden, France and Croatia) in the 1970s and 1980s showed similar phylogenetic profiles as those from Africa, Asia and South America, suggesting that socioeconomic factors such as hygiene and diet are more influential than geography (Duriez et al., 2001; Escobar-Páramo, Grenet, et al., 2004). In comparison, from 1,154 animal samples, the distribution of phylogenetic groups was 41% B1, 22% A, 21% B2 and 16% D (Tenaillon et al., 2010). Domestication, however, greatly impacted on the distribution with domesticated animals (farm and zoo) having lower levels of B2 and higher levels of A group strains (Tenaillon et al., 2010).

Longitudinal studies in humans were carried out in Sweden that differentiated between *E. coli* strains that persisted long-term within the gut (resident strains) and those that were only detected for a short period of a few days or weeks (transient strains). They showed that the B2 phylogenetic group was more likely to be a resident strain (Nowrouzian et al., 2005; Nowrouzian et al., 2006). Importantly, the dominant phylogenetic group of the resident strain for the 1970s study was the B2 group, although the A group was most prevalent (Nowrouzian et al., 2006). This highlights the possibility that different phylogenetic groups may adapt different strategies towards gut commensalism.

Serotyping was used classically to identify IPEC strains as they typically belong to a small number of O:H serotypes (Karch et al., 2005; Stenutz et al., 2006). IPEC strains are found in most phylogenetic groups within *E. coli*, though they are found to a lesser extent within group A. Although some highly virulent ExPEC strains do belong to a small number of serotypes, as with IPEC, many strains cluster with commensal strains (Smith et al., 2007). MLEE has shown that the majority of ExPEC strains belong to the B2 phylogenetic group, and to some extent the D group (Selander and Levin, 1980; Boyd and Hartl, 1998), but this was not sufficient to split the ExPEC strains from commensal strains. MLST again confirmed that the B2 phylogenetic group includes the majority of ExPEC strains (Köhler and Dobrindt, 2011). There were some allelic profiles, known as sequence types (ST), which only contained ExPEC strains and highlighted clonal lineages that have the potential to be ExPEC-associated. However, in general,

ExPEC could not be distinguished from commensal *E. coli* strains (Köhler and Dobrindt, 2011). It has also been shown that the B2 group is, however, not homogeneous, in that there are possibly sub-groups of the B2 group that possess fewer virulence factors and may have a lower potential to cause disease (Le Gall et al., 2007). An *in vivo* mouse model of ExPEC virulence showed that phylogenetic groups A and B1 were largely unable to cause infections (Picard et al., 1999). Large numbers of strains from the B2 and D groups, however, were able to cause infection and death, suggesting that the genetic background may be important for shaping virulence (Picard et al., 1999).

1.3 The Ecology of *E. coli*

1.3.1 The gut (primary) environment

The primary niche of *E. coli* is in the large intestines of warm-blooded animals and reptiles. In humans, the facultative anaerobe *E. coli* is usually one of the first bacteria to colonise the gut after birth, but its abundance decreases over time as the gut environment becomes more anaerobic as a result of oxygen consumption by colonising bacteria. Eventually the microbiota becomes dominated by anaerobic bacteria (Matamoros et al., 2013). Despite this shift in microbiota, *E. coli* can usually persist at lower levels and can be detected in approximately 90% of humans sampled (Penders et al., 2006; Tenaillon et al., 2010). Within the gastrointestinal (GI) tract the generation time for *E. coli* has been approximated to between 40 and 120 mins (Poulsen et al., 1994; Poulsen et al., 1995), which is much longer than that seen *in vitro* of about 30mins when grown on intestinal mucus (Licht et al., 1999). This reduced growth reflects that within the gut *E. coli* must compete with a large variety of other bacteria for nutrients as well as cope with any host responses. As a result, the majority of *E. coli* ingested are thought to quickly transit through the GI-tract without any negative impact on the host or stably colonising the gut (Caugant et al., 1981; Savageau, 1983). Many of these strains will only be detectable for a few days or weeks before they are released back into the external environment and are known as transients. However, some are capable of persisting for longer periods

of time, up to several years, and are called residents (Wallick and Stuart, 1943; Sears et al., 1950; Cooke et al., 1972; Touchon et al., 2009; Tenaillon et al., 2010).

Early studies on human faecal samples using MLEE indicated that one strain of *E. coli* typically constitutes more than 50% of the colonies isolated from a single individual. This dominant strain tends to also be a long-term resident of the microbiota (Caugant et al., 1981). A study looking at *E. coli* diversity in the human GI-tract found that a single host typically has 1-4 strains (or genotypes) with a diversity score (Simpson's Index) of 1.74, with 85% of individuals possessing two or more strains (Gordon et al., 2015). Of those possessing more than two strains, 85% had strains only from one or two phylogenetic groups, with the second most abundant strain being significantly more likely to be from the same phylogenetic group as the dominant strain. Also, the dominant phylogenetic groups also influenced the distribution of different strains within the lower intestines, with groups E and F resulting in greater heterogeneity and group B2 resulting in the least (Gordon et al., 2015).

The location of *E. coli* within the gut is still debated, but because *E. coli* is detected in the colon at 1,000 times the levels found in the ileum, the colon is considered to be its primary niche (Savageau, 1974). Commensal *E. coli* strains have been isolated from all areas of the lower intestinal tract in humans (ileum, colon and rectum), but with variability in diversity depending on location (Gordon et al., 2015). However, *in vitro* plug flow and chemostat cultures have suggested that *E. coli* is more adapted to the small intestine (Koch, 1987) and increased colonisation of the ileum by some strains of *E. coli* has been observed (Staley et al., 1969; Aktan et al., 2007; Barnich and Darfeuille-Michaud, 2007). Several pathogenic *E. coli*, including EPEC, ETEC, DAEC and some EAEC strains, have also been shown to predominantly colonise and cause disease in the small intestine (Kaper et al., 2004; Croxen and Finlay, 2010; Okhuysen and Dupont, 2010). Other pathotypes, including EHEC and some EAEC strains, are able to attach to epithelium in the large and small intestine and EIEC strains colonise the large intestine only (Phillips et al., 2000; Croxen and Finlay, 2010; Okhuysen

and Dupont, 2010; Lewis et al., 2015). However, the isolation of *E. coli* from different areas of the GI-tract does not necessarily equate to colonisation throughout the gut. It is possible that *E. coli* are able to attach to the mucus or epithelium of a particular region of the gut without establishing a growing and persistent population (colonisation), making it more difficult to determine exactly where in the gut the primary niche of *E. coli* is.

1.3.1.1 Gut physiology

The gut is structured both laterally and longitudinally, with the environment in each area of the gut being very different and presenting different challenges to ingested bacteria. Examples include differences in food transit time, nutrient concentrations and availability, pH and bile, oxygen availability and changes in mucus and epithelial cells. After ingestion bacteria must survive transit through the oesophagus and stomach. Once within the lower intestinal tract, to cope with the constant flow of material within the gut, it is important for *E. coli* to form an attachment to the gut lining in order to prevent being washed out of the GI-tract. Mucus is secreted by goblet cells in the intestine and forms a layer covering the gut epithelium. Commensal *E. coli* strains are thought to colonise the gut by forming microcolonies within this mucus layer. In fact, *E. coli* in the lumen do not appear to grow compared to *E. coli* that have managed to colonise the mucus layer (Poulsen et al., 1995). This could be related to the difference in persistence of transient and resident strains of *E. coli*, with transient strains unable to colonise the mucus and passing quickly through the lumen where they do not grow, resulting in them only being present for a short amount of time. Indeed, the community of bacteria within the mucus has been shown to significantly differ from that recovered from the faeces (Zoetendal et al., 2002), with a higher diversity of *E. coli* being detected from intestinal samples (Gordon et al., 2015).

The mucus in the colon can be differentiated into two layers; the inner layer which makes direct contact with the epithelium and the outer layer which extends into the lumen (Atuma et al., 2001). Colonisation by the microbiota and *E. coli* is thought to be primarily in the outer layer of the mucus. Mucus consists

of approximately 98% water and the heavily glycosylated mucin proteins, primarily MUC2 in the human GI-tract (Hansson, 2012). The mucin is much tighter in the inner layer so that pore sizes are too small for bacteria to enter through without having to degrade it (Johansson et al., 2008; Johansson et al., 2010; Johansson et al., 2011). The expansion of the mucin, and thus enlargement of the pore sizes in the outer mucus layer, allows bacteria to enter and colonise (Johansson et al., 2011). The inner layer is thicker in the proximal colon, compared to the distal colon, and as such is partially penetrable to bacteria (Ermund et al., 2013). However, within the inner mucus layer digestive enzymes and antimicrobial peptides are also secreted by epithelial cells, as well as Paneth cells in the small intestine, making it difficult for bacteria to survive. As a result of small pore size and antimicrobial secretion the inner layer of mucus is largely devoid of bacteria (Hansson, 2012). Dextran sulphate (DSS) models of bacterial invasion of the inner mucus layer have shown a breakdown of mucin which allows bacteria to enter the inner layer and access the epithelium underneath resulting in a strong induction of the intestinal inflammatory response (Johansson et al., 2010).

The structure of the mucus layer within the small intestine is less clear, but recent studies have suggested that although it is predominantly loose and easily penetrated by bacteria, a thin tight layer is directly attached to the epithelium, which prevents bacteria from attaching to host cells (Ermund et al., 2013; Bajka et al., 2015). However, gaps in the mucus may exist which allow penetration of bacteria through the mucus layer and direct contact with the epithelium (Bajka et al., 2015). The reason for this is probably related to the fact that most absorption of nutrients by the host occurs in the small intestine, so having a thinner and penetrable mucus layer would facilitate nutrient uptake.

1.3.1.2 Attachment and motility

Commensal bacterial adhesins are able to interact with and bind to host O-glycans found within mucus, facilitating attachment of *E. coli* (Hansson, 2012). It is important to note that mucus is constantly being shed into the lumen with a turn-over of 1h. Therefore, although *E. coli* can attach to the mucus, washout

still occurs. Possession of P-fimbriae and the K5 capsular polysaccharide have been linked to increased colonisation of *E. coli* in the colon of gnotobiotic rats (Herías et al., 1995; Herías et al., 1997). Both of these were also shown to be more prevalent in long persisting *E. coli* strains compared to transient strains (Nowrouzian et al. 2001a; 2001b; Nowrouzian et al., 2003). Interestingly, P-fimbriae are involved in urinary tract infections, and alongside type 1 fimbriae, mediate adherence to the epithelium (Melican et al., 2011). Type 1 fimbriae have been shown to be more prevalent in resident strains (Nowrouzian et al., 2003). The K5 capsule has also been associated with ExPEC infections (Gransden et al., 1990) and the interactions of *E. coli* strain Nissle 1917 with host epithelial cells and maintenance of gut health (Hafez et al., 2009; Nzakizwanayo et al., 2015).

S-fimbriated *E. coli* adhere to mucus glycoproteins extracted from faeces better than other strains of *E. coli* (Tuomola et al., 1999), and have been shown to be important for attaching to host colonic epithelial cells (Adlerberth et al., 1995), but the relevance of this for commensal strains in a healthy gut is unclear. Indeed, S-fimbriae knockout mutants do not show reduced colonisation in streptomycin treated rats (Herías et al., 2001). S-fimbriae were shown to be present at equal levels in transient and resident strains of *E. coli*, suggesting a limited role in long-term persistence in the GI-tract (Nowrouzian et al., 2001a). S-fimbriae may instead be virulence factors that are implicated in ExPEC meningitis in children (Stins et al., 1994) as well as being found in UPEC strains (Herías et al., 1995). The repertoire of adherence factors possessed by pathogenic *E. coli* is important in defining the pathotype, with each pathotype having its own characteristic mechanisms for host cell attachment. For example, EPEC strains possess bundle-forming pili (BFP) and intimin for attachment to epithelial cells in the small intestine, whereas UPEC strains utilise type 1 fimbriae and P-fimbriae to attach to the uroepithelium (Croxen and Finlay, 2010).

Pathogenic bacteria that attach to the epithelium are generally able to breach the mucus layer via several mechanisms, including penetration of the mucus,

secretion of mucus degrading enzymes, evading the mucus via attachment to intestinal microfold cells (M-cells) and production of toxins that disrupt mucus producing cells. It is not clear, however, if commensal bacteria are able to form close attachments to the epithelium. Many commensal *E. coli* strains are able to produce adhesins used by pathogenic strains to adhere to intestinal epithelial cells. This includes CFA/I produced by ETEC strains (Tullus et al., 1992; Herías et al., 1995) and the *E. coli* common pilus (ECP) used by many pathogenic *E. coli*, including EHEC O157:H7 strain EDL933 (Rendón et al., 2007).

Flagella are also thought to be involved in intestinal cell adherence by some strains of *E. coli*, including the probiotic strain Nissle 1917 (Haiko and Westerlund-Wikstrom, 2013). However, within the gut, non-motile mutations quickly occur in successful *E. coli* colonisers, with the GI-tract environment selecting against expression of flagella. Motility has been shown to be lost in a few days after gut colonisation in mice (Gauger et al., 2007; De Paepe et al., 2011). Non-motile mutants of the EHEC O157:H7 strain EDL932 had increased colonisation of the cattle GI-tract compared to the wild-type (Dobbin et al., 2006). However, another study showed that the H7 flagella of EHEC O157:H7 were involved in attachment to the bovine intestinal epithelium (Mahajan et al., 2009). A role for flagella in the EHEC EDL933 and EPEC E2348/69 strains in adhering to bovine mucus has also been observed (Erdem et al., 2007). Therefore, it is possible that flagella may be involved in moving to the site of attachment and initial adherence. Flagella are probably required by *Salmonella* and pathogenic *E. coli* to swim against the flow of mucus being shed into the lumen, so that it can reach and adhere to epithelial cells (Stecher et al., 2004). This is supported by the observation that flagella and adhesion molecules are not usually concomitantly expressed, but bacteria switch from one to the other as it adapts to either a motile or sessile lifestyle (Haiko and Westerlund-Wikstrom, 2013). Studies have shown that reduced motility can arise though mutations in either the EnvZ/OmpR two-component transduction system (Giraud et al., 2008) or the *flhDC* operon which encodes the flagella master regulator (Dobbin et al., 2006; De Paepe et al., 2011). The presence of flagella is

associated with reduced colonisation ability, most likely because of energy expenditure on flagella construction and function (Giraud et al., 2008). Expression of the flagella genes may also affect colonisation through other mechanisms as many genes not involved in flagella synthesis and function are down regulated in flagella mutant strains (Zhao et al., 2007). Flagellin is very immunogenic and stimulates cytokine production by the innate immune system (Ciacci-Wollwine et al., 1998; Hayashi et al., 2001; Moors et al., 2001). It has therefore been proposed that *E. coli* modulates flagella production to prevent recognition by the immune system (Kim et al., 2012; Cullender et al., 2013). Mutations in the EnvZ/OmpR system also reduce cell permeability, which may increase resistance to bile salts or other stressors in the gut (De Paepe et al., 2011).

1.3.1.3 Stressors in the GI-tract

To be able to colonise the GI-tract, bacteria must be able to adapt to stresses that are present in the gut environment. Two of the main stressors which bacteria must overcome to survive within the gut are gastric acidity and the activity of bile salts. After ingestion of food, stomach pH will decrease to a pH of approximately 2 as a result of the secretion of gastric acid, which contains hydrochloric acid. Bacteria that transmit between hosts through the faecal-oral route, such as *E. coli*, must survive transit in the stomach to reach the lower intestines where they colonise. There are several acid resistance systems that have been described in *E. coli* which generally involve the removal of protons from the cell to maintain intracellular pH levels (Foster, 2004). Acid resistance system 1 (AR1) is regulated through the alternative sigma factor *rpoS* and as a result is linked to the stationary growth phase (Arnold and Kaspar, 1995). The exact mechanism of AR1 is not fully understood, but an equivalent system in *Streptococcus* has been shown to actively pump protons across the cell membrane (Martin-Galiano et al., 2001).

The influx of protons into the cell not only influences pH, but also affects the electrical membrane potential of the cell by introducing positive charge. When *E. coli* is grown at pH7, its membrane potential is -50mV, but this can rise up to

80mV in acid resistant strains (Richard and Foster, 2004). This strategy is used by many acidophilic bacteria (Matin, 1999) and reduces damage by slowing down movement of protons into the cell (Matin et al., 1982). The AR2 and AR3 systems are both able to contribute towards this increase of membrane potential. These systems involve the combined activity of amino acid decarboxylases and antiporters. Decarboxylases are able to "consume" a proton by replacing the carboxyl group on an amino acid substrate with a proton. The product, however, still retains the charge from the proton, so the cell is able to maintain a high positive charge while also removing protons (Foster, 2004). The AR2 and AR3 have also been suggested to have a role in intracellular pH homeostasis (Foster, 2004).

As well as hydrochloric acid, gastric acid also contains enzymes that carry out proteolysis, which has been shown to be important in bacterial killing (Zhu et al., 2006). Indeed, studies have shown that acid resistant bacteria are able to survive at approximately pH 2 for several hours, but this is not frequently reached or sustained for long periods in the stomach environment (Gorden and Small, 1993; Small et al., 1994; Zhu et al., 2006). The stomach enzyme pepsin has an active pH range of 2.5-3.5, which can increase the killing of acid resistant bacteria in the stomach (Zhu et al., 2006).

After reaching the small intestine, bacteria are exposed to bile salts, which are detergents produced by the liver to facilitate dispersion and digestion of fats. They are made from bile acids, which include cholic acid and chenodeoxycholic acid in humans, and are mainly secreted as conjugates, connected to either taurine or glycine. Conjugated bile acids are able to penetrate into cells via porin channels, such as OmpF (Thanassi et al., 1997). In their unconjugated form, bile acids are able to diffuse through the cell membrane (Plésiat and Nikaido, 1992), which may be of importance as several members of the microbiota are able to deconjugate bile salts (Franklund et al., 1993; Ridlon et al., 2006). *E. coli* has been shown to actively pump out bile salts through efflux systems EmrB and AcrAB (Thanassi et al., 1997). The AcrAB system has also been shown to be upregulated in the presence of bile salts and fatty acids (Rosenberg et al., 2003).

1.3.1.4 Nutrients and carbon sources

The resource ratio model of competition proposes that the predominance of different taxa is determined by availability, demand and consumption rate of nutrients (Hibbing et al., 2010). Niches are often defined by what nutrients are available and in what quantity. The gut is a heterogeneous environment and *E. coli* is able utilise many different carbon sources. The availability of certain nutrients changes along and across the gut, influenced by host breakdown of food, uptake of nutrients and nutrient utilisation by the microbiota (figure 1.3). However, *E. coli* only accounts for a small proportion of bacteria within the healthy gut, suggesting that its preferred nutrients are only available at low levels and its niche is small (Chang et al., 2004). *E. coli* has been shown to preferentially catabolise several monosaccharides which are the breakdown products of glycans within the mucus. These monosaccharides are most likely produced by other commensal bacteria, many of which possess glycan degrading enzymes, and perhaps possibly by host cells (Chang et al., 2004).

Sugars that *E. coli* has been shown to metabolise preferentially both *in vivo* and *in vitro* include gluconate, N-acetylglucosamine, N-acetylneuraminic acid, glucuronate, mannose, fucose and ribose (Chang et al., 2004; Alpert et al., 2009). Mutants missing the catabolic pathways for any of these seven sugars are outcompeted by wild-type strains in the murine gut, though all mutants were able to colonise streptomycin fed mice when monoassociated (Chang et al., 2004). The metabolism of fucose and ribose is linked, with strains unable to utilise fucose switching to ribose metabolism in the mouse GI-tract, and double knockout mutants unable to utilise either fucose or ribose being eliminated from the gut (Autieri et al., 2007). Glycogen may also play a role in the gut as mutants unable to utilise or store glycogen had reduced colonisation abilities (Jones et al., 2008). Importantly, carbon source mutants are differentially affected in their ability to colonise the gut. N-acetylneuraminic acid and N-acetylglucosamine mutants are only affected during initial colonisation, whereas glucuronate, mannose, fucose and ribose mutations are affected during maintenance (Chang et al., 2004). These observations would indicate that

colonisation consists of two phases in terms of catabolic processes; initial and maintenance phases, both of which have different nutrient requirements. As the use of different carbon sources changes over time, new incoming *E. coli* strains may therefore not be in direct competition with resident *E. coli* for particular resources.

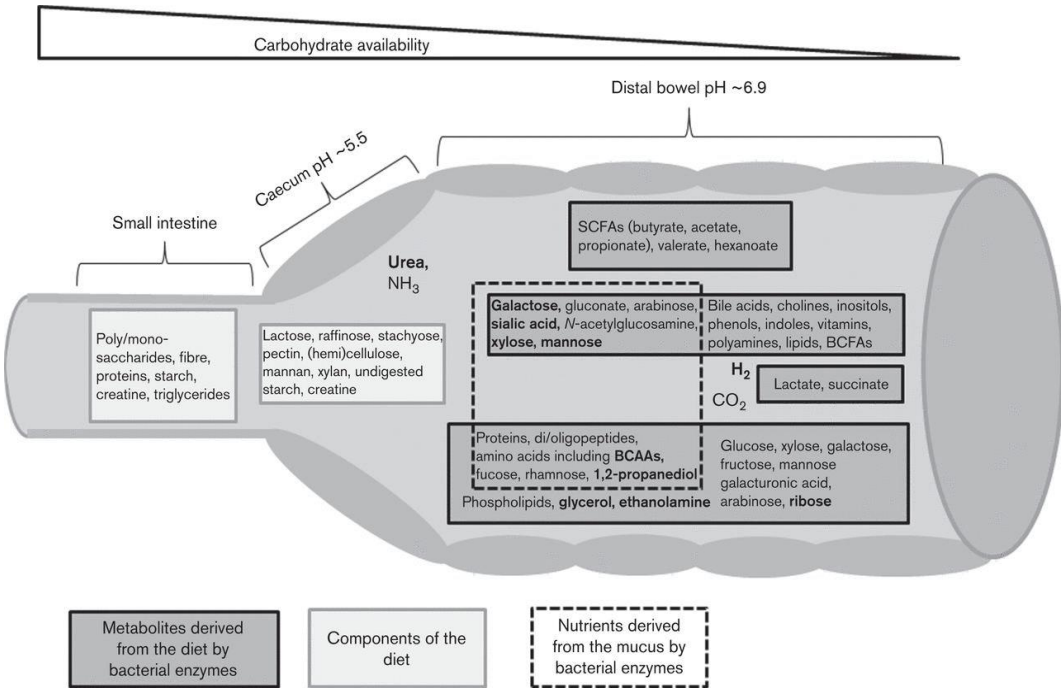


Figure 1.3: Nutrients available for commensal and pathogenic bacteria in the GI-tract. Metabolites shown in bold are specifically used by enteropathogens. SCFA (short-chain fatty acid), BCFA (branched-chain fatty acid), BCAA (branched-chain amino acid) (Staib and Fuchs, 2014). This figure is copyrighted by Microbiology Society.

Comparisons between commensal and pathogenic natural isolates and laboratory strains of *E. coli* have highlighted that core catabolic and carbon uptake operons show very little variation (Ihssen et al., 2007). This large nutritional overlap could explain why commensal strains offer protection against intestinal pathogenic strains of *E. coli* such as O157:H7 in mice (Miranda et al., 2004; Leatham et al., 2009). However, *in vivo* studies in mice have shown that the sugars used by EHEC O157:H7 are slightly different to those used by commensal strains (Durso et al., 2004; Fabich et al., 2008). The commensal *E. coli* strain HS and probiotic *E. coli* strain Nissle 1971 were both shown to utilise slightly different carbon sources in the GI-tract (Maltby et al., 2013). When combined, these two strains use all five sugars known to be important in EHEC

O157:H7 strain EDL933 colonisation of the mouse gut, and were shown to prevent colonisation by the pathogen (Maltby et al., 2013). It is possible that the overlap in carbon sources used by both commensals and pathogens is enough to grant colonisation resistance, or the strains may be competing for other resources.

As well as carbon, nitrogen is an essential nutrient for which *E. coli* has evolved multiple uptake and catabolic systems which utilise nitrogen in many forms, including amino acids, purines, pyrimidines, ammonia and nitrate. *E. coli* grown on mucus displayed repression of amino acid and nucleotide biosynthesis genes, as well as evidence for amino acid catabolism, suggesting that these are available in the GI-tract (Chang et al., 2004). However, *E. coli* K-12 strain MG1655 mutants in amino acid catabolism showed no colonisation defect in streptomycin treated mice (Chang et al., 2004). Other *E. coli* K-12 MG1655 mutants unable to synthesise purines and pyrimidines did have reduced colonisation of gnotobiotic mice (Vogel-Scheel et al., 2010), which suggests that although purines and pyrimidines may be available in mucus (Chang et al., 2004), they are limiting in the gut environment and important in gut colonisation.

A major source of nitrogen for the gut microbiota is nitrate. The majority of human dietary nitrate is absorbed in the small intestine; however, approximately a third reaches the lower intestines, of which only 1% is recovered in faeces (Bartholomew and Hill, 1984). Nitrate is thought to be converted in the gut by commensal bacteria, including *E. coli*, into nitrite and ammonia. *E. coli* is able to use nitrate as a terminal electron acceptor during anaerobic respiration to produce nitrite, which is then subsequently converted to ammonia (Tiso and Schechter, 2015). In the presence of nitrate, *E. coli* has been shown to have increased growth *in vitro* under the anaerobic and microaerobic conditions likely to be found in the GI-tract. Alongside nitrate, N-oxides ($R_3N^+–O^-$) and S-oxides ($RS^+–O^-$), which are generated during oxidation of tertiary amines and thiols respectively (Youssif, 2001; Gupta and Carroll, 2014), can be used as terminal electron acceptors, which has implications on *E. coli*

growth during inflammation. Nitrate, N-oxides and S-oxides can all be produced during inflammation from nitric oxide (NO) or reactive oxygen species (ROS) released by the host immune system (Lundberg et al., 1994; Singer et al., 1996; Enocksson et al., 2004; Winter et al., 2013). As a result, *E. coli* overgrowth may be possible, and has been reported in inflammatory bowel disease (Martinez-Medina and Garcia-Gil, 2014). As well as S-oxides, *E. coli* is able to utilise other forms of sulphur, including cysteine, sulphate, sulphite and sulphonates. *E. coli* preferentially uses cysteine, which is found in intestinal mucus where it is responsible for mucus integrity and barrier function (Gouyer et al., 2015).

1.3.1.5 Respiration and oxygen

One of the main factors that changes along the gut is oxygen availability. In the external environment and in the mouth, oesophagus and stomach, bacteria are exposed to high levels of oxygen. After entering the small intestine, however, oxygen is gradually reabsorbed through the mucosa (He et al., 1999) leading eventually to the primarily anaerobic colon (Saldeña et al., 2000). Oxygen is, however, still present in the colon next to the mucosa at levels of approximately 2-7% and a diffusion gradient occurs across the mucus leading to the anaerobic lumen (He et al., 1999; Marteyn et al., 2011; Espey, 2013). As a facultative anaerobe, E. coli is able to survive in both the oxygen rich upper GI-tract and the oxygen poor colon. Removal of oxygen by E. coli is also vital to the obligate anaerobes within the colon, which form more than 99% of the colon microbiota (Jones et al., 2011). E. coli are capable of using nitrate and fumarate as alternative terminal electron acceptors during anaerobic respiration as well as fermentation when oxygen levels are low (Jones et al., 2007). Nitrate within the gut is, however, extremely limited and only affects E. coli during initial colonisation as it is rapidly depleted. After colonisation with EHEC strain EDL933 the nitrate concentration in mucus was shown to become undetectable within 24h (Jones et al., 2011). Nitrate metabolic mutants were, as a result, unable to colonise the gut when competing with wild-type strains. In contrast, fumarate is important in long-term persistence within the gut, rather than initial colonisation, as mutants unable to utilise fumarate were able to colonise the GI-

tract of mice but numbers declined over time at a faster rate than the wild-type strains (Jones et al., 2011).

It is likely that in the colon, as E. coli occupies the mucus near the epithelium, E. coli lives in a microaerobic environment (Jones et al., 2007). Under such growth conditions it would be important for it to be able to perform both aerobic and anaerobic respiration to optimise fitness and thus increasing its chances of entering the microbiota. Indeed, mutants unable to perform either aerobic or anaerobic respiration have reduced fitness in the GI-tract of mice compared to the wild-type (Jones et al., 2007; Jones et al., 2011).

1.3.1.6 The gut microbiota and colonisation resistance

The type and amount of resident bacteria influences how well they can prevent incoming bacteria becoming stable members of the microbiota. In the human GI-tract there is a wide variety of species which form the microbiota, with over 1,000 species known to be able to colonise the colon (Rajilić-Stojanović et al., 2007). Not only is the colonic microbiota diverse, but it is also abundant, with up to 10^{11} cells per gram of faeces often being detected in healthy individuals. Recently three major distributions, named enterotypes, have been identified (Arumugam et al., 2011), though whether these are distinct groups or form part of a gradient is unclear (Jeffery et al., 2012). These enterotypes are classified based on the dominant genera: *Bacteroides*, *Prevotella* or *Ruminococcus*. Other important genera include *Bifidobacterium*, *Eubacterium*, *Lactobacillus* and *Peptostreptococcus* (Wilson, 1993; Arumugam et al., 2011). The differences in enterotypes could not be explained by host properties, but some correlations were found between microbiota composition and age, Body Mass Index (BMI) (Arumugam et al., 2011) and diet (Wu et al., 2011). Key host factors that influence the composition of the microbiota include age (Kirjavainen et al., 1998; Ley et al., 2006; Arumugam et al., 2011; Portal-Celhay and Blaser, 2012), gut morphology (Ley et al., 2008), diet (Ley et al., 2008) and immune competence (Khachatryan et al., 2008; Gulati et al., 2012).

Colonisation resistance occurs when the microbiota prevents colonisation by incoming bacteria, both pathogenic and commensal, in the GI-tract. Mouse models show that *E. coli* is rapidly excreted and can be lost from the GI-tract within 18 hours of ingestion in wild-type mice (Freter et al., 1983). However, in streptomycin treated and germ-free mice, which have a reduced or no microbiota respectively, *E. coli* is able to persist in the gut after ingestion (Freter et al., 1983; Rang et al., 1999). This persistence is not reduced after addition of a complete conventional microbiota (Freter et al., 1983; Rang et al., 1999), suggesting the main effect of colonisation resistance is during initial colonisation rather than during maintenance.

Colonisation ability is not increased if a strain is re-introduced into a GI-tract where it has already colonised (Leatham et al., 2009). This is most likely due to the incoming population directly competing with the resident population for the same niche. Unless there are vacant niches available, or the incoming bacteria have a competitive advantage over the current residential strain, colonisation is unlikely to occur. Commensal strains can also offer protection against intestinal pathogenic *E. coli*, such as EHEC O157:H7 (Leatham et al., 2009). A reduced diversity in resident *Enterobacteriaceae*, including *E. coli*, reduces the colonisation resistance of the human gut microbiota to *E. coli* infection (Apperloorenkema et al., 1990).

Several mechanisms of colonisation resistance have been proposed: killing or exclusion of incoming bacteria via oxygen consumption (Altier, 2005; Marteyn et al., 2010), growth inhibition through secretion of antimicrobials and short-chain fatty acid (SCFA) production (Shin et al., 2002; Gantois et al., 2006; Duncan et al., 2009; Fukuda et al., 2011), as well as competition for nutrients and niches (Lawley et al., 2012). For antimicrobial secretion, incoming bacteria, as well as the microbiota, may be able to produce them. Few *E. coli* strains are known to produce bacteriocins such as colicin, which is toxic against other *E. coli* strains. However, the competitive advantage lies with the resident strains in the microbiota, as they most likely have superior numbers. To release colicin bacteria must lyse and the small number of incoming bacteria may not be able

to sufficiently kill enough of the resident microbiota and retain enough cells to successfully colonise the microbiota (Durrett and Levin, 1997). The type VI protein secretion system (T6SS) has been shown to enable contact-dependent killing of competing bacteria via injection of effector proteins and virulence factors (Bingle et al., 2008). Many Gram-negative bacteria possess the T6SS including *V. cholerae*, which has been shown to use it to kill competing *E. coli* (Dong et al., 2013). Commensal strains may also be able to suppress expression of virulence factors of invading pathogenic bacteria (Kamada et al., 2013) or provide a physical barrier by heavily colonising the mucus layer and preventing access to the epithelium (Juge, 2012).

There are several strategies utilised by bacteria to optimise their ability to colonise the gut and overcome colonisation resistance. These include changing the gut environment to their advantage and to disrupt the microbiota, activation of the immune system (by some pathogenic strains), adaptation to the gut environment and increased fitness compared to residents (Rezzonico et al., 2011; Stecher and Hardt, 2011). For many commensal *E. coli*, however, although they may be adapted to and competitive within the gut environment, whether they can successfully colonise the gut and become stable members of the microbiota may depend more on being in the right place at the right time.

1.3.2 The external (secondary) environment

It has been estimated that 50% of all living *E. coli* are present in secondary environments at any one time, including soils, water and on plants (Savageau, 1983). These bacteria are released from the GI-tract into the external environment through defaecation by the host, at which point a large majority of them probably die (Winfield and Groisman, 2003). Those that are able to survive and persist within the external environment may then go on to colonise another host when re-ingested. The ability to persist in this external non-host environment is not equal for all strains of *E. coli* and as a result some can be found in much higher proportions. This suggests that life and survival in the secondary environment is not entirely stochastic and provides certain selection

pressures on *E. coli* which has a significant influence on its ecology (Bergholz et al., 2011; Méric et al., 2013).

The presence of *E. coli* is monitored by many industries to detect faecal contamination, under the assumption that *E. coli* has a quick and high level of death on release from the host and does not persist in the external environment for a long time. This assumption has, however, been questioned following multiple observations that *E. coli* is able to persist in the secondary environment, such as in soils and water, for long periods of time. DNA fingerprint profiles between *E. coli* isolated from soil and strains from surrounding wildlife (hosts) showed that the two groups were distinct, and it was suggested that *E. coli* may form persisting populations in soil, and possibly in other secondary habitats (Byappanahalli et al., 2006). Alternatively, the predominance of certain *E. coli* isolates in soil compared to host faecal samples may reflect their increased ability to survive in the external environment, rather than the presence of entirely different populations of *E. coli* strains. *E. coli* strains that might constitute true naturalised populations have been isolated from secondary environments, including tropical soils (Byappanahalli et al., 2006; Ishii et al., 2006; Goto and Yan, 2011), water (Bermúdez and Hazen, 1988; Power et al., 2005; Vital et al., 2008), sediments (Solo-Gabriele et al., 2000; Whitman and Nevers, 2003; Ishii et al., 2007) and plants (Solomon et al., 2003; Islam et al., 2004; Ibekwe et al., 2007). It has been postulated that tropical soils and waters provided a warm and moist environment that replicated the GI-tract enough to permit growth of *E. coli* in these environments (Winfield and Groisman, 2003). However, reports suggest that *E. coli* could form sustainable populations or survive for long periods in soils from temperate climates (Sjogren, 1995; Byappanahalli et al., 2006; Ishii et al., 2006; Ishii et al., 2007; Texier et al., 2008; Brennan, Abram, et al., 2010; Brennan, O'Flaherty, et al., 2010).

1.3.2.1 E. coli adaptation to plants

Plants are increasingly considered an important reservoir for pathogenic and commensal *E. coli*, with increasing numbers of EHEC outbreaks being traced to

contaminated vegetable products (Herman et al., 2015). Initial attachment of *E. coli* to plants has been studied in some detail, with a particular focus on pathogenic strains. Biofilm formation has been highlighted as a possible trait that influences plant attachment (Niemira and Cooke, 2010; Méric et al., 2013; Yaron and Romling, 2014). *csg* curli, which are an important component of extracellular matrices, when introduced into *E. coli* K-12 strains that are unable to colonise plants, resulted in increased binding to alfalfa sprouts and seed coats (Torres et al., 2005). Cellulose, another component of extracellular matrices, has been implicated in plant attachment by both S. *enterica* and *E. coli* O157:H7 strains, with cellulose biosynthesis mutants having a reduced ability to colonise alfalfa sprouts (Barak et al., 2007; Matthysse et al., 2008).

In EHEC strains, the LEE-encoded EspA filaments were shown to be important in attachment to salad leaves with no adherence observed for *ΔespA* mutants (Shaw et al., 2008). To support this observation, EHEC T3SS mutants were also shown to have reduced colonisation of spinach and lettuce leaves (Shaw et al., 2008; Xicohtencatl-Cortes et al., 2009). Due to the physical properties of the plant cell wall, it is thought that translocation of effector proteins into plant cells via the T3SS is unlikely (Shaw et al., 2008). One effector protein, EspB, was shown to not influence attachment, but was important for EHEC tropism towards stomata (Shaw et al., 2008). ETEC colonisation factors (CFA) and the associated secreted adhesin *etpA*, which are normally involved in intestinal colonisation, did not affect plant colonisation by this pathogen (Shaw et al., 2011). Strong evidence of virulence traits being involved in the association of *E. coli* with plants has been provided for the AAF (aggregative adherence fimbriae) pili, which can be used by EAEC pathogenic strains of *E. coli* to adhere to epithelial cells in the gut (Berger et al., 2009b).

There is evidence to suggest that flagella are able to influence colonisation of plants by *E. coli*, with studies reporting reduced colonisation of spinach and lettuce leaves by *ΔfliC* ETEC and EHEC mutants (Xicohtencatl-Cortes et al., 2009; Shaw et al., 2011). EHEC O157:H7 flagella have been shown to facilitate attachment via interactions with ionic lipids in spinach and *Arabidopsis thaliana*

plasma membranes (Rossez et al., 2014). Flagella have been observed to influence movement of EHEC and EAEC strains on and within plant tissues (Cooley et al., 2003; Berger et al., 2009b). *S. enterica* Typhimurium flagella have also been shown to have a role in chemotaxis-dependent internalisation of *Salmonella* into lettuce leaves (Cooley et al., 2003; Kroupitski et al., 2009). However, EAEC flagella mutants showed no reduction in attachment to plants compared to the wild-type, suggesting the role of flagella may be different depending on strain (Berger et al., 2009b). In support of this, it has been observed that *Salmonella* flagella involvement in plant colonisation is dependent on serovar (Berger et al., 2009a).

Internalisation of *E. coli* into the plant may be important in terms of persistence, as the internal environment may offer protection from external conditions. There are many observations of *E. coli* strains, both commensal and pathogenic, being recovered from inside plant tissues (Solomon et al., 2002; Warriner et al., 2003; Hora et al., 2005; Ongeng et al., 2011). There are two proposed routes of internalisation: (i) active movement of bacteria through natural openings (e.g. stomata) or sites of damage on the plant surface, and (ii) passive uptake of bacteria by plants alongside water uptake (Deering et al., 2012). Chemotrophic attraction to stomata has been shown in EHEC O157:H7 strains (Shaw et al., 2008), suggesting that some strains are able to actively penetrate plant leaves. Exposure of seeds or roots to contaminated water has shown that both *E. coli* and *Salmonella* spp. can be internalised and spread throughout the plant through the xylem, as bacteria were subsequently isolated from unexposed aerial parts of the plants (Warriner et al., 2003). The likelihood of an *E. coli* strain achieving internalisation is influenced by the type of plant, plant age, route of contamination, as well as the strain itself (Pu et al., 2009). Once inside the plant, several studies have indicated that *E. coli* is capable of replicating to high levels (Cooley et al., 2003; Jablasone et al., 2005; Deering et al., 2011). One difficulty faced by internalised *E. coli* is that the plant innate immune system may be triggered by pathogen associated molecular patterns (PAMPs), such as flagella or LPS (Melotto et al., 2014). Whether *E. coli* can suppress this immune

response like phytopathogens has not been investigated. It is important to note, that although *E. coli* is able to become internalised, there are no recordings of *E. coli* being able to cause pathogenesis in plants, and it has been hypothesised that *E. coli* are simply taking advantage of plants as a transmission vector (Holden et al., 2009; Chekabab et al., 2013).

The availability of carbon sources to *E. coli* on plants varies depending on plant age, growing conditions and wound presence (Hora et al., 2005; Brandl, 2008). Several nutrients are potentially available to bacteria on or within leaves, including carbohydrates, organic acids, amino acids, methanol and various salts (Corpe and Rheem, 1989; Mercier and Lindow, 2000). Water availability on leaves was shown to be sufficient using a *Pantoea agglomerans* bioreporter strain, with the aqueous laminar layer that coats leaves possibly providing access to enough water (Lindow and Brandl, 2003). However, the availability of some nutrients on leaves is often not homogenous, with both sugars (fucose and sucrose) and ferric iron being concentrated in certain areas (Joyner and Lindow, 2000; Leveau and Lindow, 2001). There is evidence that siderophores are important in plant colonisation, with enterobactin and salmochelin both being induced in *S. enterica* Typhimurium grown in alfalfa root exudates (Hao et al., 2012). This, however, may be influenced by the type of plant being colonised as polyphenolic compounds (e.g. tannins) are able to sequester iron to varying degrees (Karamanoli et al., 2011). Another limiting nutrient on leaves is nitrogen, which was shown to limit growth of *E. coli* O157:H7 strain H1827 (Brandl and Amundson, 2008). On well fertilised plants, however, evidence suggests that carbon sources become more limiting, due to increased availability of nitrogen (Brandl and Amundson, 2008).

1.3.3 Requirements of the "biphasic" lifestyle of E. coli

The lifestyle of *E. coli* is largely biphasic, consisting of host-associated (primary environment) and host-independent (secondary environment) phases (van Elsas et al., 2011). Two hypotheses have been proposed to explain how *E. coli* responds to the transition from the primary to secondary environment, and vice versa. "Demand theory" states that during the transition between

environments, *E. coli* adapts by altering gene regulation and expression, with high-demand genes being positively regulated and low-demand genes negatively regulated (Savageau, 1974; 1983). In contrast, it has been proposed that specific strains are primarily adapted to either the primary or secondary environment, with selection pressures influencing *E. coli* transitioning between these environments (Whittam, 1989; Gordon et al., 2002). These strains may possess certain genes or phenotypes that facilitate survival in their respective environment. Comparisons of the genetic population structures of *E. coli* in the primary and secondary environments show distinct profiles supporting selection as shaping *E. coli* populations (Gordon et al., 2002; White et al., 2011). However, it is most likely a mixture of these two processes that shape *E. coli* adaptation with both differences in gene regulation and gene content being important factors.

1.3.3.1 Trade-offs: Self-Preservation and Nutritional Competence (SPANC)

In both the primary and secondary environments, *E. coli* is exposed to many stressors, and as a result stress protection and the stress response are important to survival. Stress protection is tightly regulated by the alternative sigma factor RpoS, a key regulator of the general stress response (Hengge-Aronis, 2002). Observations that *rpoS* mutants were better competitors for nutrients but were poor at surviving environmental stressors (King et al., 2004), lead to the hypothesis that *E. coli* faces a trade-off between self-preservation and nutritional competence (SPANC) with strains in different positions on the SPANC balance possibly occupying different niches (Ferenci, 2005). Stress protection diverts energy away from metabolic processes, which could result in reduced growth. *E. coli* O157:H7 isolates with a functional *rpoS* system have increased persistence in soil, which may be linked to oxidative capacity (Franz et al., 2011; van Hoek et al., 2013). In the gut, however, RpoS may provide little benefit as *E. coli* strain BJ4 *ΔrpoS* mutants fed to mice were able to colonise the gut and outcompete the wild-type (Krogfelt et al., 2000). In good

correspondence with this hypothesis, *E. coli rpoS* mutants can be readily isolated from sources of faecal contamination (Chiang et al., 2011).

Antibiotic resistance is often acquired through changes in cell permeability, and *E. coli* K-12 strains with varying antibiotic resistance were shown to have differing colonisation in germ-free mice (Onderdonk et al., 1981). Importantly, increased stress protection against many stressors, including antibiotic agents, bile salts and desiccation, appear to be linked to reduced cell membrane permeability and possibly reduced nutrient uptake (Ferenci, 2005). Indeed, in the murine gut *E. coli* strains quickly develop mutations in membrane protein genes such as *ompB*, *envZ* and the *flhDC* operon, as well as in the *malT* gene that regulates maltose uptake (Giraud et al., 2008; De Paepe et al., 2011). These mutations affected membrane permeability leading to increased resistance against bile and increased survival *in vivo* and *in vitro* when grown in media containing bile salts compared to ancestor strains. Mutant strains showed reduced growth in the absence of bile salts compared to the parental strains, suggesting reduced uptake of nutrients (De Paepe et al., 2011). Within the gut the trade-off seems to be between nutritional competence and resistance to bile secreted into the small intestine (Giraud et al., 2008; De Paepe et al., 2011).

In summary, the SPANC balance could reflect the trade-off arising between host and non-host environmental adaptation, and might be an important factor involved in *E. coli* diversification within the GI-tract itself (De Paepe et al., 2011).

1.4 Iron homeostasis and siderophores

1.4.1 Iron homeostasis in bacteria

For most bacteria iron is essential, playing a role in many cellular processes, including DNA replication, photosynthesis, nitrogen fixation, methanogenesis, hydrogen production and consumption, protection from oxidative stress, respiration, energy generation, the tricarboxylic acid (TCA) cycle, oxygen transport and gene regulation (Andrews et al., 2003). Very few bacteria have been shown to be able to grow in the absence of iron, such as lactobacilli (Archibald, 1983), the Lyme disease pathogen *Borrelia burgdorferi* (Posey and

Gherardini, 2000) and the syphilis pathogen *Treponema pallidum* (Posey et al., 1999). The lactobacilli are thought to have evolved the ability to survive without iron as they live in milk which contains lactoferrin, a host iron binding protein (Weinberg, 1997). Instead of iron, lactobacilli probably incorporate high levels of manganese into their proteins (Archibald and Duong, 1984; Imbert and Blondeau, 1998), and intracellular levels of iron in lactobacilli can be as low as two atoms per cell (Sabine and Vaseleko, 1967; Imbert and Blondeau, 1998). *B. burgdorferi* and *T. pallidum* are obligate intracellular pathogens with minimal genomes and as such are able to rely on host cellular processes to compensate for pathways that require iron (Posey et al., 1999; Posey and Gherardini, 2000). It is possible that these intracellular parasites evolved to not use iron as a result of their hosts sequestering available iron.

Because of the high availability of iron during the Archean Eon, 4,000 to 2,500 million years ago, it has been suggested that it was incorporated into proteins early on during the evolution of life (Beinert et al., 1997). Iron can be found in a wide range of oxidation states, but the two most common are the Fe^{2+} (ferrous) and Fe^{3+} (ferric) forms. During early life on Earth, before the appearance of photosynthesis and the associated production of significant amounts of oxygen, iron would have been readily available in its soluble Fe^{2+} form. In contrast, bacteria (and other organisms) that occupy environments that are rich in oxygen cannot gain easy access to iron, as it readily oxidises to its Fe^{3+} form which is extremely poorly soluble. Another important characteristic of ferric iron that indicates it was incorporated in early life forms before the appearance of photosynthesis is that under aerobic conditions Fe^{2+} can be toxic. This is due to its ability to react with oxygen and generate free radicals that can damage cellular components including DNA and proteins (Andrews et al., 2003). As a result, following the increase of atmospheric oxygen bacteria have evolved mechanisms to tightly regulate iron levels within the cell to ensure they have access to sufficient amounts of iron, while limiting its toxicity.

There are five main mechanisms (Andrews et al., 2003) through which bacteria regulate intracellular iron levels: (i) scavenging iron from the surrounding

environment using high-affinity transport molecules and receptors (ii) storage of iron within the cell to be used when external iron levels are low (iii) control of the incorporation of iron into proteins and enzymes to reduce iron use when it is in short supply (iv) redox stress systems able to reduce or prevent damage from iron (v) regulatory systems that co-ordinate all of the mechanisms above according to iron availability. The way these general mechanisms function will depend on the bacteria, their environment and iron availability.

1.4.2 Siderophores in E. coli

To survive, *E. coli* must maintain an intracellular iron concentration between 10^{-7} and 10^{-5}M, or 10^5 to 10^6 ferric ions each generation (Braun and Braun, 2002; Andrews et al., 2003; Raymond et al., 2003). To facilitate uptake of Fe^{3+} bacteria have evolved several different strategies, including lowering the external pH as Fe^{3+} becomes more soluble in acidic environments, reducing Fe^{3+} into the more soluble Fe^{2+} which can be taken up using less costly ferrous transporters, or using Fe^{3+} chelators known as siderophores (Guerinot, 1994). The main strategy employed by *E. coli* to facilitate Fe^{3+} uptake is the use of siderophores. Many species use these molecules, and more than 500 different siderophores have been described (Ratledge and Dover, 2000; Wandersman and Delepelaire, 2004). There are five major classes of siderophore: catecholates, phenolates, hydroxamates, α-hydroxy-carboxylates and mixed type (reviewed in Garenaux et al., 2011) which all have slightly different characteristics.

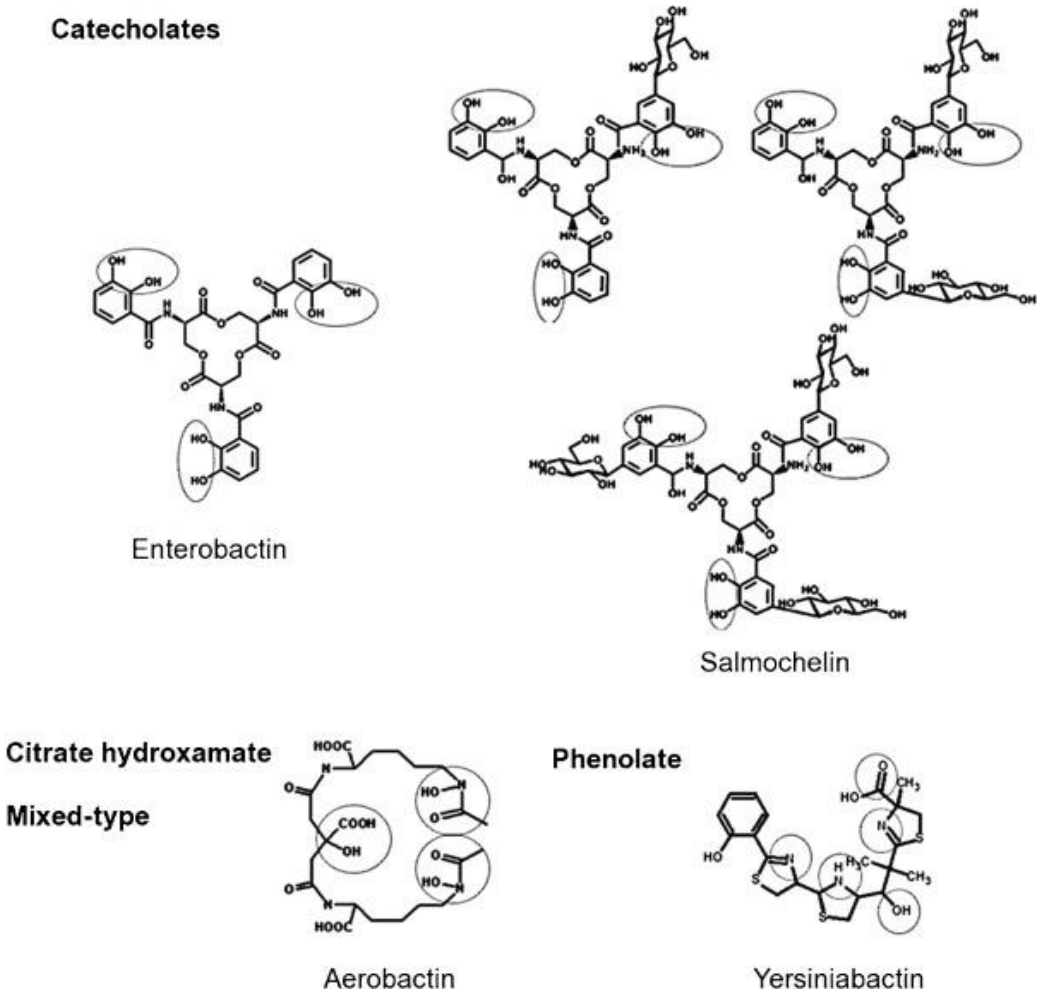


Figure 1.4: Chemical structure of the four siderophores produced by *E. coli*. Salmochelin, a glucosylated form of enterobactin, has three possible structures, depending on the number of glucose molecules attached. Circles denote the iron-binding regions of each siderophore (Garenaux et al., 2011). This figure is copyrighted by Elsevier B.V.

E. coli can synthesise up to four siderophore systems; enterobactin, salmochelin, yersiniabactin and aerobactin (figure 1.4). Enterobactin, a catecholate type siderophore has the highest binding constant for ferric iron of any known siderophore (Grass, 2006). Enterobactin is widely synthesised within enterobacteria and has been found in almost all *E. coli* strains, both pathogenic and commensal (Ratledge and Dover, 2000; Crosa and Walsh, 2002). Salmochelin is a modified version of enterobactin, and as such, shares most of its synthetic pathway (Grass, 2006). Salmochelin may be important in the host environment, as enterobactin is bound and inhibited by the host molecule lipocalin-2. Modifying enterobactin into salmochelin changes its shape,

preventing lipocalin-2 from binding, and allowing iron uptake to occur (Fischbach et al., 2006). Yersiniabactin was first described in *Yersinia enterocolitica* and is a phenolate type siderophore (Heesemann et al., 1993). Finally, aerobactin is a mixed-hydroxymate type siderophore that, similarly to salmochelin and yersiniabactin, is often associated with pathogenic strains of *E. coli* as a virulence factor (Köhler and Dobrindt, 2011). Siderophore-dependent iron uptake generally follows the same steps, even though they can display a wide range of differences in chemical properties. The main stages involved in siderophore use and Fe^{3+} uptake include synthesis, export, binding of the ferri-siderophore to outer membrane receptors, internalisation and iron release into the cytoplasm.

The fact that some strains of *E. coli* encode multiple siderophore systems raises the question of redundancy. However, each siderophore displays unique characteristics, which may result in differing optimum activity in different environments. Enterobactin has been shown to have the highest binding affinity for iron of any known iron chelator, with a log β_{110} of 49 (Loomis and Raymond, 1991). β_{110} represents the stability constant, and is calculated based on the ratio of siderophore-bound iron to unbound iron in mixed solutions (Boukhalfa and Crumbliss, 2002). Higher values for log β_{110} indicate that the siderophore-iron complex has a higher association rate (Boukhalfa and Crumbliss, 2002). The binding affinity of salmochelin is unknown, but as its Fe^{3+} binding site is identical to that of enterobactin, it is most likely able to bind Fe^{3+} with an equivalent affinity. In contrast, aerobactin has one of the weakest binding affinities for $Fe^{3+,}$ at log β_{110} = 22.5 (Harris et al., 1979). However, under acidic conditions the binding affinity for both enterobactin and aerobactin has been shown to be approximately equal, up to a pH of 5.6, after which the binding affinity of aerobactin plateaus while enterobactin increases further (Valdebenito et al., 2006). To reflect these differences, production of aerobactin by *E. coli* strain Nissle 1917 was increased at pH 5.6 compared to pH 7 (Valdebenito et al., 2006). In contrast to aerobactin, yersiniabactin, which has a binding affinity of log β_{110} of 36.6 (Perry et al., 1999), has reduced stability under acidic conditions

(Drechsel et al., 1995) and decreased production by *E. coli* strain Nissle 1917 was observed at pH 5.6 (Valdebenito et al., 2006). To support these observations, an *E. coli* Nissle 1917 *Δybt* mutant which cannot synthesise yersiniabactin was outcompeted by the wild-type at pH 7 and pH 7.6, but not at pH 5.6 (Valdebenito et al., 2006). Salmochelin production has also been shown to be altered by pH, with the salmochelin locus in several *S. enterica* Typhimurium strains being induced under alkaline conditions (Foster et al., 1994). Higher salmochelin production has also been observed at pH 7 and 7.6 by *E. coli* strain Nissle 1917 (Valdebenito et al., 2006). The observation that each siderophore is produced optimally at different pH suggests that possessing multiple siderophore systems may improve fitness in environments where the pH fluctuates, such as the host urinary tract and gut.

1.4.2.1 Siderophore synthesis and export

Siderophores are initially synthesised within the cell, with each siderophore having its own unique biosynthesis gene cluster (figure 1.5). The enterobactin gene cluster consists of 15 genes, six of which are involved in biosynthesis (Crosa and Walsh, 2002). Enterobactin is synthesised in two stages, firstly 2,3-dihydroxybenzoate (2,3-DHB) is synthesised from chorismate, and then converted with serine into enterobactin, a cyclised DHBS (2,3-dihydroxybenzoyl serine) triester (Gehring et al., 1997). Salmochelin is a modified form of enterobactin and as such it requires the enterobactin biosynthesis machinery. Additionally, salmochelin production requires a gene cluster of 5 genes which are responsible for the glucosylation of the 2,3-DHB molecule before addition of serine and formation of the triester (Mueller et al., 2009). This results in three forms of salmochelin (mono-, di- and tri-glucosyl-C-enterobactin) depending on how many glucosylated 2,3-DHB molecules are incorporated (figure 1.4) (Lin et al., 2005).

Yersiniabactin, similarly to enterobactin and salmochelin, is synthesised from chorismate. However, yersiniabactin belongs to a different class of siderophore, the phenolates, which give it different characteristics. The yersiniabactin gene cluster is formed of 11 genes and biosynthesis is carried out in seven stages, a

description of which can be found in Pfeifer (2003). Finally, the aerobactin gene cluster consists of five genes, four of which are involved in synthesis of aerobactin from L-lysine (De Lorenzo et al., 1986). After synthesis, siderophore export is carried out by transport proteins unique to each siderophore, except for enterobactin and salmochelin, which have some overlap. The EntS MFS (major facilitator superfamily) and IroC proteins export enterobactin and salmochelin, respectively, across the inner membrane into the periplasm (Ozenberger et al., 1989; Furrer et al., 2002; Crouch et al., 2008). The TolC protein exports enterobactin across the outer membrane (Bleuel et al., 2005), however, the equivalent transporter for salmochelin has not been identified. The exact mechanisms for secretion of aerobactin and yersiniabactin have also yet to be determined.

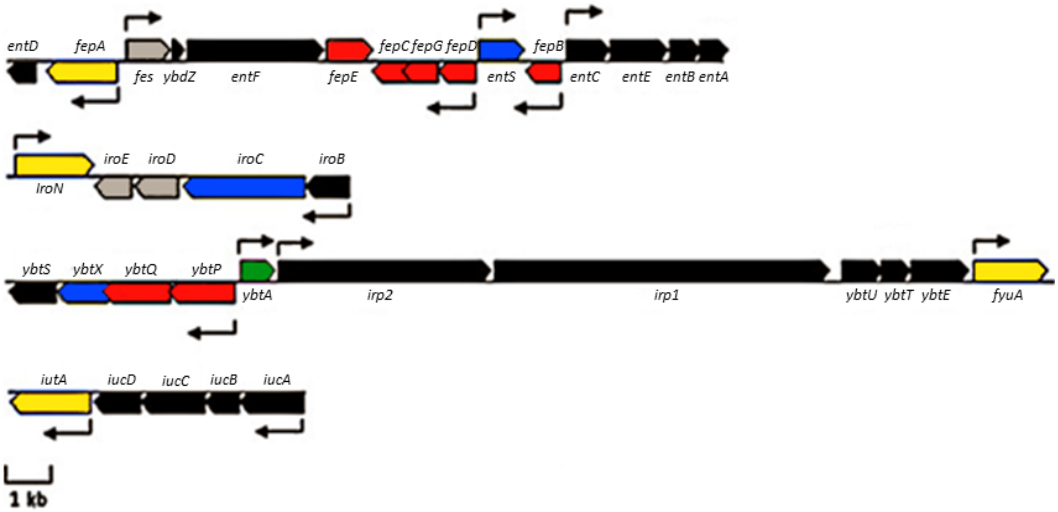


Figure 1.5: Gene clusters for all four siderophore systems in *E. coli*. Biosynthesis (black), receptor (yellow), export (blue), import (red), degradation (grey), regulation (green). Small arrows indicate promoter regions containing a Fur binding domain. From top to bottom: enterobactin, salmochelin, yersiniabactin and aerobactin (adapted from Garenaux et al., 2011). This figure is copyrighted by Elsevier B.V.

1.4.2.2 Siderophore uptake, internalisation and release of iron

After binding Fe^{3+}, ferri-siderophore complexes are imported into the cell via siderophore-specific receptors; FepA for enterobactin, IroN for salmochelin, IutA for aerobactin and FyuA for yersiniabactin (reviewed in Garenaux et al.,

2011). These receptors require the TonB-ExbB-ExbD energy transduction system to induce conformational changes and thus facilitate internalisation of the ferri-siderophore into the periplasm (Moeck and Coulton, 1998). Once internalised into the periplasm, ferri-siderophores must be transported across the inner membrane into the cytoplasm. Salmochelin and enterobactin share the transporter FepB and the inner membrane transport system FepDGC which together form an ABC transporter system (Chenault and Earhart, 1991). Aerobactin is transported into the cytoplasm using the FhuBCD system, which is also implicated in ferrichrome transport (Köster, 1991). Finally, yersiniabactin is transported using the YbtP-YbtU system (Fetherston et al., 1999).

After entering the cytoplasm, iron release from most siderophores occurs by reducing bound iron to Fe^{2+}, which results in either spontaneous release of iron from the siderophore or acquisition of the iron by other molecules (Miethke and Marahiel, 2007). This is how iron is released from aerobactin. The strong affinity that enterobactin and salmochelin have for Fe^{3+} means that the redox potentials of ferri-enterobactin and ferri-salmochelin are too low for reduction to occur (Ratledge and Dover, 2000). As a result these siderophores must be degraded to allow the Fe^{3+} to be released. Enterobactin and salmochelin can both be hydrolysed by either the Fes esterase or the IroD and IroE hydrolases (Lin et al., 2005; Zhu et al., 2005). Degradation of enterobactin and salmochelin is irreversible; however, the DHBS monomer produced can still bind ferric iron, although with much lower affinity (Hantke, 1990). It is currently unknown what the exact mechanism of iron release from ferri-yersiniabactin is, as no specific hydrolases have been identified (Brem et al., 2001), and no ferric reductases have been identified that are involved in iron acquisition via yersiniabactin (Garenaux et al., 2011).

1.4.2.3 Siderophores as virulence factors

In pathogenic *E. coli*, particularly ExPEC strains, siderophores have been described as virulence factors, more specifically the siderophores salmochelin, aerobactin and yersiniabactin. As enterobactin is in practically all strains of *E. coli*, both pathogenic and commensal, it is unclear whether it has a role in

virulence. Several studies have highlighted the fact that enterobactin synthesis does not affect virulence of pathogenic *E. coli* in chicken virulence models (Dozois et al., 2003; Caza et al., 2011). In *Salmonella* spp., some animal virulence models suggest that enterobactin increases virulence (Yancey et al., 1979; Furman et al., 1994; Nagy et al., 2013), however, other studies have shown no role (Benjamin et al., 1985; Rabsch et al., 2003). This indicates the relevance of individual siderophores for pathogenesis could be influenced by the host and the genetic background of the pathogen. One study of enterobactin production by *S. enterica* Typhimurium showed that although enterobactin production was not essential for virulence, iron uptake via its incomplete form, DHBS, was still required (Rabsch et al., 2003), suggesting that some form of ferric iron uptake system is still necessary for pathogenesis. Avian pathogenic *E. coli* (APEC) strain χ7122 enterobactin secretion mutants, but not biosynthesis mutants, had reduced virulence in a chicken virulence model (Caza et al., 2011), indicating a cost for enterobactin production.

A major problem with relying on enterobactin-mediated iron acquisition during the infection process is that it is bound and inhibited by lipocalin-2, also called neutrophil gelatinase-associated lipocalin (NGAL), a protein that is involved in innate immunity by sequestering iron to limit bacterial growth. Lipocalin-2 is secreted by host cells, particularly neutrophils, but also by epithelial cells in the GI-tract, in response to bacterial PAMPs (Chassaing et al., 2014). The glycosylated form of enterobactin, salmochelin, cannot be bound by lipocalin-2 (Fischbach et al., 2006) enabling *E. coli* to acquire iron more readily during inflammation and infection. Salmochelin has been shown to be involved in virulence for both ExPEC and APEC. APEC strain E058 and UPEC strain U17 salmochelin biosynthesis mutants showed reduced virulence in chicken models (Gao et al., 2012). Deletion of the *iro* locus in APEC strain χ7122 resulted in reduced virulence in chickens (Dozois et al., 2003; Caza et al., 2008). The salmochelin receptor gene, *iroN*, was shown to be overexpressed in urothelial cell intracellular bacterial communities (IBCs) in a mouse UTI infection model (Reigstad et al., 2007). IroN was also shown to be involved in virulence of UPEC

strains CFT073 and CP9 (Russo et al., 2002; Feldmann et al., 2007) and NMEC strain C5 (Negre et al., 2004).

As well as salmochelin, yersiniabactin and aerobactin have been shown to affect virulence in UPEC and APEC. Both have been shown to influence the establishment of infections by UPEC strains CFT073 and 536 in a UTI infection model (Garcia et al., 2011). Aerobactin is also more frequently detected in pathogenic strains of *E. coli*, particularly ExPEC strains, compared to commensals (Lafont et al., 1987; Linggood et al., 1987; Dozois et al., 1992).

1.4.2.4 Alternative functions of siderophores

Alongside binding Fe^{3+} and facilitating its uptake, several other functions for siderophores have been reported, including non-iron metal transport, non-metal transport, sequestration of toxic metals, protection from oxidative stress, molecular signalling and antibiotic activity (Johnstone and Nolan, 2015). Yersiniabactin has been shown to act as a zincophore, i.e. it can transport zinc instead of iron (Bobrov et al., 2014). Zinc is the second most abundant transition metal after iron found intracellularly in *E. coli* (Rouf, 1964). *E. coli* does possess a zinc transporter system ZnuABC, and the exact role of yersiniabactin in zinc transport is uncertain (Hantke, 2005). However, during infections, humans secrete zinc-sequestering proteins calprotectin and psoriasin to limit bacterial growth, and yersiniabactin may facilitate zinc uptake during infections (Kehl-Fie et al., 2011; Hood and Skaar, 2012). The Ybt-Zn complex is not taken up through the yersiniabactin receptor *fyuA* (Bobrov et al., 2014), suggesting a different uptake pathway is utilised.

Copper is secreted during infections by mammals to protect against microbial infection (White et al., 2009; Chaturvedi and Henderson, 2014). Copper secreted by the host during UTI reacts with enterobactin, producing toxic cuprous ions that can damage bacterial cells (Chaturvedi et al., 2014). Yersiniabactin has been shown to bind with Cu^{2+} before it is reduced, preventing any potential damage (Chaturvedi et al., 2012). The Ybt-Cu complex is stable and can be detected in the urine of UTI patients (Chaturvedi et al., 2012). This Ybt-Cu complex can also

work as a superoxide dismutase (Chaturvedi et al., 2014) by helping to reduce the level of NADPH oxidase-derived superoxide produced by macrophages, again protecting the bacteria from the host immune response. Enterobactin and salmochelin have also been shown to have a protective role against toxic compounds in the host environment. They have been shown in *Salmonella* to provide protection against oxidative stress in macrophages (Achard et al., 2013). When enterobactin and salmochelin are hydrolysed during iron release, hydroxyl groups are exposed which can scavenge free radicals (Adler et al., 2014). Enterobactin has also been shown to inhibit myeloperoxidase (MPO), a bactericidal enzyme secreted by the host during infection (Singh et al., 2015).

One additional role has been proposed for Yersiniabactin as a signalling molecule. It has been shown to regulate its own transcription, upregulating *fyuA*, *irp2* and *ybtP*, and downregulating the regulator gene *ybtA* (Perry et al., 2003; Anisimov et al., 2005a). The levels at which yersiniabactin elicits an effect on gene regulation are much lower than that required for a nutritional effect in *Yersina pestis* (Perry et al., 2003). Siderophores can also potentially be used as targeted antimicrobial delivery systems. Some bacteria are able to synthesise sideromycins, a class of antibiotics that consists of a siderophore covalently bonded to an antibiotic. Notable examples of sideromycins are salmycins produced by several *Streptomyces* spp. and albomycin produced by particular *Streptomycetes* (Braun et al., 2009). These two sideromycins are transported through the ferrichrome receptor (FhuA), but both show different activity, with salmycins being active against mainly Gram-positive bacteria, and albomycin against Gram-negative bacteria including *E. coli* (Braun et al., 2009). Once internalised into the target cell, albomycin is cleaved in the cytoplasm by peptidase N (PepN) to release the antibiotic from the siderophore (Braun et al., 2009). Synthetic sideromycins are currently being developed to utilise this specificity (Miethke and Marahiel, 2007).

1.4.3 Alternative iron uptake systems

Aside from siderophores, *E. coli* have access to multiple other iron uptake systems to secure iron scavenging under different conditions in a variety of environments. Most *E. coli* possess the Feo transport system that is able to take up Fe^{2+}. The *feoABC* genes that encode the system are induced under anaerobic conditions as well as under iron repression, when Fe^{2+} is expected to be more abundant than Fe^{3+}. FeoB has recently been shown to have G-protein functionality (Marlovits et al., 2002) and requires less energy than siderophore systems for iron uptake. Secretion of extracellular Fe^{3+} reductases has been documented in *E. coli* (Cowart, 2002). Reduced iron would then be taken up via FeoABC as Fe^{2+}. The second ferrous uptake system found in *E. coli* is EfeUOB, which is upregulated under low iron, aerobic and acidic conditions (Cao et al., 2007). Fe^{2+} is more abundant than Fe^{3+} not only in anaerobic environments, but also in aerobic environments when the pH is low. The two ferrous uptake systems of *E. coli*, therefore, allow optimal uptake of Fe^{2+} under both of these conditions. A third ferrous uptake system, SitABCD, is an ABC transporter that is prevalent in *Shigella* and EIEC strains (Johnson and Nolan, 2009). The role of SitABCD in the pathogenesis of these strains is unclear, however, it has been shown to contribute to virulence and resistance against oxidative stress in APEC O78 strain χ7122 (Sabri et al., 2008).

Receptors for siderophores usually only recognise a specific siderophore (Hohnadel and Meyer, 1988; Rabsch and Winkelmann, 1991; Liu et al., 1993; De Chial et al., 2003), but multiple siderophores can be used by a strain if they have multiple corresponding receptors (Rabsch and Winkelmann, 1991; Barelmann et al., 2002; Ghysels et al., 2004). One very interesting characteristic of siderophores is their vulnerability to being exploited by non-producing bacteria. As siderophores are secreted “public goods”, it is possible for non-producing bacteria, called cheaters, to take up siderophores produced by neighbouring cells as long as they have the required receptor. Cheating between *E. coli* cells has never been documented within natural populations of *E. coli*, however, it has been observed in a wide range of bacterial populations, from *Pseudomonas*

aeruginosa in the cystic fibrosis lung (De Vos et al., 2001; Buckling et al., 2007) to marine bacteria (D'Onofrio et al., 2010; Cordero et al., 2012). *E. coli* can, however, use xenosiderophores, siderophores produced by other species that they cannot make themselves, such as the fungal siderophore ferrichrome which binds to receptor FhuA (Köster and Braun, 1990; Braun et al., 2004). Ferrichrome is a hydroxymate type and shares the same cytoplasmic and periplasmic binding proteins as aerobactin, but the cell surface receptors remain specific to ferrichrome and aerobactin (Köster and Braun, 1990).

Other forms of iron, ferric citrate and haem, can also be bound to surface receptors. The FecA receptor binds ferric citrate, and internalises it using energy from the TonB-ExbB-ExbD inner membrane complex, much like siderophore receptors. The ChuA receptor binds haem and also requires energy from the TonB-ExbB-ExbD complex for internalisation. ChuA has been associated with ExPEC virulence (Torres et al., 2001; Garcia et al., 2011), and shown to be encoded by faecal isolates at lower frequencies compared to UPEC strains (Lloyd et al., 2007).

1.4.4 Iron uptake in the GI-tract

Most dietary iron in the gut will be ferric due to the instability of ferrous iron in the presence of oxygen, and is converted into Fe^{2+} by the host for absorption. However, within the mostly anaerobic colon it is possible for ferrous iron to stay in solution. *E. coli* uses several metal permeases, FeoB, EfeB, MntH and ZupT, which use passive transport to import Fe^{2+} into the cell (Kammler et al., 1993; Makui et al., 2000; Grass et al., 2005; Cao et al., 2007). FeoAB mutants in both *E. coli* and *S. enterica* Typhimurium are outcompeted by wild-type strains in the streptomycin-treated mouse (Kammler et al., 1993). Dietary iron also exists in the form of haem, which can be utilised by *E. coli* through the haemophores, such as the *chuA* transporter (Cescau et al., 2007), but a role for these in gut colonisation has not been described.

Enterobactin production has been shown to increase fitness *in vivo* with biosynthesis and receptor mutants being outcompeted by wild-type *E. coli* in the

murine GI-tract (Pi et al., 2012). Mutants were able to colonise the gut when given to mice alone, suggesting that although not essential to survival within the gut, siderophores increase competitiveness. The receptor knockout mutants were particularly less fit, possibly because they were still able to synthesise and export enterobactin. Exported enterobactin would then be able to sequester any surrounding Fe^{3+} and prevent any other iron uptake mechanisms from acquiring it. Within the gut, lipocalin-2 is secreted by immune or epithelial cells into the intestinal lumen during inflammation (Raffatellu et al., 2009). Even in the healthy gut, low levels of lipocalin-2 can still be detected in faeces (Chassaing et al., 2012), though whether this is able to influence the ability of *E. coli* to acquire iron through enterobactin is unclear.

E. coli K-12 strain MG1655 *ΔtonB* mutants, which are unable to take up ferric iron, had reduced CFUs and persistence in the GI-tract in competitive index assays with the wild-type, suggesting that Fe^{3+} is available in the GI-tract and is important in establishing a stable population in the gut (Pi et al., 2012). However, Kupz et al. (2013) found no significant reduction in fitness for a Δ*entC E. coli* Nissle 1917 mutant. There was, however, a trend towards the mutant having lower CFUs compared to the wild-type and becoming undetectable in some mice after 3-7 days. As *E. coli* strain Nissle 1917 has all four siderophore systems, it is possible that other siderophore systems, yersiniabactin and aerobactin, were able to compensate for the loss of enterobactin and salmochelin production. In contrast, the *E. coli* K-12 MG1655 strain used by Pi et al. (2012) only has the enterobactin siderophore, which may be why there was a discrepancy between these two studies. The *E. coli* Nissle 1917 strain has, however, been shown to outcompete pathogenic *S. enterica* in the mouse gut by presumably competing for iron using siderophores (Deriu et al., 2013). TonB and siderophore receptor knockout mutants were able to colonise the gut and reduce inflammation at a comparable level to the wild-type, but did not reduce S. *enterica* abundance. The contrast in *tonB* mutant colonisation seen between Pi et al. (2012) and Deriu et al. (2013) could be as a result of *E. coli* strains K-12 MG1655 and Nissle 1917 possessing a different complement of iron uptake

systems which do not require TonB. Alternatively, the presence of S. *enterica* in the GI-tract may affect the gut environment, or *E. coli* Nissle 1917 directly, that results in changes in iron uptake activity.

The presence of multiple types of siderophore in *E. coli* is important both in UTI strains, and possibly within the gut, as the different types work optimally under different environmental conditions (Valdebenito et al., 2005; Valdebenito et al., 2006). For example, aerobactin forms a more stable complex under acidic conditions, whereas salmochelin is better in more alkaline conditions. As the pH changes along the GI-tract as it does along the urinary tract, having siderophores that are able to function in various pH levels may confer a competitive advantage. It may be for this reason that yersiniabactin and salmochelin genes have been highly associated with UPEC strains (Henderson et al., 2009).

1.4.5 Regulation of siderophores and iron management systems

To tightly control intracellular iron levels based on iron availability, many iron uptake, utilisation and storage systems are under the regulator Fur (ferric uptake regulator) protein. The regulon of Fur is known to comprise of approximately 90 genes or more (Hantke, 2001; McHugh et al., 2003; Seo et al., 2014). When bound to its co-factor Fe^{2+}, Fur acts as a repressor. However, in conditions where Fe^{2+} availability is low, apo-Fur is no longer able to bind to DNA, causing de-repression of genes. This regulation ensures that iron uptake is modulated depending on intracellular levels of iron, with uptake being increased as iron becomes depleted. Binding of Fe^{2+} to Fur is thought to increase the binding affinity of Fur to its DNA binding site by 1,000 times (Andrews et al., 2003).

The main function of Fur is to repress iron acquisition genes, including siderophores, under conditions of iron sufficiency. The Fur regulon, however, also includes genes involved in other cellular pathways linked to iron utilisation, including respiration, flagella chemotaxis, the TCA cycle, glycolysis, methionine

biosynthesis, phage DNA-packing, DNA synthesis, purine metabolism and oxidative stress (Stojiljkovic et al., 1994; Park and Gunsalus, 1995; Vassinova and Kozyrev, 2000). Fe-S clusters, which are found in a wide variety of proteins, are assembled in *E. coli* by two systems ISC and SUF, which are differentially regulated by Fur. The ISC system which is responsible for housekeeping Fe-S synthesis is regulated by the IscR regulator independently of Fur which specifically monitors Fe-S cluster assembly status of the cell (Schwartz et al., 2001; Giel et al., 2006). The SUF system, however, responds to oxidative stress and iron starvation, and is regulated by Fur as well as OxyR during the oxidative stress response (Outten et al., 2004).

Fur has been shown to act as a positive regulator of some cellular functions. This can occur through two mechanisms, either direct binding of Fur, or through the activity of the small RNA RyhB. There are only a small number of genes that have been shown to be directly induced by Fur. These include the ferritin *pfr* gene of *Helicobacter pylori*, the OmpT porin in *Vibrio cholerae*, the *ftnA* and *acnA* genes in *E. coli* and the *fur* gene itself in *Vibrio vulnificus* (Delany et al., 2001; Lee and Helmann, 2007; Nandal et al., 2010; Craig et al., 2011; Seo et al., 2014). The sRNA RyhB is under repression by Fur-Fe^{2+}, so when iron levels are sufficient, RyhB activity is reduced. RyhB modulates the utilisation of iron by non-essential proteins, ensuring that when intracellular iron levels are low, iron utilisation is restricted to core functions (figure 1.6). As an example, the two operons implicated in Fe-S cluster formation, *suf* and *isc*, show differential regulation by RyhB, with the essential *suf* operon being unaffected by RyhB, whereas the non-essential *isc* operon was repressed (Outten et al., 2004). RyhB acts a post-transcriptional repressor, by increasing degradation of mRNA that it binds to. It does this through recruiting the multi-protein complex RNA degradosome. The degradosome breaks down both the target RNA and RyhB at the same time (Massé et al., 2003), ensuring that RyhB activity can be switched off quickly if necessary. As RyhB targets mRNA, transcription of regulated genes is unaffected, so production can quickly resume once RyhB repression is removed (Massé and Gottesman, 2002). Fur and RyhB work together to regulate

iron balance within cells, maintain intracellular iron levels and optimise iron utilisation (Massé et al., 2005).

As iron plays such a central and important role in cell function, iron homeostasis is influenced by global regulators other than Fur. *E. coli* strain BW25113 *Δcrp* mutants which cannot make cyclic AMP receptor protein (CRP), a regulator which is involved in carbon source utilisation and carbon catabolite repression (CCR), were shown to have decreased expression of enterobactin (Zhang et al., 2005). Enterobactin and salmochelin production by *E. coli* strain Nissle 1917 was also observed to be affected by carbon source, with higher amounts of both siderophores being produced in the presence of glycerol compared to glucose at pH 7 and 37°C (Valdebenito et al., 2006). Aerobactin and yersiniabactin also showed increased production in glycerol, although this was at a lower pH and temperature respectively (Valdebenito et al., 2006). These results suggest that siderophore production is under CCR, although the exact mechanism through which CRP does this is unclear.

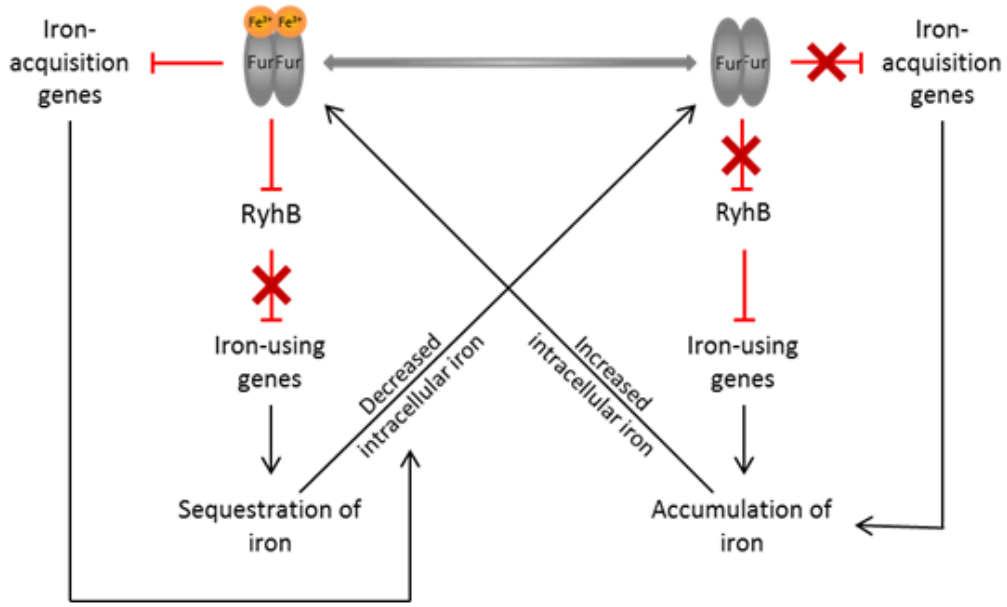


Figure 1.6: Intracellular iron regulation through Fur and RyhB activity. In iron limiting conditions, Fur is not bound to iron due to low intracellular iron levels. As a result, iron-acquisition genes and *ryhB* are de-repressed. By repressing iron-using proteins, RyhB is able to increase free intracellular iron levels. This increased free iron binds to and activates Fur, resulting in repression of *ryhB* and iron acquisition genes, thus preventing excess iron accumulating within the cell (adapted from Massé et al., 2005).

Iron also plays a large role in oxidative stress, where it can react with ROS to form the highly reactive hydroxyl radical (OH). Two key regulators in the oxidative stress response, OxyR and SoxRS, have been shown to upregulate *fur* expression (Zheng et al., 1999; Blanchard et al., 2007; Faulkner et al., 2012). Fur has also been shown to influence genes within the oxidative stress response, such as *sodAB*, which ensures that high iron levels coincide with tight control of ROS (Niederhoffer et al., 1990; Massé and Gottesman, 2002).

1.4.6 Limiting damage from iron

The partially reduced forms of oxygen, superoxide (O_2^-) and hydrogen peroxide (H_2O_2), both arise as a result of aerobic respiration. Iron can, however, react with both O_2^- and H_2O_2 through the Fenton reaction to form the HO· radical which can damage iron-sulphur clusters, DNA, and cysteine and methionine protein residues (Storz and Imlay, 1999; Imlay, 2003).

Iron reduction by superoxide: $O_2^- + Fe^{3+} \longrightarrow Fe^{2+} + O_2$

Fenton reaction: $Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + OH^- + HO\cdot$

Haber-Weiss reaction: $O_2^- + H_2O_2 \longrightarrow HO\cdot + OH^- + O_2$

In *E. coli* iron is present in the cell at 0.1mM (Outten and O'Halloran, 2001); however, only approximately 20µM is considered freely available to take place in Fenton chemistry (Keyer and Imlay, 1996). This free iron is predominantly in the more reactive ferrous form (Woodmansee and Imlay, 2002). Intracellular iron is kept to a minimum through several mechanisms, including utilising storage proteins. Storage in *E. coli* uses the FtnA (Ferritin), Bfr (bacterioferritin) and Dps proteins which bind iron as Fe^{2+} which then becomes oxidised and forms a ferric core (Andrews et al., 2003). The larger Bfr and FtnA molecules can store at least 2,000-3,000 iron atoms each, with the smaller Dps storing approximately 500 iron atoms (Andrews et al., 2003). Also, as the iron uptake regulator Fur protein is relatively abundant, it may be possible that Fur acts as a storage protein to bind free Fe^{2+} in the cytoplasm that may otherwise damage the cell (Andrews et al., 2003).

E. coli Δfur mutants have reduced total cellular levels of iron (Abdul-Tehrani et al., 1999). The majority of this iron in *Δfur* mutants is thought to be in its free form due to unregulated iron uptake and low levels of storage proteins (Keyer and Imlay, 1996; Abdul-Tehrani et al., 1999). As a result, these mutants display an increased sensitivity to redox stress reagents, although their sensitivity can be reduced via decreasing free intracellular iron (Touati et al., 1995). This can be achieved through a *tonB* mutation to prevent Fe^{3+} uptake, or overexpression of the *ftnA* ferritin gene to increase iron storage (Touati et al., 1995). Upregulation of other intracellular iron chelators, such as dipicolinate, can also protect mutants sensitive to superoxide (Maringanti and Imlay, 1999).

1.5 Introduction to research

E. coli is an important member of the microbiota that colonises a wide variety of hosts. As detailed in this introduction, several traits influencing *E. coli* fitness and colonisation of the GI-tract have been identified. However, more understanding is required of what makes an isolate a good coloniser of the gut and what facilitates its successful integration into the gut microbial community. Recent evidence indicates that the environment shapes the associated *E. coli* populations (Bergholz et al., 2011; Méric et al., 2013). Traits conferring a fitness advantage in a particular environment should therefore be enriched in the associated *E. coli* strains. We thus hypothesised that the comparison of *E. coli* isolated from the faeces of healthy hosts with isolates from the external non-host environment would reveal traits involved in gut adaptation and colonisation.

The aim of this project was to use comparative approaches to investigate traits associated with colonisation of the gut by taking advantage of two collections of natural isolates of *E. coli*, the ECOR and GMB collections. The ECOR collection (host associated strains) consists primarily of human and animal commensal strains of *E. coli* isolated from faecal samples (Ochman and Selander, 1984), whereas the GMB collection (non-host associated strains) includes isolates taken from salad crops (Méric et al., 2013), a food-safety relevant environment.

Assessing the fitness of large numbers of strains in complex environments such as the gut is extremely time-consuming and expensive. In Chapter 3, a new method for barcoding and identifying *E. coli* isolates will be described. This technique allows the parallel monitoring of several isolates in mixed *E. coli* populations in diverse environments such as the mouse GI-tract or soil columns. Competition studies carried out in the GI-tract should indicate whether there are specific strains that have persistent increased fitness, or whether dominant strains are determined largely by stochastic variations/factors. If individual strains do show increased fitness or dominance, this should enable traits and phenotypes associated with increased colonisation to be identified.

It has been hypothesised that siderophore production is an important *in vivo* fitness determinant. Indeed, enterobactin has been shown to increase fitness in the GI-tract (Pi et al., 2012). Chapter 4 will address a gap in knowledge relative to whether siderophore production in *E. coli* is influenced by the environment. Salmochelin, aerobactin and yersiniabactin have all been linked with virulence in pathogenic strains, but their presence in commensal *E. coli* strains suggest a wider role in *E. coli* lifestyle. To determine whether any siderophore system is associated with a particular environment where it possibly confers a fitness advantage, we investigated the distribution of siderophore production and related genes in the ECOR and GMB collections.

In Chapter 5, the hypotheses derived from the observed differences between the ECOR and GMB collections in terms of siderophore production and gene distribution described in Chapter 4 were investigated further. The expression of siderophore biosynthesis genes was analysed *in vitro*, under iron limitation, and *in vivo*, in the mouse GI-tract, to establish whether there was differential expression between *E. coli* strains and to determine whether additional siderophore systems to enterobactin were utilised in the host environment. Siderophore biosynthesis and *tonB* mutants were constructed to further elucidate the differences between each siderophore system and to establish whether siderophore production increases fitness in a mouse model.

2. Materials and methods

2.1 Bacterial strains

2.1.1 E. coli reference (ECOR) and GMB collections

In this study, two collections of *E. coli* natural isolates were used to investigate the phenotypic traits and genetic elements that may confer a fitness advantage within the gut, the primary environment of *E. coli*. The first collection is the well characterised *E. coli* reference (ECOR) collection which comprises 72 strains that were isolated primarily from faecal samples in USA and Europe in the early 1980s (see appendix A for full list of isolates). 61 strains were isolated from healthy human and zoo animal faecal samples (29 and 32 strains respectively), 10 strains (ECOR11, ECOR14, ECOR40, ECOR48, ECOR50, ECOR56, ECOR60, ECOR62, ECOR64 and ECOR72) were isolated from the urine of women with urinary tract infections (UTI) and one strain (ECOR 71) was isolated from an individual with asymptomatic bacteriuria. The ECOR collection was created in 1983 from a larger collection of 2,600 *E. coli* isolates, with strains being selected to represent the genotypic diversity of *E. coli* based on MLEE profiles (Ochman and Selander, 1984). The second collection, the GMB collection, comprises 96 strains isolated mostly from the aerial parts of salad crops (76 from spinach and rocket) grown predominantly in the UK during spring and summer of 2008 and 2009 (see appendix A for full list of isolates) (Méric et al., 2013). It has been shown previously that these host (ECOR) and plant (GMB) associated *E. coli* isolates display phenotypic differences that may confer a competitive advantage in their respective environments (Méric et al., 2013).

2.1.2 Additional strains

Several other strains of *E. coli* were used in this study, mainly for molecular biology, cloning and mutant generation purposes. For the development of the multiplex PCR for siderophore gene detection, the probiotic *E. coli* Nissle 1917 strain was used as it is known to have all four siderophore systems. This strain was kindly given by Ulrich Sonnenborn (Ardeypharm GmbH, Germany). The *E. coli* strain DH5α containing the pGRG36 plasmid (Addgene plasmid #16666) was

used for the construction of barcode mutants (McKenzie and Craig, 2006). To generate the siderophore knockout mutants, the *E. coli* BT340 strain containing the pCP20 plasmid as well as three *E. coli* BW25141 strains that contained the pKD46, pKD4 and pKD3 plasmids were used, all of which were kindly provided by the Coli Genetic Stock Center (CGSC, Yale University, USA).

2.1.3 Long term storage of strains, RNA and DNA

Strains were all stored as frozen glycerol stocks at -80°C. Glycerol stocks were prepared by transferring 1ml of overnight culture (LB media) into a 2ml Eppendorf tube containing 1ml 40% glycerol and freezing at -80°C. Extracted DNA and RNA were stored in water at -20°C and -80°C respectively.

2.1.4 Growth media

E. coli isolates were grown from glycerol stocks on Tryptone Bile X-glucuronide (TBX) (Oxoid) plates and incubated overnight at 37°C, unless otherwise stated. TBX plates select for bacteria that are able to grow in the presence of bile salts thus preventing the growth of most Gram-positive bacteria. Due to the presence of X-glucuronide, strains that have D-glucuronidase activity (encoded by the *uidA* gene) elicit a colour change to blue/green. The majority of *E. coli* produce D-glucuronidase, differentiating them from other coliforms that can grow on TBX plates. Mutant *E. coli* strains that possessed an antibiotic resistance cassette were grown on Lysogeny Broth (LB) plates with the required antibiotic (100µg/ml ampicillin, 100µg/ml kanamycin or 25µg/ml chloramphenicol) unless otherwise stated.

Liquid media used in this study include LB medium, Nutrient Broth (NB) (Oxoid), Super Optimal broth with Catabolite repression (SOC) and Modified M9 medium (MM9). LB medium was prepared by dissolving 10g Difco Bacto tryptone, 5g Difco Bacto yeast extract, 10g NaCl in 1 litre of water, adjusting to pH7 and autoclaving. SOC medium was prepared by dissolving 20g Bacto tryptone, 5g Bacto yeast extract, 0.5g NaCl in 1 litre of water and autoclaving. After cooling, 10mM $MgCl_2$, 2.5mM KCl and 55mM glucose (filter sterilised) were added. The MM9 medium was made according to Watts et al. (2012), firstly a 10x MM9 salt

solution was prepared by dissolving 5g NaCl, 10g NH_4Cl, 54.8g MOPS free acid, 51.0g MOPS sodium salt, 0.59g $Na_2HPO_4 \cdot 2H_2O$, 0.45g KH2PO4 in 1 litre of water and adjusting to pH 7.0. 100ml of the 10x MM9 salt solution was then mixed with 2ml 1M $MgSO_4$, 0.1ml 1M $CaCl_2$, 10ml 20% glucose solution, 20ml 10% sodium succinate solution, 10ml 2% thiamine-HCl solution, 150ml 2% casamino acid solution and then made up to 1 litre with autoclaved milliQ water. An alternative MM9 medium was used where 4ml of 50% glycerol was added as the carbon source in place of glucose.

2.1.5 Generation of mutants

2.1.5.1 Plasmid extraction and digests

The pGRG36, pKD46, pKD4, pKD3 and pCP20 plasmids were all purified from the appropriate strain for use in generating barcoded strains and siderophore knockout mutants. These were extracted from 10ml LB (with relevant antibiotic) overnight cultures of *E. coli* DH5α, BW25141 and BT340 strains using the Omega Biotek EZNA Plasmid Mini Kit I as per manufacturer's instructions for low copy-number plasmid spin protocol. Purified plasmids were quantified using a NanoDrop spectrophotometer (Thermo Scientific).

To confirm plasmid size and correct composition, restriction enzyme digests were routinely carried out and visualised on 1% agarose gels. The EcoRI enzyme was used to digest both pGRG36 and pKD46 plasmids at 37°C for 2h with the reaction mixture as follows: 10U EcoRI enzyme (Thermo Scientific), 2µl REact 3 Buffer (Thermo Scientific), 200ng plasmid DNA and water up to a final volume of 20µl. For the pKD3 and pKD4 plasmids, a double digest was performed using ApaLI and NdeI restriction enzymes (New England Biolabs). The reaction mixture was incubated at 37°C for 1h and contained 200ng plasmid DNA, 2U ApaLI enzyme, 2U NdeI enzyme, 1µg BSA (NEB), 1µl NEBuffer 4 (NEB) and was made up to 10µl using milliQ water. Finally, for the pCP20 plasmid, a digest using the AflIII enzyme (New England Biolabs) was carried out. The reaction mixture was 500ng plasmid DNA, 2.5U AflIII enzyme, 2.5µg BSA (NEB), 2.5µl NEBuffer 3 (NEB) and milliQ water up to 25µl, which was incubated for 1h at 37°C.

2.1.5.2 Preparing competent cells

To make barcoded isolates and siderophore knockout mutants, strains needed to be made competent for transformation via electroporation. LB overnight cultures of selected strains were diluted to a final OD_{600} of 0.05 in 250ml flasks containing 25ml of LB media. They were incubated at 37°C in a shaking incubator (250rpm) until they reached an OD_{600} of 0.4-0.6 and then placed on ice for 20min. Bacteria were centrifuged at 4°C for 10min at 3,226 x *g* and then washed twice with 10% filter sterilised glycerol (Sigma-Aldrich). Cells were then resuspended in 40µl of 10% glycerol, at which point they could be used for electroporation.

40µl of competent cells were mixed with 500ng of plasmid or 300ng of linear DNA insert and placed on ice for 1min. Electroporation settings were 2.5kV, 200Ω and 25µF. Immediately after electroporation, 1ml SOC medium was added to the bacteria and they were incubated for 1-3h at 37°C to recover. Bacteria were then plated out on LB plates containing the required antibiotic and incubated overnight at 37°C. For plasmids pGRG36, pCP20 and pKD46, ampicillin was used at 100µg/ml in LB plates to select for successful transformants. For the linear pKD3 and pKD4 inserts, LB plates containing chloramphenicol (25µg/ml) and kanamycin (100µg/ml) were used respectively. The pGRG36, pKD46 and pCP20 plasmids all contain temperature sensitive origins of replication, so incubation temperatures were lowered to 30°C.

2.1.5.3 Construction of barcoded strains

To be able to investigate the differences between individual isolates of *E. coli* when they are grown in mixed cultures, each strain needs to be uniquely tagged. 20 isolates of *E. coli* were selected to have a 25nt long unique sequence, called a barcode (Xu et al., 2009), inserted into their chromosomes (table 2.1). These DNA barcodes were designed for large scale genetic screening of mutations or genes; however, we adapted their use for identification of individual strains of *E. coli*. Barcodes were designed to be of homogeneous length and melting temperature to enable pooling of samples for microarrays or high-throughput sequencing (Xu et al., 2009). The strains used in this study were

selected from both ECOR and GMB collections to cover a variety of *E. coli* based on phylogenetic group (8 from Group A, 4 from Group B1, 4 from Group B2 and 4 from Group D) and phenotypic differences in biofilm formation, siderophore production and nutrient competence as previously described (Méric, 2011; Méric et al., 2013). Each barcode can be targeted by specific pair of primers allowing the amount of each individual strain in a mixed population to be quantified using real-time PCR (RT-PCR) or high-throughput sequencing.

Table 2.1: List of barcoded strains used in this study. 20 isolates representing a variety of phylogenetic groups were selected for barcode insertion. Strain GMB45 was barcoded twice with different barcodes to determine whether barcode sequence could influence transposition. All five ECOR isolates were isolated from faeces of healthy hosts as listed.

Strain	Phylogeny	Plant/Host of isolation	Barcode	Barcode Sequence
ECOR16	A	Leopard	B1	GGTACTTAAGGTTTGCCCATTCCCT
ECOR18	A	Celebese ape	B23	AGATGCACGGACTGAATTCAGCAGG
ECOR32	B1	Giraffe	23457	GGCATGAATTCGGTAACGTCACCAT
ECOR49	D	Human	B9	AACGAATTCCTGACTGGGGAGGGTG
ECOR55	B2	Human	B11	TGACGCTCACGGAATTCTAGGTAAA
GMB02	A	Rocket	23796	TAATGAATTCTCCGCCTTGGGTGAA
GMB07	B1	Spinach	B19	TGGCCTGAAACGTGAATTCAGCGTA
GMB104	A	Spinach	29884	GCCGCCATTTACGAGAATTCGTGAC
GMB16	B1	Mizuna	19427	ATCCACAGGGGCATAGGAATTCCCA
GMB18	B1	Spinach	B5	ATTGATGAATTCTGCGGTTGTCGTA
GMB23	A	Spinach	23277	TCTACCATTTTGAATTCACGCCGCA
GMB32	A	Spinach	20239	GAGGCAAGCGAATTCTGGAATCCTT
GMB34	A	Spinach	33233	CCACCGTACATCCAGAATTCTGAAT
GMB40	B2	Spinach	B6	TCACCGAATTCTGACCACTGGACTA
GMB45 (1)	B2	Spinach	19439	GTGTGCCCACGAATATGAATTCCCC
GMB45 (2)	B2	Spinach	23353	CCCTGAGAATTCAGCAGCTGACAAC
GMB48	A	Spinach	21118	TGGCTCTGTCGGAATTCGATGGGTA
GMB54	D	Spinach	33301	GCAGAGGAATTCATCCCACCTTAGT
GMB71	D	Rocket	B10	GCCGAGGCGCACGAATTCTACCCTA
GMB72	D	Rocket	B8	AGAATTCACTCCGGCTTAATTGGGC
GMB98	B2	Spinach	30438	AGGAGAATTCGTGGCGTGGTGTCAA

Insertion of barcodes into the chromosome was carried out using the Tn*7*-based transposition system as described by McKenzie and Craig (2006). Derivatives of the pGRG36 plasmid containing barcodes inserted into the multiple cloning site (MCS) within the Tn*7* transposon were transformed into the appropriate isolates. The pGRG36 plasmid contains the *tnsABCD* genes for insertion of the Tn*7* transposon at the *attTn7* site in the chromosome of *E. coli* which is located at the end of the *glmS* gene (figure 2.1). This system was selected as it has been previously shown that insertion of the transposon does not affect the phenotype (Peters and Craig, 2001; McKenzie and Craig, 2006). Insertion is highly specific as the pGRG36 plasmid lacks the *tnsE* gene which is responsible for insertion events at other sites within the chromosome which are unrelated to the *attTn7* site (Kubo and Craig, 1990).

To induce the insertion of the Tn*7* transposon into the chromosome of strains containing the pGRG36 plasmid (see section 2.5.1.2), colonies were streaked onto LB plates containing 0.1% arabinose to induce the *tnsABCD* genes which are under the arabinose inducible P_{BAD} promoter and incubated overnight at 30°C. Following induction of the Tn*7* transposon element, the pGRG36 plasmid was removed from cells by streaking colonies on LB plates and incubating overnight at 42°C. The pGRG36 plasmid contains the pSC101 *ori ts* origin of replication which functions at 32°C, but not at 42°C. Colonies were re-streaked on LB plates and incubated at 42°C a second time to ensure loss of the pGRG36 plasmid, which was confirmed by the loss of resistance to 100µg/ml ampicillin.

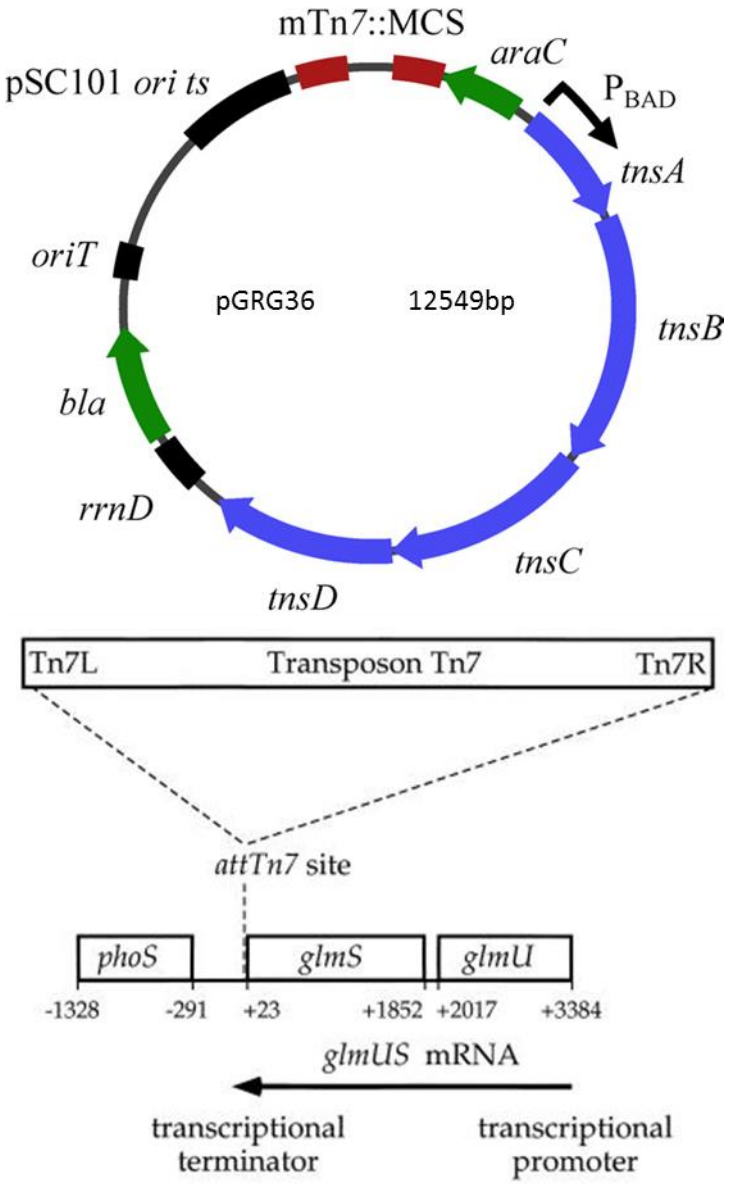


Figure 2.1: pGRG36 plasmid and Tn*7* transposon insertion into the *E. coli* chromosome. The pGRG36 plasmid multiple cloning site (MCS) contains several restriction sites, *AvrII*, *NotI*, *PacI*, *XhoI* and *SmaI*, flanked by the terminal repeats of the Tn*7* transposon. This facilitates the insertion of the gene of interest into the chromosome. The preferred site of insertion in the *E. coli* chromosome is the *attTn7* site, which is found in the transcriptional terminator of the *glmUS* genes, with a preferred orientation with the right end (Tn7R) joining close to the end of the *glmS* gene. Transposition of Tn*7* requires the *tnsABCD* genes. Transposition is regulated through the *araC* gene and the arabinose-inducible promoter P_{BAD}. pSC101 is a temperature sensitive origin of replication which ceases activity at temperature of 42°C or higher (DeBoy and Craig, 2000; McKenzie and Craig, 2006). These images are copyrighted by BioMed Central Ltd and American Society for Microbiology.

Successful insertion of the transposon and barcode was initially confirmed by colony PCRs. Colonies were first suspended in 50µl of milliQ water and heat lysed by incubation at 99°C for 10min. The PCR reaction mixture included 21µl of heat lysed bacterial suspension, 2µl of forward and reverse primers (0.4µM final concentration) and 25µl GoTaq Green Master Mix to give a final volume of 50µl (see table 2.2 for primers). Amplification was as follows: 35 cycles of 94°C for 30s, 54°C for 30s, 72°C for 1min, and 1 cycle of 72°C for 5min. PCR products were run on 1% agarose gels to determine whether the transposon had correctly inserted into the *attTn7* site. The PCR products that were of the correct predicted size were cleaned using the QIAGEN QIAquick PCR purification kit and then prepared for sequencing using the BigDye Terminator Sequencing Kit (Thermo Scientific) to determine whether the recombination had occurred without any unwanted mutations or errors. The reaction mixture was 0.5µl BigDye Terminator v3.1 Reaction Mix, 2µl 5x BigDye Buffer, 1µl primer (2µM), 40ng DNA and water to bring the final volume to 10µl (see table 2.2 for primers). The reaction settings were as follows: 1 cycle of 96°C for 1min, and 25 cycles of 96°C for 10s, 50°C for 5s, 60°C for 4min. Samples were then sent for sequencing by Eurofins MWG. Glycerol stocks were prepared for barcoded strains that showed the correct insertion had occurred.

Table 2.2: Primers used to amplify the Tn*7* transposon and barcode region of transformed cells. Primers target the region around the Tn*7* insertion site. Different forward and reverse primers were used depending on strain sequence around the *attTn7* site. Primers were used for both colony PCRs and sequencing to confirm the successful insertion of Tn*7* and the barcode.

Primer		**Sequence**
Forward	attTn7-Left	GATGCTGGTGGCGAAGCTGT
	f-attTn7	TGTGACAGAGAAAAAGTAGCC
	f-50/34	AAAGGAATAGTACACCAAAGA
Reverse	attTn7-Right	GATGACGGTTTGTCACATGGA
	r-attTn7	ATCTTCTACACCGTTCCGC

2.1.5.4 Construction of siderophore knockout mutants

To generate the required siderophore knockout mutants (see table 2.3 for list of mutants), the lambda red recombinase system was used (Datsenko and Wanner, 2000). This recombinase is very efficient and sequence replacements (e.g. a target gene by a resistance cassette) by a double-crossover can be achieved by providing short flanking homologous sequences (approximately 50nt). The lambda red system is provided by transforming strains with the pKD46 plasmid which encodes the lambda red genes *exo*, *bet* and *gam* (Murphy, 1998). These genes are under the promoter P_{araB} which is inducible by arabinose. When the lambda red recombinase is induced in the presence of arabinose, it is able to switch the gene of interest (thus removing it from the chromosome) with specially constructed linear dsDNA inserts which encode an antibiotic cassette, allowing mutants where recombination has occurred to be selected (figure 2.2). The protocol used in this study to generate mutants is a modified version of that described by Datsenko and Wanner (2000).

To prepare linear DNA inserts for recombination, primers were designed to amplify the antibiotic resistance genes present on the pKD3 (Cm) and pKD4 (Km) plasmids. Importantly, the primers included 5' extensions of approximately 50bp homologous to the flanking regions of the gene to be replaced by the antibiotic resistance cassettes on the *E. coli* chromosome (table 2.4). The PCR reaction mixture was prepared by adding 50ng plasmid DNA, 2µl of each primer (0.4µM final concentration), 25µl GoTaq Green Master Mix and milliQ water added to a final volume of 50µl (see table 2.4 for primers). Amplification for the PCR was as follows: 35 cycles of 94°C for 15s, 50°C for 30s, 72°C for 90s, and 1 cycle of 72°C for 5min. PCR products were run on 1% agarose gels for product size confirmation and cleaned using the QIAGEN QIAquick PCR Purification Kit as per manufacturer's instructions. If multiple PCR products were visible on the agarose gel, the QIAGEN QIAquick Gel Extraction Kit was used as per manufacturer's instructions to purify the correct PCR product. Clean linear DNA inserts were then quantified using a NanoDrop spectrophotometer (Thermo Scientific).

Table 2.3: List of siderophore deletion and ferric uptake mutants.

Siderophore biosynthesis gene deletion mutants:	
GMB104a	GMB104 Δ*iucABCD*; kan^r
GMB104s	GMB104 Δ*iroB*; cm^r
GMB104y	GMB104 Δ*ybtS*; kan^r
GMB104es	GMB104 Δ*entB*; cm^r
GMB104as	GMB104 Δ*iucABCD* Δ*iroB*; cm^r
GMB1104sy	GMB104 Δ*iroB* Δ*ybtS*; kan^r
GMB104ay	GMB104 Δ*iucABCD* Δ*ybtS*; kan^r
GMB104asy	GMB104 Δ*iucABCD* Δ*iroB* Δ*ybtS*
GMB104eas	GMB104 Δ*entB* Δ*iucABCD*; kan^r cm^r
GMB104esy	GMB104 Δ*entB* Δ*ybtS*
GMB104easy	GMB104 Δ*entB* Δ*iucABCD* Δ*ybtS*; kan^r cm^r
GMB91e	GMB91 Δ*entB*; cm^r
Siderophore *tonB* deletion mutants:	
GMB104t	GMB104 Δ*tonB*; kan^r
GMB91t	GMB91 Δ*tonB*; kan^r

Before introducing the linear DNA fragments, target strains were transformed with the pKD46 plasmid (section 2.5.1.2). Overnight cultures of the strains containing the pKD46 plasmid grown in LB medium with 100µg/ml ampicillin were diluted in 25ml LB to an OD_{600} of 0.005. The LB medium was supplemented with 10-50mM arabinose once the cultures reached an OD_{600} of 0.1 to induce recombinase expression. Cultures were then incubated at 37°C in a shaking incubator (250rpm) until they reached an OD_{600} of 0.4-0.6. The cells were then made electrocompetent as described in section 2.1.5.2, and transformed with linear DNA fragments. Colony PCRs were carried out on transformed colonies to determine whether recombination had successfully occurred as described in section 2.1.5.3. The PCR products that were of the correct predicted size were cleaned and prepared for sequencing as described in section 2.1.5.3. Sequences were compared to known sequences for each siderophore gene as well as the resistance cassettes in both pKD4 and pKD3. Glycerol stocks were prepared for mutant strains that showed correct insertion.

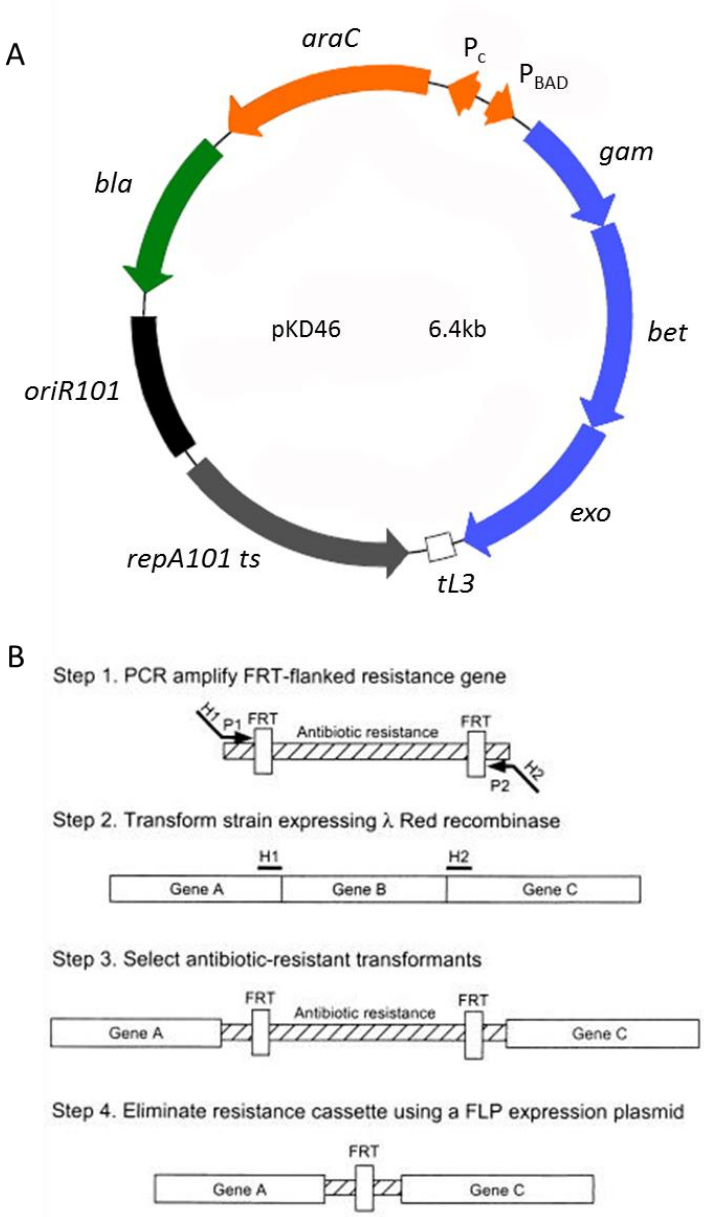


Figure 2.2: The pKD46 plasmid and lambda red recombinase system. A) pKD46 encodes the lambda red genes *gam*, *bet* and *exo*, with the native terminator *tL3*, which are inducible by arabinose via the promoter P_{BAD}. It also possesses the *bla* gene for ampicillin resistance and the temperature sensitive origin of replication *repA101ts*. B) A linear DNA insert is constructed using the antibiotic resistance cassette on plasmids pKD3 and pKD4, chloramphenicol and kanamycin respectively, and primers possessing extensions that are homologous to the gene of interest. Recombination occurs via the lambda red recombinase, replacing the gene of interest with the resistance cassette. The pKD3 and pKD4 resistance cassettes are both flanked by FRT sites, enabling removal using a Flp recombinase. P1 and P2 labels denote primer binding sites. H1 and H2 refer to the chromosome homologous extensions or regions (Datsenko and Wanner 2000). Image B is copyrighted by National Academy of Sciences.

When it was necessary to delete more than two genes from the same strain, we first removed the kanamycin and chloramphenicol resistance cassettes inserted in the genome. To do so, we exploited the fact that both resistance genes were flanked by FRT sites, which are recognised by the site-specific Flp recombinase. Recombination results in the excision of the resistance gene. The Flp recombinase was provided by transforming the target strains with the pCP20 plasmid (Cherepanov and Wackernagel, 1995; Datsenko and Wanner, 2000). Removal of these antibiotic resistance cassettes enabled further genes to be knocked out using pKD4 and pKD3 linear DNA inserts as mutants no longer had antibiotic resistance. After transformation with the pCP20 plasmid, the strains were incubated at 30°C overnight on LB and ampicillin (100µg/ml) to induce the Flp recombinase and then restreaked on LB plates and incubated at 42°C to ensure loss of the pCP20 plasmid. Loss of the plasmid was confirmed by streaking colonies on LB and ampicillin plates and the Flp-FRT recombination was confirmed by sequencing.

Table 2.4: Primers used for the construction of siderophore deletion mutants. (Next page) Deletion primers target the resistance cassette of both pKD3 and pKD4 plasmids. This PCR was used to generate linear DNA for the transformation and recombination of strains. Screen primers were used in colony PCRs of transformed strains to confirm successful recombination. Deletion primer extensions homologous to the *E. coli* chromosome are underlined. The deletion primers are from Watts et al. (2012), with the exception of the *entB* and *tonB* primers which were designed for this study. All screen primers were designed for this study.

Siderophore	Gene	Primer	Sequence
Aerobactin	*iucABCD*	deletion F	GAGCTGTTTGATTATGATCCTGCCCTCTGAAAAATCCGCCACAGATGTGGGTGTAGGCTGGAGCTGCTTC
		deletion R	GCGGGCGGCATACTGAGATCGAATAAATCACGTCCCATTACGCGATTAAGCATATGAATATCCTCCTTAG
		screen F	TTCTGACAATAACATTTCTCGTTGA
		screen R	ACACCACGAACGTTTCATCA
Yersiniabactin	*ybtS*	deletion F	AGGTTAGAAAACAGTTACTCCTACACCATTAAATAGGGCGCAATGCTCGCGTGTAGGCTGGAGCTGCTTC
		deletion R	CCGGCTGCATTCGGCAAGAGAACTGATGACAAGCGCAATTGATTGATCCGCATATGAATATCCTCCTTAG
		screen F	CAATGTACTGATGCGGCAAT
		screen R	GGTCGGATTATCTGGAAGCA
Enterobactin	*entB*	deletion F	GAAAGTCGATAAAAAACAATTACGTCAGTGGCTGGCGTCACGCGTGTAGGCTGGAGCTGCTTC
		deletion R	TGAAATCCATTATTTCACCTCGCGGGAGAGTAGCTTCCACCACATATGAATATCCTCCTTAG
		screen F	TGAGCATGGAAGATGAGCTG
		screen R	GTCGCAAAGGGATATTGCTC
Salmochelin	*iroB*	deletion F	TCTGTAAAATACGATCCACTGGCCGGATCGTTCCGCAAAAAAGCCAGCACGTGTAGGCTGGAGCTGCTTC
		deletion R	TGCGTCGACTGCCTGATTTAGATCGTCAAGCGGAGAGGGATTTTCTCATGCATATGAATATCCTCCTTAG
		screen F	GAGCTGTCCATAACGCTGGT
		screen R	TGTGATTCGCAGGCATTAAG
Ferric iron uptake	*tonB*	deletion F	GGCGAAGATCTGCAACGGAAAGATGATGTCTTTGTTAAGGCCATGCATAAGTGTAGGCTGGAGCTGCTTC
		deletion R	CCTTACCTGTTGAGTAATAGTCAAAAGCCTCCGGTCGGAGGCTTTTGACTCATATGAATATCCTCCTTAG
		screen F	GCGTTTTTCGAGGCTATCAG
		screen R	AAGTATGTCGCGGTTGATCC

2.1.6 Characterisation of mutants

When grown under laboratory conditions, bacteria can rapidly accumulate adaptive mutations which may cause phenotypic changes (Spira et al., 2011; Eydallin et al., 2013). As one of the main aims of this study was to investigate the differences in fitness of *E. coli* strains based on their source of isolation, it was important to assess whether the genetically manipulated strains had characteristics associated with domestication to the laboratory environment. Several assays were therefore performed to assess catalase activity (RpoS function), growth patterns, colony morphology and additional phenotypes such as nutritional competence.

2.1.6.1 Growth curves

Culturing bacteria within the laboratory often involves growing bacteria in nutrient rich media designed to achieve maximal growth rates (Neidhardt, 1999). As a result, it is likely that *E. coli* adapt to this new laboratory habitat, optimising growth within these media and displaying altered growth rates compared to ancestor strains (Eydallin et al., 2013). To screen for laboratory adaptation, barcoded isolates were grown in triplicate alongside their wild-type counterpart in LB medium. LB overnight cultures were used to inoculate 25ml LB in 250ml conical flasks to a final OD_{600} of 0.005. Siderophore mutants were grown in triplicate alongside the wild-type in MM9 media both with and without the iron chelator 2,2'-dipyridyl (Sigma-Aldrich) to induce low iron conditions. Strains that showed reduced growth in MM9 media were grown in MM9 supplemented with ferrous sulphate ($FeSO_4$) to determine whether reduced growth was due to reduced iron uptake. Overnight cultures grown in LB with the relevant antibiotic were centrifuged at 9,240 x *g* for 5 min at room temperature (RT) and washed using PBS to remove residual iron present in the LB medium. Resuspended cells were then used to inoculate 25ml of MM9 media with or without either 100µM 2,2'-dipyridyl or 100µM iron(II) sulphate heptahydrate ($FeSO_4 \cdot 7H_2O$) to a final OD_{600} of 0.05. Cultures were all incubated at 37°C in a shaking incubator and OD_{600} readings taken every hour to monitor growth.

2.1.6.2 Catalase assays

RpoS is a sigma factor (σ^{38}) that is responsible for regulating the general stress response by reducing growth and increasing resistance to stressors, such as nutrient starvation, extreme temperatures, osmotic shock, pH, oxidative stress and other environmental fluctuations that may damage the cell (Battesti et al., 2011). RpoS mutations have been observed in natural and laboratory populations of *E. coli* (Notley-McRobb et al., 2002) and also readily occur within *in vitro* cultures of *E. coli* (Farrell and Finkel, 2003). Bacteria that are grown under nutrient limitation have been shown to have increased growth rates when they possessed an RpoS that was non-functioning (Notley-McRobb et al., 2002). RpoS regulates the enzyme catalase (KatE) which catalyses the breakdown of H_2O_2 to water and oxygen to prevent damage to the cell. A simple catalase assay was used to determine whether mutants had retained catalase activity, and thus RpoS function, which would greatly affect their phenotype. The assays were performed by picking single colonies of each mutant as well as the wild-type for that strain. The colonies were transferred to a plastic petri dish and 20µl of H_2O_2 was applied to each colony. Catalase positive colonies produce oxygen which forms as bubbles that can be identified visually.

2.1.6.3 YESCA plates

Some laboratory domesticated *E. coli* strains have been shown to possess altered curli fimbriae formation (Eydallin et al., 2013). To assess changes in colony morphology and curli formation, barcoded and parental isolates were grown on YESCA agar plates. This agar contains two dyes, congo red and coomassie brilliant blue G, which stain curli. YESCA plates were made as per Hammar et al. (1996), with 10g casamino acids, 1g yeast extract and 20g Bacto agar dissolved in 1 litre water and autoclaved. 20mg congo red and 10mg coomassie brilliant blue G were added after autoclaving. Both barcoded and parental *E. coli* strains were streaked out from glycerol stocks onto YESCA plates and incubated for 48h at 37°C and 7 days at RT. Whole plate images were taken using a Panasonic Lumix DMC-F25 digital bridge camera. Individual colonies were imaged using a Leica M165C Stereo microscope using the software Leica

Application Suite (LAS) V4.2. Parental and barcoded strains colony morphologies were visually compared to determine whether any changes had appeared.

2.1.6.4 BIOLOG plates

To test nutritional competence of barcoded strains compared to the wild-type, BIOLOG GN2 microplates were used to profile carbon source utilisation. GN2 microplates are 96 well plates where each well contains a unique carbon source, as well as a blank control. These 95 carbon sources can be used to generate a metabolic profile and identify aerobic gram-negative bacteria. Colonies of both mutants and the wild-type were suspended in BIOLOG GN inoculating fluid and bacterial suspensions used to inoculate the GN2 microplates which were incubated at 37°C. Each well contained tetrazolium violet (2,5-diphenyl-3-(1-naphthyl)tetrazolium chloride), a redox dye, which generates a purple colour in response to the production of NADH generated during metabolic activity, and thus carbon source utilisation (Bochner and Savageau, 1977; Bochner, 2009). OD_{600} for the plates was measured at 24h, 48h and 72h incubation to confirm any differences observed were a result of metabolic changes rather than differences in growth rate.

2.2 Barcode mixed population competition studies

2.2.1 Preparation of mixed strain inocula

To prepare mixed cultures of barcoded *E. coli*, strains were initially grown individually and then mixed together at the start of the mixed culture. LB overnight cultures were used to inoculate 10ml of LB to an OD_{600} of 0.005 for each strain to be used in the mixed culture. These were then incubated at 37°C in a shaking incubator until they reached an OD_{600} of 0.2 at which point the individual cultures were mixed to give an equal number of each strain according to OD_{600}. To prepare mixed populations for use in soil and mice experiments, LB overnight cultures of barcoded strains were used to make mixed cultures. The barcoded strains for soil and mouse experiments displayed different growth rates (data not shown) that made it not feasible to use log cultures to prepare mixed cultures. Prior to carrying out competition studies, overnight cultures

were spread on TBX plates and incubated overnight at 37°C to confirm that OD_{600} accurately represented bacteria number. The OD_{600} absorbance of each overnight culture was measured and 4 OD_{600} of each culture harvested and mixed together to get an equal number of each strain in the mixed culture. Cells were then centrifuged at 3,226 x *g* at RT for 10min and washed twice using 10ml PBS. The OD_{600} was measured after the second wash and bacteria were resuspended in the required amount of PBS to give OD_{600} = 2 (approximately 1-2x10^9 CFU/ml). These bacteria could then be used to inoculate soil or given to mice via gavage.

2.2.2 In vitro competition studies

To test the ability to quantify and identify the amount of each barcoded strain in a mixed population, an *in vitro* competition study was carried out in LB medium. The eight strains GMB71-B10, GMB34-33233, ECOR49-B9, ECOR55-B11, GMB07-B19, GMB45-23353, GMB45-19439 and ECOR16-B1 were used to create a mixed *E. coli* inoculum as described above (section 2.2.1). The inoculum was then used to initiate an LB culture, which was grown at 37°C in a shaking incubator and sampled every hour for the first 8h followed by samples taken at 24h and 44h (98h in the second culture). During sampling OD_{600} was measured and 500µl was taken for gDNA extraction. These 500µl were centrifuged for 5min at 9,240 x *g* at RT and the pellet stored at -20°C prior to gDNA extraction. To extract gDNA the QIAGEN Blood and Tissue Kit was used as per manufacturer's instructions for gram-negative bacteria. gDNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific) and then used in RT-PCR.

2.2.3 Mouse GI-tract competition studies

To investigate whether barcoded strains could be monitored in the gut and determine which strains had increased fitness and competitiveness in the GI-tract, mixed cultures were given to mice via oral gavage. Wild-type female C57BL/6 mice were used for the competition experiments, and all were 4-8 weeks old. All mice were bred and maintained at a conventional animal unit at the University of East Anglia (UEA). Mice were specific pathogen-free (SPF) and had access to a standard chow mouse diet and sterilised water *ad libitum*. Mice

were given the antibiotics ampicillin and neomycin (1mg/ml of each in drinking water) for 1 week prior to gavage to enable bacteria to colonise the gut. *E. coli* has been shown to colonise poorly in conventional mice that have an intact microbiota due to colonisation resistance (van der Waaij et al., 1972; Freter et al., 1983), so removal of the microbiota by antibiotics greatly facilitates colonisation. Ampicillin is active against many Gram-positive bacteria and some Gram-negatives, especially *Enterococci*, *Streptococcus* and some *Enterobacteriaceae*. Neomycin is particularly effective against Gram-negative and some Gram-positive bacteria. Together these antibiotics are able to remove a significant amount of the resident microbiota that would otherwise prevent *E. coli* from colonising (van der Waaij et al., 1972; Vijay-Kumar et al., 2010). During the antibiotic treatment mice remained on the standard chow diet. 24 hours before the bacterial challenge, mice were taken off the antibiotics and returned to standard water. Mice were then given 10^8 CFU of *E. coli* by single oral gavage. Faecal samples were also collected when the gavage was given and spread on TBX plates to confirm the absence of *E. coli* within the GI-tract of the mice.

For the first mouse competition study the eight strains ECOR49-B9, ECOR16-B1, GMB45-19439, ECOR55-B11, GMB45-23353, GMB34-33233, GMB07-B19 and GMB71-B10 were used. For the second experiment 13 strains, GMB23-23277, GMB32-20239, GMB34-33233, GMB98-30438, GMB45-19439, GMB45-23353, GMB104-29884, GMB72-B8, GMB18-B5, GMB71-B10, GMB54-33301, GMB07-B19 and GMB02-23796, were used to expand on what had been seen previously with the first mouse experiment and to compare with the soil competition study. Following gavage faecal samples were collected on the following days for mouse experiment 1: 1, 2, 3, 9, 16 and 22. For this first experiment, there was one cage of 4 mice which were all given the same gavage. As the mice were cohoused, it is possible that the microbiota was shared between them as mice exhibit coprophagia, consumption of faecal matter (Barnes et al., 1963). Mice were then euthanized on day 27 using schedule 1 techniques (carbon dioxide and dislocation of the neck) and GI-tract contents (ileum, caecum and colon) collected. For the second experiment, two groups of 5 mice were given the

same gavage, to account for the possibility of coprophagia. Faecal samples for both groups of mice were collected on days 1, 2, 3, 4, 6, 8, 10 and 12. Further samples were collected on days 14, 18, 22, 26 and 30 for the first group only. The second group had faecal samples taken up to day 12, but on day 14 they were euthanized using schedule 1 techniques (carbon dioxide and dislocation of the neck) and the GI-tract contents (ileum, caecum, proximal colon and distal colon) were collected to determine whether strains were evenly distributed across the gut or whether they inhabited different areas which would reduce direct competition between different *E. coli* strains. The mucosae from the proximal and distal colon were also collected to determine whether there were any differences between *E. coli* composition in the lumen and that associated with the mucosa.

Two samples were collected from each mouse at each time point, with one sample being used to determine CFUs and one for gDNA extraction. Faecal samples for CFUs were homogenised in 1ml PBS on the day of collection. A dilution series was made by diluting the homogenised faeces in PBS in factors of 10. These dilutions were then spread on TBX plates and incubated overnight at 37°C. Colonies were counted the following day and the CFU/g calculated. gDNA samples were initially stored at -80°C for gDNA extraction at a later date. gDNA was extracted using the QIAGEN QIAamp DNA Stool Mini Kit as per manufacturer's instructions. Faecal samples were homogenised and cells lysed using a FastPrep FP120 benchtop homogeniser (Thermo Savant) using the following settings: 4 runs of speed 6.5 for 45s with 60-90s on ice in between each run, and then manufacturer's instructions were followed. For the second experiment, the QIAGEN QIAamp Fast Stool Mini Kit was used as per manufacturer's instructions but a fast-prep was not required for increased DNA yields. Extracted gDNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific) and used for either RT-PCR or further preparation for high-throughput sequencing.

2.2.4 Soil competition study

To assess the viability of using and detecting barcoded strains in natural environments, a mixed culture was used to inoculate columns of soil. The following 12 strains were used in the soil experiment: GMB23-23277, GMB32-20239, GMB34-33233, GMB98-30438, GMB45-19439, GMB104-29884, GMB72-B8, GMB40-B6, GMB18-B5, GMB71-B10, GMB54-33301 and GMB07-B19. Soil was kindly provided by Philip Poole (John Innes Centre, UK). The method for the soil competition study was adapted from Williams et al. (2007). Soil was placed in 50ml falcon tubes and had a moisture content of 15%, with 40g of soil in each tube at a density of $1gcm^{-3}$. 2ml of three dilutions of bacteria (10^7, 10^8 and 10^9 CFU/ml) were applied to the top of the soil column. The lids of the tubes were left loose to ensure adequate oxygen levels at the soil surface and then soil columns were incubated at 20°C in the dark over a period of 24 days. Soil columns were divided into three sections (top, middle and bottom) and 0.25g soil used for gDNA extraction and 0.5g used to make a 10x dilution series using PBS which was plated out on TBX plates and incubated overnight at 37°C to estimate CFUs. gDNA extraction from soil samples was performed using the PowerSoil (MO BIO) kit as per manufacturer's instructions. To increase DNA yield, a FastPrep FP120 homogeniser (Thermo Savant) was used with the following settings: 4 runs at speed 4 for 30s with 1min on ice in between each run. Extracted gDNA was then cleaned as contaminants, such as humic acids, are often coextracted from soil alongside DNA using the OneStep PCR Inhibitor Removal Kit (Zymo Research) as per manufacturer's instructions. gDNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific) and then used in RT-PCRs.

2.2.5 Quantification of barcoded strains using RT-PCR

Primers for the RT-PCR were designed using the Primer3 software (http://frodo.wi.mit.edu/) based on the barcode sequence and on part of the Tn*7* transposon region. The reverse primer was designed to contain the whole barcode sequence as well as part of the surrounding Tn*7* region (table 2.5). The forward primer was designed to be within the Tn*7* region common to all

barcoded strains. The amplicon sizes were designed to be about 150bp in length to ensure optimum detection during the RT-PCR. To determine whether there were any biases in the primers in terms of efficiency, standard curves were performed. gDNA was extracted from LB overnight cultures of each barcode strain used in the competition studies using the QIAGEN DNeasy Blood and Tissue Kit as per manufacturer's instructions. This gDNA was then diluted using milliQ water to give a concentration of 10ng/µl and a dilution series prepared from this down to a dilution factor of 10^{-8}. 10µl of each dilution was used for each reaction (100ng of gDNA for the highest concentration). The reaction mixture per well was 12.5µl SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich), 0.25µl Reference dye (in kit), 0.05µl primer 1 (0.2µM final concentration), 0.05µl primer 2 (0.2µM final concentration), 10µl gDNA, and water up to a final volume of 25µl. Reactions were carried out in a 96 well plate which was centrifuged prior to the RT-PCR at 2,760 x *g* for 1min at RT to remove air bubbles that might interfere with signal detection. The RT-PCR reaction settings were 1 cycle of 94°C for 2mins, 40 cycles of 94°C for 30s, 60°C for 30s, 72°C for 45s, 1 cycle of 72°C for 5mins, and 1 cycle of dissociation curve (1 cycle of 95°C for 15s, 60°C for 1min, 95°C for 15s and 60°C for 15s). Fluorescence was measured during the 45s elongation step at 72°C. RT-PCRs were performed on the ABI TaqMan 7500 System. ROX indicator settings were used for detection of the reference dye used in this study, which was provided with the Green JumpStart Taq ReadyMix (Sigma-Aldrich). The standard curves were used to determine how the amount of gDNA translated into Ct values. Assuming 100% efficiency of primers, for a 10x dilution series the Ct value is expected to increase by 3.3 for every 10-fold reduction of gDNA. Primers with standard curves that had gradients between 3.0-3.6 (90%-110% efficiency) were deemed suitable for use in the RT-PCR.

The measured sensitivity for the RT-PCR using purified gDNA diluted in milliQ water (for the standard curves) was 0.01pg gDNA per reaction. To ensure that the sensitivity was not severely altered by the presence of gDNA from other bacteria or other factors present in samples isolated from the mouse gut,

barcode gDNA was mixed with faecal gDNA extracted from wild-type C57BL/6 mice which did not have an altered microbiota kindly provided by Ida Porcelli (Institute of Food Research, UK). Mouse faecal gDNA was extracted using the QIAGEN Stool Kit as per manufacturer's instructions. Increasingly diluted barcoded gDNA was added to a constant concentration of faecal gDNA (10ng/µl) and then used in an RT-PCR (reaction mixture and settings as for the standard curves) to generate a standard curve to determine primer efficiency and sensitivity. Primer efficiency was unchanged from the standard curves using pure barcode gDNA. Sensitivity was still as low as with pure gDNA with 0.01pg required for an accurate measure of gDNA quantity.

Table 2.5: RT-PCR primers used for the detection of barcoded strains. The transposon P3 primer was used as the forward primer for all RT-PCR reactions and targeted a region of the Tn*7* transposon common to all barcoded strains. All other primers were unique to a specific barcode strain.

Strain	Barcode	Primer Sequence
GMB45 (1)	19439	GGCCGCGGGGAATTCATATTCGTGGGCACAC
GMB32	20239	GGCCGCAAGGATTCCAGAATTCGCTTGCCTC
GMB48	21118	GGCCGCTACCCATCGAATTCCGACAGAGCCAC
GMB23	23277	GGCCGCTGCGGCGTGAATTCAAAATGGTAGAC
GMB45 (2)	23353	GGCCGCTGTTGTCAGCTGAATTCTCAGGGTTC
ECOR32	23457	GGCCGCGGCATGAATTCGGTAACGTCACC
GMB02	23796	GGCCGCTTCACCCAAGGCGGAGAATTCATTAC
GMB104	29884	GGCCGCGTCACGAATTCTCGTAAATGGCGGC
GMB98	30438	GGCCGCTTGACACCACGCCACGAATTCTCCTC
GMB34	33233	GGCCGCATTCAGAATTCTGGATGTACGGTGGC
GMB54	33301	GGCCGCACTAAGGTGGGATGAATTCCTCTGC
ECOR16	B1	GGCCGCAGGGAATGGGCAAACCTTAAGTACC
GMB18	B5	GGCCGCGATTGATGAATTCTGCGGTTGTCGTGC
GMB40	B6	GGCCGCGTAGTCCAGTGGTCAGAATTCGGTGAGC
GMB72	B8	GGCCGCGCCCAATTAAGCCGGAGTGAATTCTGC
ECOR49	B9	GGCCGCGCACCCTCCCCAGTCAGGAATTCGTTGC
GMB71	B10	GGCCGCGTAGGGTAGAATTCGTGCGCCTCGGC
ECOR55	B11	GGCCGCGTTTACCTAGAATTCCGTGAGCGTCAGC
GMB07	B19	GGCCGCGTACGCTGAATTCACGTTTCAGGCCAGC
All	Transposon P3	TGCCCGTCGTATTAAAGAGG

For competition study samples, 100ng (concentration of 10ng/µl) of gDNA was used for each reaction, which were carried out in duplicate or triplicate. The RT-PCR reaction mixture and settings were as described for the primer standard curves (see above). Standard curve line equations were used to determine how much of each barcode strain was in a sample. The proportion of each barcode strain in a sample was calculated to determine the population dynamics.

2.2.6 High-throughput sequencing for detection of barcoded strains

For high-throughput sequencing, an amplicon based system was used. Samples were prepared using amplicon primers designed to target the Tn*7* region around the barcode common to all barcoded isolates, so that primers could amplify all barcode strains at the same time (table 2.6). The amplicon primers included the Illumina flowcell adapter sequences for use on the Illumina platforms. Reverse primers contained a 6nt unique index sequence that enabled pooling of up to 48 samples. The PCR reaction mixture was 200ng template DNA, 1U phusion DNA polymerase (NEB), 10µl 5x buffer (contains 1.5mM $MgCl_2$ final concentration), 1µl dNTPs (200µM final concentration), 2µl each primer (0.4µM final concentration) and water (Sigma-Aldrich) up to a final volume of 50µl (see table 2.6 for primers). Amplicon PCR settings were 1 cycle of 95°C for 30s, 35 cycles of 94°C for 15s, 55°C for 30s, 72°C for 45s and 1 cycle of 72°C for 5min. PCR products were run on a 1.5% agarose gel to confirm successful amplification and correctly sized amplicons purified using the QIAGEN QIAquick PCR purification kit as per manufacturer's instructions. Samples were then sent to TGAC (The Genome Analysis Centre, UK) for size fractionation and sequencing. A 100bp SE (single-end) read was performed on the Illumina HiSeq. From the sequencing data, the Unix grep command was used to count the frequency of each barcode and determine the proportion of each barcoded strain in each sample.

Table 2.6: Amplicon PCR primers and index sequences. Forward and reverse amplicon primers contain adaptor sequences for use in the Illumina MiSeq or HiSeq platforms (underlined). The remainder of these primers is homologous to the Tn*7* region surrounding the barcode, which can be used on all barcoded strains, generating a 300bp amplicon. The reverse amplicon primer contains a 6nt unique index sequence which enables identification of individual samples when pooled (highlighted in red). The index sequences listed are those used for the first group of mice from mouse competition study 2. The sequencing primer targets the adaptor sequence in the forward amplicon primer. Only one sequencing primer was used as single-end (SE), rather than paired-end (PE) reads were performed. The index primer targets the adaptor region in the reverse amplicon primer.

Amplicon primers	**Sequence**		
Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCTTGCCCGTCGTATTAAAGAGG		
Reverse	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTAAATGATATCGGATCCTAGTAAGC		
Index sequences			
	CGTGAT	TCCGCG	CGGATA
	GTGTCG	GGACTA	CCAGAG
	AGACCT	TCGAAC	TGTGCT
	TACCAA	AGATGA	CACTCG
	CCGAGC	CTGGCA	
	CATTGT	TTCAGG	
	GCGCTA	GACAGT	
	TGAACT	ACTTAG	
	GGATGC	ATTCGA	
	ATCATA	GCGTAT	
	AATGGA	GAACGC	
Illumina primers	**Sequence**		
Sequencing	ACACTCTTTCCCTACACGACGCTCTTCCGATCT		
Indexing	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC		

2.3 Identification of siderophore genes by multiplex PCR

2.3.1 ClustalW multiple sequence alignment

Following the initial observation that siderophore production is increased in faecal strains of *E. coli* compared to plant isolates (Méric, 2011), the presence or absence of siderophore genes within the ECOR and GMB collections was assessed. A multiplex PCR was designed that could detect four biosynthesis and one receptor gene for all four siderophore systems in *E. coli*. The salmochelin

locus, however, only includes one biosynthesis gene, so export and degradation genes were included in the PCR. To ensure accurate identification of siderophore gene presence, alignments of siderophore gene sequences were performed to find conserved regions that could be targeted by the multiplex PCR. Publically available whole genome sequences from Ecocyc (Keseler et al., 2013) and NCBI were used to construct these alignments. Initial alignments were performed in Ecocyc using the in-built multi-genome alignment software. An nBLAST (nucleotide Basic Local Alignment Search Tool) was then performed using default parameters on the NCBI website to check siderophore regions and get more sequences for alignments as the NCBI has a greater collection of sequences. A Perl script was used to manipulate the output from the NCBI nBLAST alignment so that it could be used in the Bioedit Sequence Alignment Editor software (Hall, 1999). A multiple alignment was then performed using ClustalW v1.4 (Thompson et al., 1994). These alignments highlighted conserved regions for primer design. Primers were designed using the Primer3 software (http://frodo.wi.mit.edu/). PCR product sizes were designed to be of different lengths within one siderophore system to ensure that bands could be distinguished on a gel.

2.3.2 Multiplex PCR

Template DNA was extracted from overnight cultures using the QIAGEN DNeasy Blood and Tissue extraction kit as per manufacturer's instructions. 10ng of template DNA was used for each multiplex PCR in a 25µl reaction volume containing 12.5µl Go-Taq Green MasterMix (Promega) and 0.1µM of each primer (table 2.7). Amplification for each PCR was as follows: 35 cycles at 95°C for 30s, 55°C for 30s, 72°C for 1 min, and 1 cycle at 72°C for 5 min. Yersiniabactin and aerobactin multiplex PCRs had slight alterations, with the annealing temperature raised to 60°C for the yersiniabactin PCR and elongation step shortened to 40s for the aerobactin PCR. Multiplex PCR products were run on a 1% agarose gel to determine siderophore gene presence or absence. Strains where all five genes (4 biosynthesis genes and 1 receptor) were detected for a specific siderophore were considered to possess that siderophore system.

Table 2.7: Multiplex PCR primers for detection of all four siderophore systems in *E. coli*.

Siderophore	Gene	Primer Sequences		Product Size
Enterobactin	*entA*	F	GTGCGCTGTAATGTGGTTTC	184
		R	CAGAGGCGAGGAACAAAATC	
	entB	F	GCGACTACTGCAAACAGCAC	382
		R	TTCAGCGACATCAAATGCTC	
	entC	F	GACTCAGGCGATGAAAGAGG	438
		R	TGCAATCCAAAAACGTTCAA	
	entE	F	CGTAGCGTCGAGATTTGTCA	776
		R	CCCATCAGCTCATCTTCCAT	
	fepA	F	TTTGTCGAGGTTGCCATACA	349
		R	CACGCTGATTTTGATTGACG	
Salmochelin	*iroB*	F	CAACCATCGGTTTGACAGTG	166
		R	GACGTAACACCGCCGAGTAT	
	iroC	F	TGCCACACAGGATTTTACCA	388
		R	CTCACTCTGGGTGCAGCATA	
	iroD	F	GGTAAGCAGTTGTCCGGTGT	227
		R	GTTACTGCGGCTCCTATTCG	
	iroE	F	ATCATAACCTCTGCCCAACG	300
		R	ACCAACCTCCCTTTCGATCT	
	iroN	F	CTTCCTCTACCAGCCTGACG	648
		R	GCTCCGAAGTGATCATCCAT	
Yersiniabactin	*irp1*	F	AGAGCGGAAATAACCGAACA	221
		R	GTAAACAGGCCGTGACGATT	
	irp2	F	CTGGTGATGGTGATGGAAAA	247
		R	CCATCGCGATAAATTGTCCT	
	irp3	F	GTATACCTCGCCGGAACAGA	177
		R	GCCAGCGTTTGTAAGGAACT	
	irp4&5	F	GCGCCACAAGGACTGATTAT	905
		R	GTCTCTCCAGCGACCAGAAC	
	fyuA	F	GGGAATGTGAAACTGCGTCT	791
		R	CGGGTGCCAAGTTCATAGTT	
Aerobactin	*iucA*	F	ATAAGGGAAATAGCGCAGCA	212
		R	TTACGGCTGAAGCGGATTAC	
	iucB	F	CCACGAATAGTGACGACCAA	339
		R	GTTTTTGATGCAGAGCGTGA	
	iucC	F	ATTTCGGGAAACGCTTCTTT	158
		R	GTGGTTCCGCTGTATCACCT	
	iucD	F	TCTTCCTTCAGTCCGGAGAA	630
		R	TCCTCATTTTTCCTGGCATC	
	iutA	F	CCAGCCTCAAACTCCATCAT	157
		R	ACAGCCGACAACTGGACTCT	

2.4 Chemical assays to determine siderophore production *in vitro*

2.4.1 Liquid CAS assay

To confirm observed increased siderophore production on Chrome Azurol S (CAS) agar plates (Méric, 2011), a CAS liquid assay was performed on a subset of ECOR and GMB strains. This assay, unlike CAS agar, is not influenced by potential differences in diffusion by individual siderophores. Liquid CAS assays were also used to characterise siderophore production in siderophore biosynthesis and *tonB* mutants. Chrome Azurol S liquid (CAS) assay solution was prepared as per Payne (1994). 2mM CAS solution (0.121g CAS in 100ml water) and 1mM Fe stock solution (1mM $FeCl_3 \cdot 6H_2O$ in 10mM HCl) were made and mixed at a ratio of 5:1. Hexadecyltrimethylammonium bromide (HDTMA) solution (0.0219g HDTMA in 50ml water) and piperazine buffer (4.307g piperazine in 30ml water and add HCl to bring pH to 5.6) were then made and 9ml of the CAS-Fe mixture was added to the HDTMA solution. Finally the piperazine buffer was added alongside milliQ water to bring the volume up to 100ml. A shuttle solution, 0.2M 5-Sulfosalicylic acid, was used to enhance the colour change.

Single colonies were used to inoculate 10ml of MM9 media and grown in a shaking incubator at 37°C for 16h. OD_{600} absorbance was measured and then cultures were centrifuged at 9,240 x *g* for 5min at RT and the supernatant filtered (0.2µm filter) to remove bacteria. 0.5ml of supernatant was added to 0.5ml CAS assay solution, followed by 10µl of shuttle solution. These were mixed by inversion and colour allowed to develop for 10mins. Siderophores present in the supernatant will cause the CAS solution to change in colour from blue to orange by removing iron from the dye complex. A spectrophotometer was then used to measure absorbance at OD_{600} using the MM9 medium as a blank and MM9 medium with CAS and shuttle solutions added as a reference measurement. The amount of siderophore present in the supernatant was

calculated using the following equation and then standardising to culture OD_{600} readings as previously described (Payne, 1994; Watts et al., 2012):

$[(A_{reference} – A_{sample})/A_{reference}] \times 100 = \%$ siderophore units

2.4.2 Csàky assay

To specifically assess the production of aerobactin by siderophore biosynthesis mutants, a Csàky assay was performed which can measure hydroxamate in used media (Csàky, 1948). MM9 overnight culture OD_{600} absorbance was measured and then cultures were centrifuged at 9,240 x *g* at RT for 5mins and supernatant transferred to bijoux. The supernatant was hydrolysed with 1ml 6N sulphuric acid in a boiling water bath filled with mineral oil. 1ml of both MM9 medium and water were also acid hydrolysed as controls. 3ml 35% sodium acetate solution was added to hydrolysed supernatant to buffer the solutions. 1ml sulfanilic acid solution (1g $C_6H_7NO_3S$ in 100ml 30% acetic acid) and 0.5ml iodine solution (1.3g iodine in 100ml acetic acid) were added and the mixture incubated at RT for 5mins. 1ml of sodium arsenite (2g NaAsO2 in 100ml water) was then added to neutralise excess iodine (turning the solution transparent) followed by 1ml α-naphthylamine solution (3g α-naphthylamine in 30% acetic acid) and 1.5ml distilled water to give a final volume of 10ml. The solution was incubated at RT for 30mins to allow the colour to develop and then absorbance was measured at 543nm. This absorbance was then standardised to overnight culture OD_{600} to give a relative measure of hydroxamate production.

2.4.3 Arnow assay

To measure the production of catecholate-type siderophores, enterobactin and salmochelin, by siderophore biosynthesis mutants, the Arnow assay was performed (Arnow, 1937). MM9 overnight culture OD_{600} absorbance was measured and then cultures were centrifuged at 3,226 x *g* at RT for 10mins and supernatant transferred to glass assay tubes. A blank of 1ml MM9 medium and a negative control of 1ml water were also transferred to glass assay tubes. A standard curve was prepared using a catechol solution (192mg catechol in 1 litre water) and making a 10x dilution series down to 10^{-5} dilution. 1ml of 0.5N HCl

and 1ml nitritre-molybdate reagent (10g sodium nitrite and 10g sodium molybdate in 100ml water) were added. Presence of catechols in the supernatant causes a colour change to yellow. 1ml 1M sodium hydroxide was then added, which in the presence of catechols would cause a colour change from yellow to red, and finally 1ml water added to give a final volume of 5ml. Solutions were incubated for 5mins at RT to allow the red colour to fully develop and absorbance was measured at 521nm. The amount of catechol present in the supernatants was calculated using the standard curve and standardised according to overnight culture OD_{600}.

2.5 *In vitro* gene expression measurement

2.5.1 Bacteria and growth conditions

The expression of a biosynthesis gene for each siderophore system was determined to investigate possible variations in siderophore regulation between strains. LB overnight cultures of eight GMB isolates (GMB23, GMB30, GMB40, GMB53, GMB88, GMB91, GMB100 and GMB104) were washed once with water and diluted to a final OD_{600} of 0.05 in 25ml of MM9 medium without addition of iron. Overnight LB cultures of strain GMB104, which possessed all four siderophore systems, was also used to inoculate 25ml of nutrient broth (NB) with or without the addition of 200µM or 500µM of iron chelator 2,2'-dipyridyl (DIP). The cultures were incubated at 37°C with aeration until reaching the exponential phase of growth (OD_{600} = 0.2).

2.5.2 RNA extraction, DNase treatment and cDNA synthesis

2.0 OD_{600} units of bacteria were harvested from cultures grown as described in section 2.5.1 and transferred to 50ml falcon tubes containing ice cold 1/5 volume phenol/ethanol stop solution (10% (v/v) phenol pH 4.3 (Sigma-Aldrich) 90% (v/v) ethanol (Sigma-Aldrich)). Cultures were left on ice for 30min and centrifuged at 3,226 x *g* for 10min at 4°C. The supernatant was then discarded and bacteria resuspended in residual liquid in the falcon tube. The bacteria were then transferred to Eppendorf tubes and centrifuged again at 1,700 x *g* and 4°C

for 10min. Any remaining supernatant was removed and pellets were stored at -80°C if necessary for RNA extraction at a later time.

For RNA extraction, pellets were resuspended in 100µl TE buffer containing 50mg/ml lysozyme and incubated at RT for 4min. From this point the Promega SV total RNA purification kit was used. 75µl of lysis buffer (provided with the Promega kit) were added to cells and mixed by inversion. 350µl RNA dilution buffer (Promega) was added and mixed by inversion and samples were heated at 70°C for 3min to lyse bacteria. Samples were then centrifuged for 10min at 15,616 x *g* at RT and supernatant transferred to RNase free 1.5ml tubes (Axygen). 200µl 95% ethanol was added to the lysate and then the Promega SV total RNA purification kit spin columns were used to extract RNA as per manufacturer's instructions. Extracted RNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific).

A DNase treatment is included in the Promega SV total RNA purification Kit protocol. However, significant amounts of DNA were still found and an additional DNase treatment was required: 1U of DNase I (Thermo Scientific) was added to each µg of RNA to be treated in the following reaction mixture: 1µl 10x buffer (100 mM Tris-HCl, 25 mM $MgCl_2$, 1 mM $CaCl_2$) (Thermo Scientific), 1µg of RNA and made up to a final volume of 10µl using RNase free water (Sigma-Aldrich). The amount of RNA could be scaled up as necessary, increasing the amount of DNase I used so that 1U was used for every 1µg of RNA in the reaction mixture. The DNase mixture was incubated at 37°C for 30min and then 0.1 volume of EDTA (Fermentas/Thermo Scientific) added. The mixture was then incubated for 10min at 65°C to stop the DNase reaction.

To remove the DNase from the RNA, the RNA was extracted using phenol-chloroform. QIAGEN water was added to DNase treated RNA to bring the volume up to 400µl and then 400µl of phenol-chloroform acid pH4.3 (Ambion) were added and the solution vortexed to form an emulsion. Samples were then centrifuged at 13,800 x *g* for 30min at 4°C and the upper phase transferred to an RNase free 1.5ml tube (Axygen). 2.5 volume of ice cold 95% ethanol (Sigma-

Aldrich) and 0.1 volume sodium acetate 3M pH5.2 (Ambion) were added and the solution mixed via inversion. The samples were then incubated at -80°C for 1h to allow the RNA to precipitate. RNA was pelleted by centrifugation at 17,982g for 30min at 4°C and the supernatant removed. RNA pellets were washed using 1ml ice cold 70% ethanol and then centrifuged again at 17,982 x *g* for 10min at 4°C. The supernatant was removed and the pellet dried at 95°C for 2min to remove all residual ethanol. RNA was resuspended in 100µl RNase free water (Sigma-Aldrich) and dissolved by heating for 10mins at 65°C. RNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific).

To determine whether DNase treatment was successful, a PCR was performed using untreated and treated RNA. Universal primers targeting the 16S gene were used (see table 2.8) as it is well conserved in *E. coli*. Successfully treated RNA should not yield a product in the PCR due to the absence of DNA. 100ng of treated or untreated RNA was mixed with 12.5µl GoTaq Green Master Mix, 1µl each primer (0.4µM final concentration) and water up to a final volume of 25µl. PCR reaction settings were as follows: 1 cycle of 95°C for 5min, 40 cycles of 94°C for 30s, 55°C for 1min, 72°C for 1 min and 1 cycle of 72°C for 5min. PCR products were run on a 1% agarose gel.

Table 2.8: Universal 16S primers used to detect DNA contamination of DNase treated RNA.

Region	Primer	Sequences
V3	341F	CCTACGGGAGGCAGCAG
	534R	ATTACCGCGGCTGCTGG
V1-V3	63F	CAGGCCTAACACATGCAAGTC
	338R	GCTGCCTCCCGTAGGAGT

cDNA synthesis was performed using the Fermentas RevertAid premium kit. 5µl RNA were mixed with 1µl Random Hexamer primer, 1µl 50x dNTP mix and made up to 14.5µl with RNase free water (Sigma-Aldrich). This mixture was incubated at 65°C for 5min and then chilled on ice briefly and centrifuged to remove any condensation from the Eppendorf lid. 4µl 5x RT buffer and 1µl (200U) RevertAid Reverse Transcriptase were added and the mixture incubated for 10min at RT

followed by 30min at 50°C. The reaction was terminated by heating at 85°C for 5min. cDNA could then be used directly for RT-PCR.

2.5.3 Real-time PCR

It was assumed that 1µg of RNA yielded 1µg cDNA (100% yield). The reaction mixture per well was 12.5µl SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich), 0.25µl Reference dye (in kit), 0.05 primer 1 (100µM), 0.05 primer 2 (100µM), 100ng cDNA, and water up to a final volume of 25µl. Reactions were carried out in 96 well plates which were centrifuged prior to the RT-PCR at 2,760 x *g* for 1min at RT to remove air bubbles. The RT-PCR reaction settings were 1 cycle of 94°C for 2mins, 40 cycles of 94°C for 30s, 60°C for 30s, 72°C for 45s, 1 cycle of 72°C for 5mins, and 1 cycle of dissociation curve (1 cycle of 95°C for 15s, 60°C for 1min, 95°C for 15s and 60°C for 15s). Fluorescence was measured during the 45s elongation step at 72°C. RT-PCRs were performed on an ABI TaqMan 7500 System. ROX indicator settings were used for detection of the reference dye used in this study, which was provided with the Green JumpStart Taq ReadyMix (Sigma-Aldrich).

Primers used in the RT-PCR were designed in silico using the Primer3 software (http://frodo.wi.mit.edu/) to target conserved regions of one biosynthesis gene for each siderophore system (*entC*, *iroB*, *irp2*, *iucA*) and the internal standard *rpoB* (table 2.9). *rpoB* was used to standardise gene expression as previously described (Lopez-Velasco et al., 2010). *rpoB* gene expression in strain GMB104 was analysed under conditions of different levels of iron limitation including MM9 medium and NB medium supplemented with 2,2'-dipyridyl at various concentrations (0µM, 100µM, 200µM and 500µM). Results showed limited variation in *rpoB* expression, with differences of less than 1 Ct between conditions (data not shown). Amplicons were designed to be 100-150bp in size and their efficiency was determined by carrying out standard curves using known amounts of DNA. The RT-PCR reaction mixture and settings for these standard curves were the same as described above. Assuming 100% efficiency of primers, for a 10x dilution series the Ct value is expected to increase by 3.3 for every 10-fold reduction of cDNA. Primers with standard curves that had

gradients between 3.0-3.6 (90%-110% efficiency) were suitable for use in the RT-PCR. Any primers that fell outside of this efficiency were redesigned to optimise their efficiency.

Table 2.9: RT-PCR primers for the detection of siderophore biosynthesis gene expression.

Siderophore	Gene	Primer Sequences		Product Size
Enterobactin	*entC*	F	CGAGCGTTTTAGCTCCATTC	143
		R	CCTCTTTCATCGCCTGAGTC	
Salmochelin	*iroB*	F	TATACCGGTCGTGATGCAAA	150
		R	ATACTCGGCGGTGTTACGTC	
Yersiniabactin	*irp2*	F	TAAAACTGAAGCCGGGTCAC	122
		R	CCGTTGTGTCACCAGAAATG	
Aerobactin	*iucA*	F	CTGCCGGTCGGATTTATTTA	138
		R	ATAAGGGAAATAGCGCAGCA	
RpoB	*rpoB*	F	GTGGTGAAACCGCATCTTTT	138
		R	CGATGTACTCAACCGGGACT	

The relative expression of each gene compared to the housekeeping gene *rpoB* was calculated. Comparisons between expression in MM9 and NB were calculated using the Livak (ΔΔCt) method (Livak and Schmittgen, 2001). This method works with the assumption that each primer pair has an equal efficiency and that amplification efficiency is 100%. To calculate gene expression ratios, the following equations were used:

$$\Delta Ct_{(\text{gene of interest})} = Ct_{(\text{gene of interest})} - Ct_{(rpoB)}$$

$$\Delta\Delta Ct = \Delta Ct_{(\text{treated})} - \Delta Ct_{(\text{control})}$$

$$\text{Ratio of gene expression} = 2^{-\Delta\Delta Ct}$$

Treated samples were those grown under iron limitation (MM9, NB with 200µM DIP and NB with 500µM DIP) and the control was expression measured in NB (iron replete conditions). For comparison with siderophore production on CAS plates, mRNA levels were normalised to the lowest corresponding value for each gene tested.

2.6 Semi-quantitative PCR to determine siderophore expression *in vivo*

2.6.1 Mouse faecal sampling

As enterobactin production and ferric iron transport have been shown to be important during GI-tract colonisation (Pi et al., 2012), and the possible role of yersiniabactin, aerobactin and salmochelin production in the gut by commensal *E. coli* is not known, we assessed whether siderophore genes are expressed within the mouse GI-tract. Wild-type female C57BL/6 mice were used for the *in vivo* siderophore expression experiment, and all were 12 weeks old. Mice, housing and food and water access were as described in section 2.2.4. Mice were given the antibiotics ampicillin and neomycin (100mg/ml of each in drinking water) for 1 week prior to gavage to enable bacteria to colonise the gut. During the antibiotic treatment mice remained on the standard chow diet. 24h before the bacterial challenge, mice were taken off the antibiotics and returned to standard water. Mice were then given 10^8 CFU of *E. coli* strain GMB104 by single oral gavage. The strain GMB104 was used as it has all four siderophore systems. Faecal samples were also collected when the gavage was given to spread on TBX plates to confirm the absence of *E. coli* within the GI-tract of the mice.

E. coli colonisation was monitored 1, 3, 8 and 14 days after bacterial challenge by collection of faecal samples for CFU counts. Faecal samples were homogenised in 1ml PBS and a 10x dilution series prepared. Dilutions were spread on TBX plates which were incubated overnight at 37°C. Colonies were counted the following day and faecal bacterial load (CFU/g) was calculated. At 14 days after the bacterial challenge, mice were euthanized using schedule 1 techniques (carbon dioxide and dislocation of the neck) and intestinal contents collected. Contents were taken from the colon, caecum and ileum and transferred to 15ml flacon tubes containing 1ml RNAlater solution (Ambion) and homogenised. Samples were snap frozen on dry ice to minimise degradation of RNA and stored at -80°C until RNA extraction was performed.

2.6.2 RNA extraction from faecal samples

To extract RNA, 2ml phenol-chloroform acid pH4.5 (Ambion) and 1ml RNase free water were added to frozen homogenised intestinal contents. Samples were then vortexed until they thawed completely and the solutions were fully mixed. Samples were then transferred to 2ml tubes containing a mixture of glass beads, sizes 710-1,180µm and ≤106µm (Sigma-Aldrich), macaloid clay and 70µl 10% sodium dodecyl sulphate (SDS). To lyse bacteria, a FastPrep FP120 (Thermo Savant) was used as follows: 45s at speed 6.5 5-6 runs with 1min incubation on ice between each run. Samples were centrifuged at 9,240 x *g* for 6min at RT and the upper phase was transferred to 1.5ml RNase free tubes (Axygen). 1.25 volume of ethanol was added and the solution mixed. This solution was then transferred to mirVana spin columns for RNA extraction. The MirVana miRNA Isolation Kit (Ambion) was used as per manufacturer's instructions. RNA was analysed using the Agilent RNA 6000 Nano Kit and the Agilent 2100 Bioanalyser. RNA was prepared for analysis on the Bioanalyser as per manufacturer's instructions. RNA was then DNase treated and cDNA synthesised as previously described (section 2.5.2).

2.6.3 Semi-quantitative PCR

100ng template cDNA was used to perform semi-quantitative PCR on *in vivo* intestinal contents. The reaction mixture also included 25µl GoTaq Green Master Mix (Promega), 0.1µl each primer (0.2µM final concentration) and water (Sigma-Aldrich) up to a final volume of 50µl. PCR amplification was as described for the *in vitro* siderophore expression RT-PCR (section 2.5.2), except that there was no dissociation curve. During the PCR, samples were taken at 20, 25, 30, 35 and 40 cycles and were visualised on a 2% agarose gel. Gel images were analysed using the ImageJ software (http://imagej.nih.gov.ij) (Abramoff et al., 2004) to determine band intensity, from which relative gene expression could be calculated.

2.7 Competitive index assays *in vivo*

Following on from siderophore expression experiments in the mouse model, competitive index assays were performed using siderophore biosynthesis and *ΔtonB* mutants to determine whether there was a fitness advantage for ferric iron uptake through siderophores. Overnight LB cultures, containing the required antibiotics, of wild-type GMB91 and GMB104 and mutants GMB91e, GMB91t and GMB104t were centrifuged at 3,226 x *g* at RT for 10min and washed twice with PBS. OD_{600} was measured following washing and cultures mixed to give 1:1 mixtures of the wild-type strain (GMB91 or GMB104) with one of its mutants.

10^8 CFU of a single mixture was given to groups of 5 mice (15 mice total) via oral gavage. C57BL/6 wild-type mice aged 4-6 weeks were used for the competitive index assays. Mice were housed and fed as described in section 2.2.4. Prior to gavage, mice were pre-treated for one week with ampicillin (1mg/ml) and neomycin (1mg/ml) in drinking water, which was removed 24h before gavage. Faecal samples were taken during gavaging to confirm successful antibiotic treatment and no detectable *E. coli* when samples were grown on TBX plates. Faecal samples were collected on days 1, 4, 9, 14 and 28. Faecal samples were homogenised in PBS, diluted and plated out onto both MacConkey and LB with either kanamycin (100μg/ml) or chloramphenicol (25μg/ml) agar plates. The MacConkey plates were used to determine total *E. coli* levels, whereas the LB with antibiotics plates would only grow mutants possessing resistance cassettes. All plates were incubated at 37°C overnight before counting. A minimum of one log difference in CFUs between the wild-type and a mutant was required to be considered as a significant difference in colonisation ability.

2.8 Statistical analyses

Statistical analysis was performed using GraphPad Prism (version 5) or Microsoft Excel (2010) software. A Student's *t*-test was used for comparisons between two groups, including analysis of differences between the ECOR-F and GMB collections (liquid CAS assays, number of siderophore genes), as well as changes

in siderophore production and gene expression between wild-type strains and/or mutants (Csàky and Arnow assays, liquid CAS assays, siderophore gene expression, competitive index assays). For comparisons of multiple groups, the Benjamini and Hochberg false discovery rate method was performed following a Student's *t*-test for the correction of possible type I errors (Benjamini and Hochberg, 1995). The Benjamini-Hochberg critical value, (i/m)Q, was calculated, where i = *p*-value rank, m = total number of statistical tests and Q = false discovery rate. A false discovery rate of 5% was used for our analysis. *p*-values of less than 0.05 and lower than their corresponding (i/m)Q value were considered to be significant.

Contingency tables produced from calculating the prevalence of siderophore systems in ECOR-F and GMB strains were analysed using Fisher's exact test. For multiple comparisons, the Benjamini and Hochberg false discovery rate method was performed as described above (Benjamini and Hochberg, 1995).

For the analysis of Biolog plates, OD_{600} readings were used to calculate Spearman correlation coefficients (r_s) between carbon source utilisation profiles. These were analysed for both barcoded and siderophore mutants to establish possible differences to wild-type strains. For siderophore mutants, principal component analysis (PCA) was also performed on OD_{600} readings, both for comparisons between the wild-type and mutants, and between individual carbon sources. PCA enables possibly correlated variables to be converted into a set of variables that are linearly uncorrelated variables called principal components. These principal components are ranked in terms of how much variation in the data they account for. The two principal components that were responsible for the highest amount of variation were used to construct plots to visualise the variation in the data.

3. Dynamics of *E. coli* populations in mixed cultures and in primary and secondary environments

3.1 Introduction

The factors involved in the diversity, transition and persistence between and within hosts of commensal *E. coli* and other enteric bacteria are not fully understood. To address this gap in knowledge, epidemiological approaches can be used to assess and compare the diversity of bacterial communities associated with various hosts (Hayashi et al., 2002; Zhu et al., 2002) and/or in the external environment (Lim et al., 1993; Gray and Herwig, 1996; Gilbert et al., 2012). 16S sequencing is frequently used in bacterial identification as it is present in most strains of bacteria. As the 16S rRNA gene is not changing in function, it can be used to give an indication of the evolution and relatedness of different species (Janda and Abbott, 2007). In terms of distinguishing different strains within single bacterial species, however, 16S rRNA sequencing gives mixed results. In some bacteria it has been shown to differentiate between strains (Moghadam et al., 2010), but in others accurate identification cannot be achieved at the species or even genera levels (Bosshard et al., 2006; Mignard and Flandrois, 2006). This is due to variations in heterogeneity that exists in the 16S rRNA gene between strains of the same species (Janda and Abbott, 2007). This can result both in closely related strains being classified as different species, as well as some distantly related species being classed as the same species (Coenye and Vandamme, 2003). 16S amplicon sequencing of complex environments, including the gut, has yielded important information about which bacteria are present and in what amount, but it often cannot discriminate down to the strain level. This, however, is important in understanding *E. coli* populations, as several studies have indicated that most humans possess more than one strain of *E. coli* in their gut microbiota (Smati et al., 2013; Gordon et al., 2015), as well as multiple strains of *E. coli* being found together in external environments (Méric et al., 2013).

E. coli are typically identified as lactose-positive and citrate-negative colonies on selective agars such as Simmon's citrate agar. Studies that have investigated the diversity of *E. coli* in the GI-tract have mainly focussed on the prevalence and distribution of the major phylogenetic groups as they can be determined fairly quickly and easily using a quadruplex PCR targeting the *chuA*, *yjaA* and *aprA* genes and the TspE4.C2 gene fragment (Clermont et al., 2013). Studies have shown that the phylogenetic groups A and B2 are prevalent in humans (Duriez et al., 2001; Zhang et al., 2002; Escobar-Páramo, Grenet, et al., 2004; Nowrouzian et al., 2005; Pallecchi et al., 2007; Gordon et al., 2015), with the distribution of the phylogenetic groups being influenced by geographic and socioeconomic factors (Tenaillon et al., 2010). Although *E. coli* is widely distributed among humans, being detected in over 90% of individuals sampled to date, its occurrence in other animals varies significantly, from 56% in wild mammals to 23% in birds and 10% in reptiles (Gordon and Cowling, 2003; Tenaillon et al., 2010). This can be linked to host characteristics, such as host gut morphology, body mass, diet and sex (Gordon and Cowling, 2003). From 1,154 animal samples, the distribution of the main phylogenetic groups of *E. coli* was 41% B1, 22% A, 21% B2 and 16% D (Tenaillon et al., 2010). Some strains have only been isolated from specific hosts, suggesting some host specialisation, but these only make up a very small proportion of all *E. coli* (Escobar-Paramo et al., 2006; Clermont et al., 2008). The B1 phylogenetic group has been shown to be prevalent not only in animals, but also in the external environment (Gordon et al., 2002; Walk et al., 2007; Méric et al., 2013), possibly reflecting the movement of *E. coli* between the host and environment.

Few studies have, however, looked at what happens during colonisation of the gut by *E. coli*, and how competing strains interact with one another. This assessment of the fitness of large numbers of *E. coli* strains in complex environments, such as the gut or external environment, is extremely time consuming and expensive. Studies that do identify individual strains from faecal samples use several different techniques, but all involve randomly selecting *E. coli* colonies grown from faecal samples (Nowrouzian et al., 2005; Lautenbach et

al., 2008; Gordon et al., 2015; Wassenaar and Gunzer, 2015). This reduces the ability to fully analyse the entire population and detect all strains present. The quadruplex PCRs used to identify phylogenetic groups described above also involve characterisation of a subset of colonies. It has been estimated that a phylogroup must occupy at least 10% of the whole population to be accurately detected in sampled colonies (Lautenbach et al., 2008; Smati et al., 2013), leading to possible underestimations of the number of phylogroups or strains. Recently, the phylogroup PCR has been adapted for use in qPCR, reducing the threshold for detection from 10% down to 0.1% of the population (Smati et al., 2013). Methods for identification include multilocus enzyme electrophoresis (MLEE), multilocus sequence typing (MLST), random amplified polymorphic DNA (RAPD) PCR, Repetitive extrageneic palindromic elements (REP) PCR and enterobacterial repetitive intergenic consensus sequences (ERIC) PCR.

Studies on *E. coli* colonisation of the GI-tract of babies have shown that those strains that colonised shortly after birth are more likely to persist for longer time periods in the gut than those that colonise later (Kühn et al., 1986). This could highlight that the first strains introduced into the sterile infant gut can establish themselves more readily due to absence of the microbiota and the resulting reduced competition. Exposure of adults with a healthy microbiota to a single high dose of probiotic *E. coli* (Symbioflor2), however, can result in prolonged persistence and detection in faeces of *E. coli* (Wassenaar et al., 2014; Wassenaar and Gunzer, 2015). Indeed, the stability and persistence of early colonisers is not observed in all children, with infants from Pakistan possessing more resident strains of *E. coli* in their first 6 months compared to children from Sweden (Kühn et al., 1986; Adlerberth et al., 1998; Nowrouzian et al., 2003). This difference may reflect increased exposure of Pakistani children to *E. coli* as a result of poor hygienic conditions (Adlerberth et al., 1998).

Some strains have been shown to be better at persisting in the GI-tract, with *E. coli* isolates acquired from the mother more likely to become residents compared to strains from other sources (Adlerberth et al., 1998). Several ExPEC virulence factors have been found to be increased in strains that are able to

persist for long periods, suggesting they are acting as colonisation factors. These include aerobactin, hemolysin, K5 capsule, the *iha* siderophore receptor and *eib* immunoglobulin binding proteins (Nowrouzian et al., 2001a; Nowrouzian et al., 2006; Wassenaar and Gunzer, 2015). These virulence factors are found more frequently in the B2 phylogenetic group (Lee et al., 2010), which is often observed as the most predominant group isolated from the GI-tract (Zhang et al., 2002; Escobar-Páramo, Grenet, et al., 2004; Gordon et al., 2015), as well as frequently being found as a resident strain (Nowrouzian et al., 2005; Nowrouzian et al., 2006; Gordon et al., 2015).

To identify genes that may influence colonisation ability or pathogenicity in the host by several strains of bacteria, many studies have utilised transposon insertion sequencing (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009; van Opijnen and Camilli, 2010). Transposons are used to introduce knockouts throughout the genome via insertional mutagenesis. Mutants are pooled and screened together alongside the parental strain in a host model and mutations that are decreased, or increased, in the population indicate which genes affect colonisation or pathogenicity. These studies are limited by the fact that mutations can only be studied in one specific strain. This is important, as the genetic background of a strain influences the phenotype produced when a gene is knocked out due to epistatic effects. Also, individual gene mutations can only be studied in isolation, which again does not allow epistatic effects to be taken into account.

To address the current limitations in monitoring individual strains in mixed populations, we developed a new method for uniquely tagging and identifying *E. coli* isolates that enables parallel monitoring of several isolates in mixed *E. coli* populations to determine population dynamics. To do this unique 20nt barcodes (Xu et al., 2009) were inserted into several isolates using the Tn*7* transposon (McKenzie and Craig, 2006). Insertion by Tn*7* of the barcode was at a specific site, the end of the *glmS* gene, to prevent any possible changes in phenotype of mutants carrying the barcode. Detection of the barcode was carried out either by RT-PCR or using high-throughput sequencing. Competition studies were

carried out in the mouse GI-tract to determine whether persistence of particular strains is determined by stochastic factors or specific genetic elements. If individual strains show increased fitness or dominance, this should enable traits and phenotypes associated with increased colonisation to be identified. The barcoding method was also assessed in soil columns, to determine the application of this technique to other environments.

3.2 Barcode mutant characterisation

To minimise known problems associated with strain domestication in the laboratory environment, a series of phenotypic tests was used to exclude the most likely changes. This is particularly important as domestication is likely to result in changes altering *E. coli* survival both in the host and external environment. It is also possible for transposon insertion to occur outside of the site of interest (the *attTn7* site located at the end on the *glmS* gene). However, the Tn*7* transposon used in this study has been shown to have a non-specific insertion frequency of 6.8×10^{-5} (McKenzie and Craig, 2006), so this type of mutation should be rare. The barcoded strains were characterised alongside the parental strains to confirm that barcoding had no effect on phenotype. Four traits were measured, growth rate in LB medium, carbon source utilisation, catalase activity and colony morphology, all of which can be altered or lost in laboratory-adapted strains of bacteria.

3.2.1 Growth curves

Barcoded strains were grown alongside the parental strains in LB medium to confirm that insertion of the barcode did not affect growth. The majority of barcoded strains showed comparable growth to the parental, however, three barcoded strains (ECOR49-B9, ECOR55-B11 and ECOR18-B23) displayed a slight reduction in growth (figure 3.1).

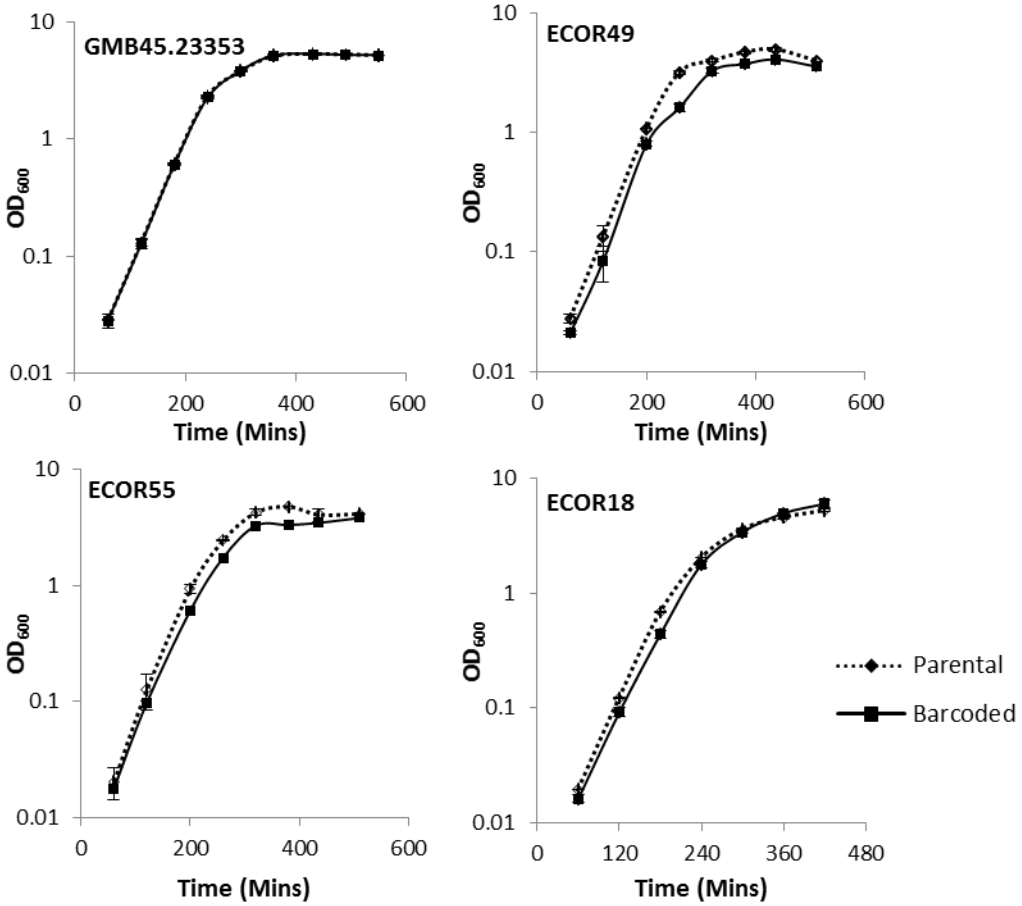


Figure 3.1: Four example growth curves of barcoded and parental strains grown in LB medium at 37°C. A) GMB45 and GMB45-23353. B) ECOR49 and ECOR49-B9. C) ECOR55 and ECOR55-B11. D) ECOR18 and ECOR18-B23. ECOR49-B9, ECOR55-B11 and ECOR18-B23 showed a reduction in growth compared to their parental strains. All other barcoded strains showed equivalent growth to their parental strains, represented here by GMB45-23353. Results show the mean ± standard error (*n* = 3). Growth curve graphs for other barcoded strains are shown in appendix B.1.

3.2.2 Catalase assays

A decrease in RpoS activity is often observed when *E. coli* strains experience protracted nutrient starvation, for example, when cultures reach stationary phase. One simple method to assess RpoS activity is to perform catalase assays which measure the activity of KatE, which is strongly regulated by RpoS (Mulvey et al., 1990). Catalase assays were performed on single colonies of both parental and barcoded versions of each strain. Results showed that catalase activity, and thus RpoS activity, was not altered for any of the mutants (table 3.1). ECOR49-B9 was not included in the assay due to significant changes in phenotype compared to the wild-type (see section 3.2.3).

Table 3.1: Catalase activity of parental and barcoded strains. KatE activity was determined based on visible oxygen production on exposure of colonies to H_2O_2. Activity levels were determined based on whether oxygen production occurred immediately or was delayed after H_2O_2 exposure. Strains where a large volume of O_2 was produced (larger in size than the colony) were also considered to have strong activity. ++ = strong activity, + = weak activity and - = no visible activity.

Strain	Catalase Activity	Barcode strain	Catalase Activity
ECOR16	+	ECOR16-B1	+
ECOR18	++	ECOR18-B23	++
ECOR32	-	ECOR32-23457	-
ECOR55	+	ECOR55-B11	+
GMB02	+	GMB02-23796	+
GMB07	+	GMB07-B19	+
GMB104	++	GMB104-29884	++
GMB16	+	GMB16-19427	+
GMB18	++	GMB18-B5	++
GMB23	+	GMB23-23277	+
GMB32	++	GMB32-20239	++
GMB34	+	GMB34-33233	+
GMB40	-	GMB40-B6	-
GMB45	++	GMB45-19439	++
		GMB45-23353	++
GMB48	++	GMB48-21118	++
GMB54	+	GMB54-33301	+
GMB71	+	GMB71-B10	+
GMB72	+	GMB72-B8	+
GMB98	+	GMB98-30438	+

3.2.3 BIOLOG plates

Domestication has also been shown to affect the utilisation of nutrients (Eydallin et al., 2013) as assessed by measuring growth using the BIOLOG system. GN2 BIOLOG plates were therefore used to assess carbon source utilisation of barcoded strains (list of carbon sources on BIOLOG plates in appendix B.2). The results indicated that the majority of strains had unaltered carbon source utilisation (figure 3.2). ECOR55-B11 and ECOR18-B23, which both showed reduced growth in LB, did not show a significantly altered carbon utilisation profile compared to their parental strains. GMB16-19427, however, although displaying the same growth curve as the wild-type, had a different

profile to the parental strain, caused by being unable to metabolise D- and L-serine.

ECOR49-B9, which as well as displaying reduced growth, showed a significantly different carbon utilisation profile compared to both ECOR49 and all other *E. coli* strains tested. This difference was caused by changes in metabolism of 10 different carbon sources, all of which could be used by the parental, but not the barcoded strain. ECOR49-B9, however, displayed reduced growth on almost all carbon sources on the BIOLOG plates compared to other strains, so these differences in metabolism may be as a result of ECOR49-B9 being unable to grow sufficiently.

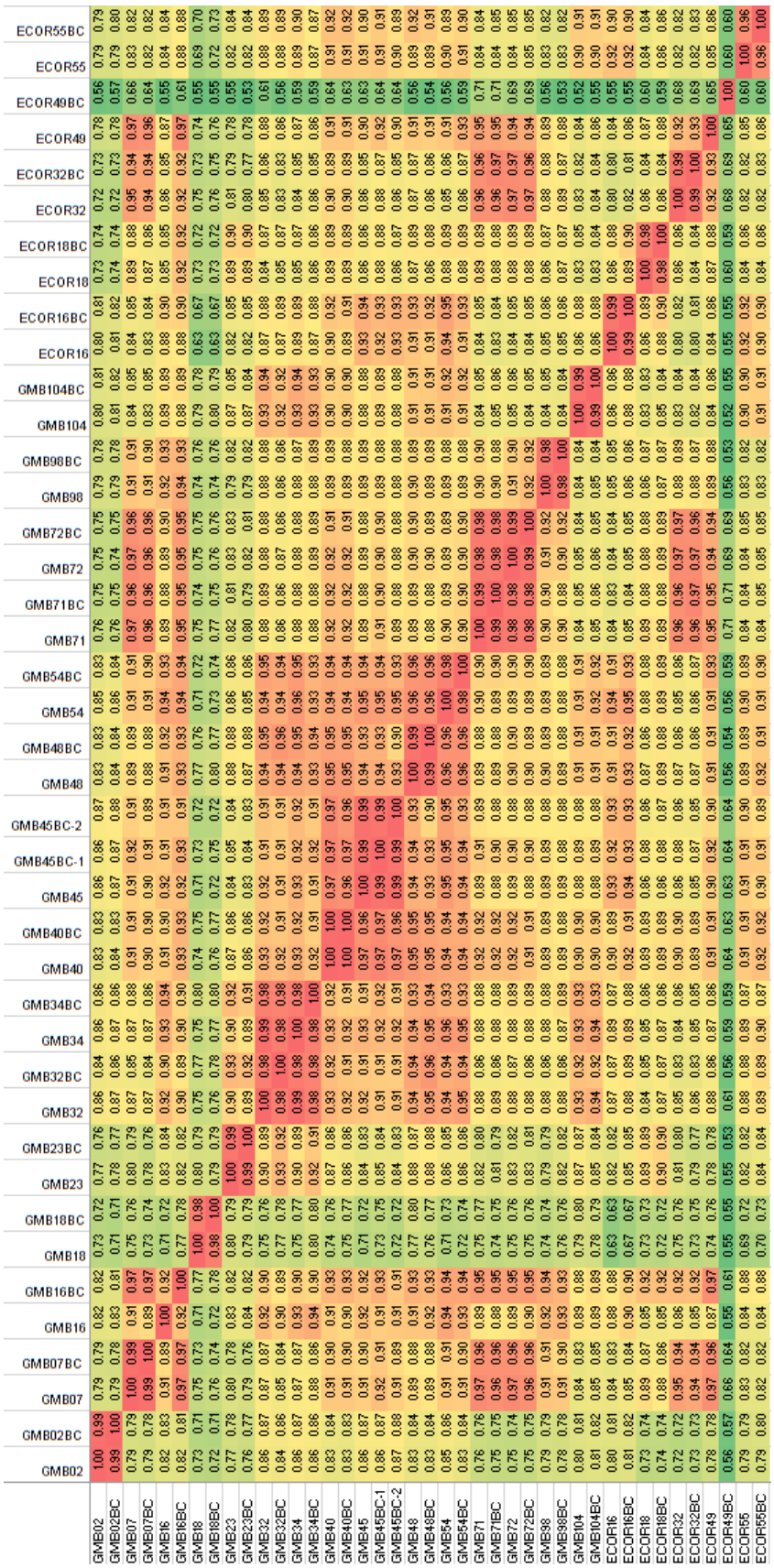


Figure 3.2: Correlation matrix between BIOLOG GN2 profiles of parental and barcoded strains after incubation at 37°C for 24h. Values in the table are Spearman correlation coefficients (r_s) between whole carbon source utilisation profiles. High correlation values are indicated in red and low values in green. The barcoded strains are indicated with the 'BC' suffix.

3.2.4 Colony morphology on YESCA plates

To determine whether transforming the strains to contain the barcode tag influenced colony morphology, bacteria were grown on YESCA plates, which can be used to visualise curli formation. Barcoded strains ECOR18-B23, ECOR55-11, ECOR49-B9 and GMB16-19429 were excluded from the colony morphology assay due to their altered phenotypes in earlier assays. Colony morphology was not affected in all the other barcoded strains compared to the parental strains (figure 3.3), resulting in 16 out of 20 (80%) of the barcoded strains tested being identical for all four phenotypic assays used.

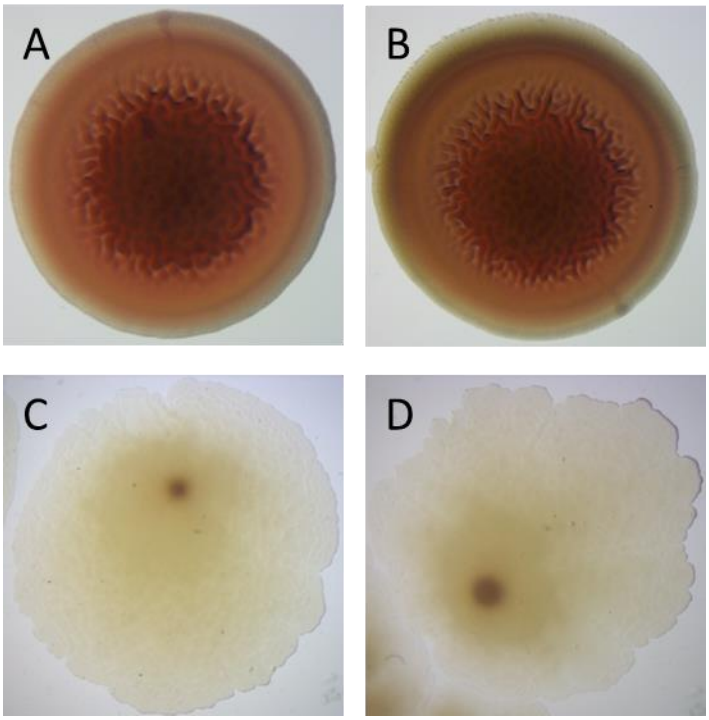


Figure 3.3: Colony morphology on YESCA plates of two strains of *E. coli* and their barcoded versions. (A) GMB40, (B) GMB40-B6, (C) GMB07 and (D) GMB07-B19 were imaged using a Leica M165C Stereo microscope after YESCA plates were incubated at 37°C for 48h followed by 7 days at RT. All other images are shown in appendix B.3.

3.3 Characterisation of barcode detection performance

3.3.1 RT-PCR standard curves

To be able to accurately determine the relative proportions of each strain, it was essential to perform standard curves on all barcoded strains to ensure that the efficiency of the RT-PCR reaction was comparable for each primer. Primer efficiency was determined by calculating the gradient of the graph, with Ct

values expected to reduce by 3.3 for every 10-fold reduction in gDNA concentration at 100% efficiency. Equations generated from these standard curves were used to calculate the amount of each barcode strain in the competition studies below (figure 3.4A). To assess the limit of quantification (LOQ) of the RT-PCR, standard curves using lower dilutions of gDNA were performed. The efficiency of several primers decreased when the amount of gDNA was under 0.01pg, indicating this quantity of gDNA represents the lower limit for accurate quantification of barcode strains by RT-PCR.

Two barcoded strains were used to determine whether primer specificity or the sensitivity of the RT-PCR was reduced in the presence of exogenous gDNA. gDNA extracted from healthy mouse faecal samples was used to contaminate gDNA extracted from pure barcode strains and standard curves performed. Faecal gDNA was kept at a constant concentration (10ng/µl), as used for competition study RT-PCRs, while the barcode gDNA was reduced, reflecting that faecal samples may contain low levels of barcode gDNA compared to gDNA derived from the rest of the microbiota. The results suggest that the primers were not affected by the presence of exogenous gDNA and that we can accurately detect each barcode with a very large dynamic range of 6-7 orders of magnitude (figure 3.4).

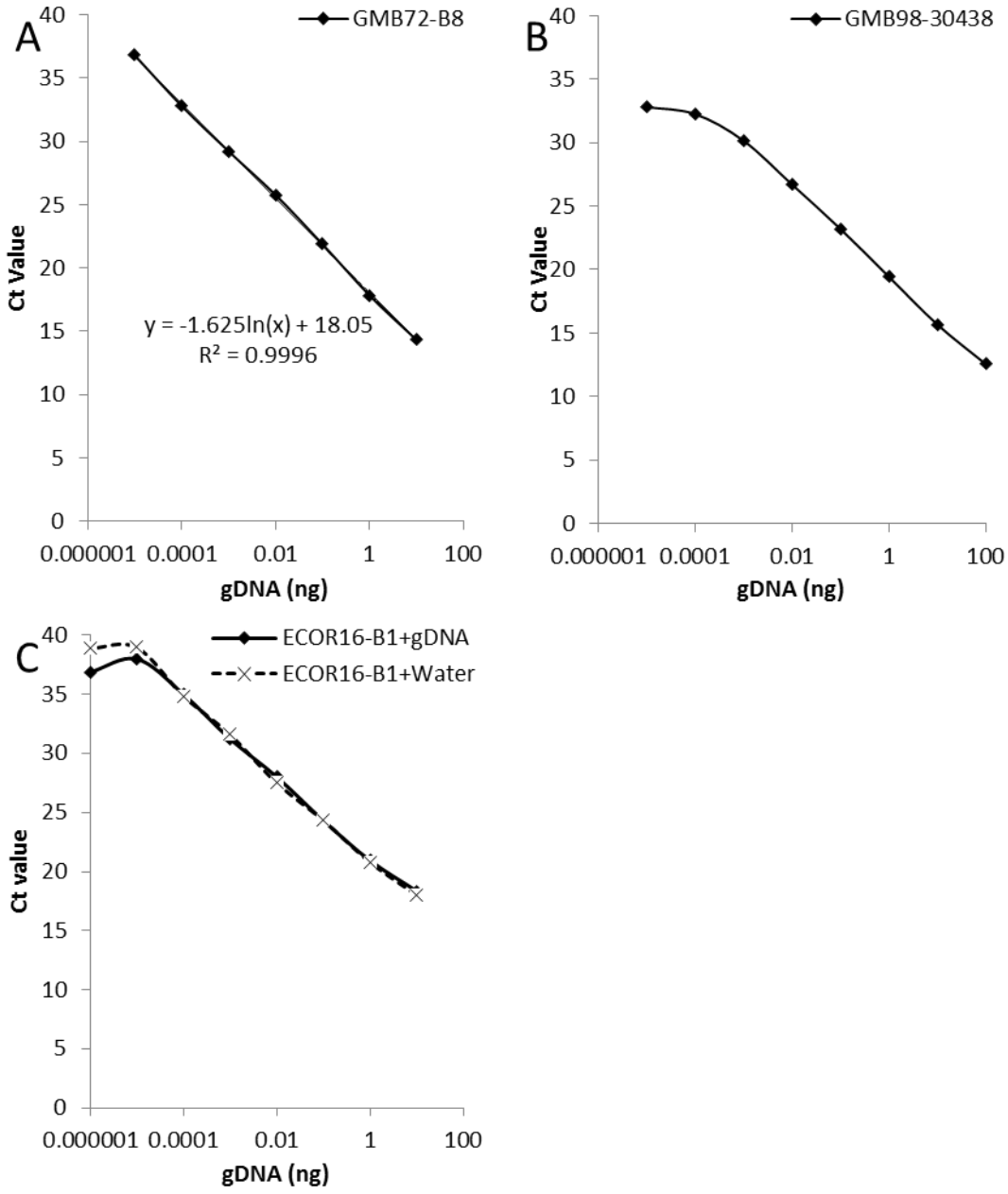


Figure 3.4: RT-PCR example standard curves for barcode detection primers. (A) Equation for determining gDNA concentration from RT-PCR Ct values (B) Loss of primer efficiency at low gDNA concentrations (C) Different concentrations of barcode gDNA mixed with a constant concentration of contaminating mouse faecal gDNA (10ng/µl) showed no loss of primer sensitivity.

3.4 Assessing the barcoding method *in vitro* using a culture competition study

To assess the feasibility of the barcoding approach, two multi-strain competition experiments in independent LB cultures were performed using eight strains (see table A.1 for strain information). Two different barcoded versions of GMB45,

GMB45-19439 and GMB45-23353, were included in this experiment to determine whether the barcode sequence influenced either the fitness or detection of a strain. The ECOR49-B9 and ECOR55-B11 barcoded strains were also used. These two strains, although displaying a different phenotype to the wild-type during barcode mutant characterisation (section 3.2), are still of value for testing the reproducibility of the method, which was the aim of the *in vitro* competition study.

The population dynamics were similar in both independent experiments, with strains ECOR16 and GMB45 becoming dominant during the early exponential growth phase, followed by GMB45 being dominant in the late exponential phase, and finally GMB71 being the most abundant during stationary phase. Strain ECOR16 grew well initially, but quickly declined after three hours of incubation to a level barely above the detection limit of the RT-PCR. The remaining four strains in the mixed cultures remained at fairly stable proportions for the duration of the experiment (figure 3.5). These results were reproducible between cultures, suggesting that competitive fitness changes depend on growth phase, and that ECOR16 was competitively excluded during stationary phase growth. There was also evidence of fluctuations within a strain which were probably stochastic (displaying an unpredictable pattern as a result of a random variable). This was observed in the two barcode versions of GMB45 which were present at different proportions in each mixed culture. The overall population of GMB45, however, is consistent between the two cultures.

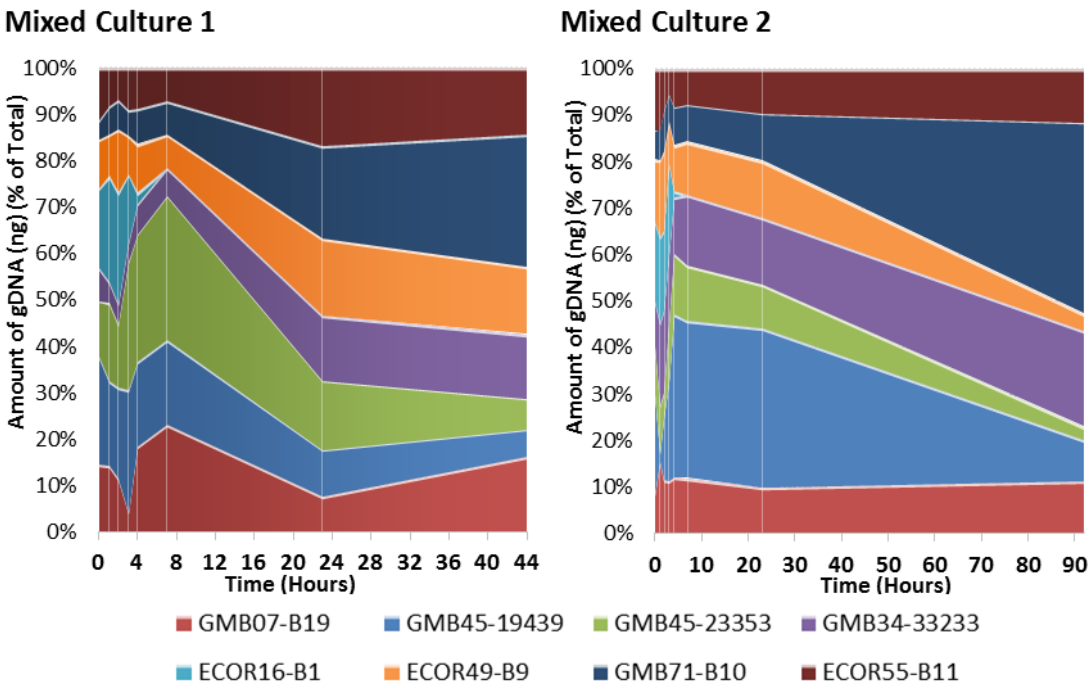


Figure 3.5: Stacked area graphs of the proportions of each barcode strain as a percentage of the whole *E. coli* population in mixed-strain LB cultures grown at 37°C. Graphs show proportions for both independent mixed cultures, which displayed similar population dynamics. ECOR16-B1 declined rapidly after 3h to a level barely above the detection limit. The cultures displayed a shift in dominant strains dependant on growth phase. Samples were taken at 1, 2, 3, 4, 7 and 23h after inoculation for both cultures. The final samples were taken at 44h for culture 1 and 92h for culture 2.

3.5 Determining the competitiveness of *E. coli* isolates in the gut using competition studies in the mouse model

The GI-tract is a highly complex and dynamic environment. To determine whether the barcoding method could be applied *in vivo*, two independent experiments were performed where barcoded strains were given to groups of four to five cohoused C57BL/6 mice. These mice were treated with antibiotics ampicillin and neomycin one week prior to gavage to facilitate colonisation by *E. coli*. This combination of antibiotics has been shown to disrupt approximately 90% of the microbiota (Vijay-Kumar et al., 2010). Loss of *E. coli* and other Gram-negative bacteria after antibiotic treatment was confirmed by spreading faecal samples on TBX and MacConkey plates. CFU counts showed that following a single oral gavage of 10^8 CFU, *E. coli* were able to persist within the GI-tract for

several weeks (figure 3.6). Samples were taken at regular intervals and the gDNA was extracted from faecal samples taken alongside CFUs.

3.5.1 Mouse competition experiment 1

The GI-tract is a highly complex and dynamic environment. To determine whether the barcoding method could be applied *in vivo*, seven barcoded strains were given to four cohoused C57BL/6 mice (see table A.1 for strain information). The number of *E. coli* isolated from faecal samples peaked at an average of 7.65×10^9 CFU/g on day two, which then decreased to approximately 10^4-10^5 CFU/g and stabilised at about that level for the remainder of the experiment (figure 3.6). gDNA extracted from faecal samples taken alongside CFUs was used in RT-PCRs to quantify the amount of each strain present.

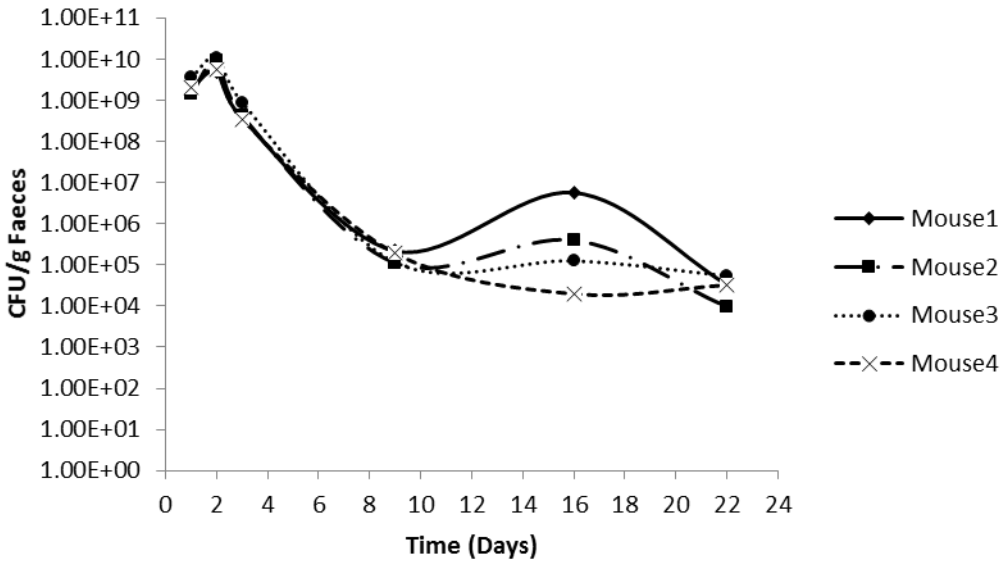


Figure 3.6: CFUs for total *E. coli* from mouse faecal samples. CFU counts peaked at day 2, and then declined to stabilise at approximately 10^5 CFU/g at the end of the experiment. Mice were monitored for 22 days. Results show faecal CFU/g for four mice ± standard error of the mean.

The results show there was reproducibility between mice, with the *E. coli* strains displaying similar trends in population dynamics in all four mice (figure 3.7). GMB34-33233 and GMB71-B10 initially became the dominant strains, with over 90% of *E. coli* belonging to these two isolates after three days in every mouse. After the first three days, a shift in the dominant *E. coli* was observed, to GMB45, both GMB45-19439 and GMB45-23343. During this shift in dominance

the total *E. coli* CFUs measured in faecal samples decreased from 10^9 CFU/g on day three to 10^5 CFU/g on day nine (figure 3.6).

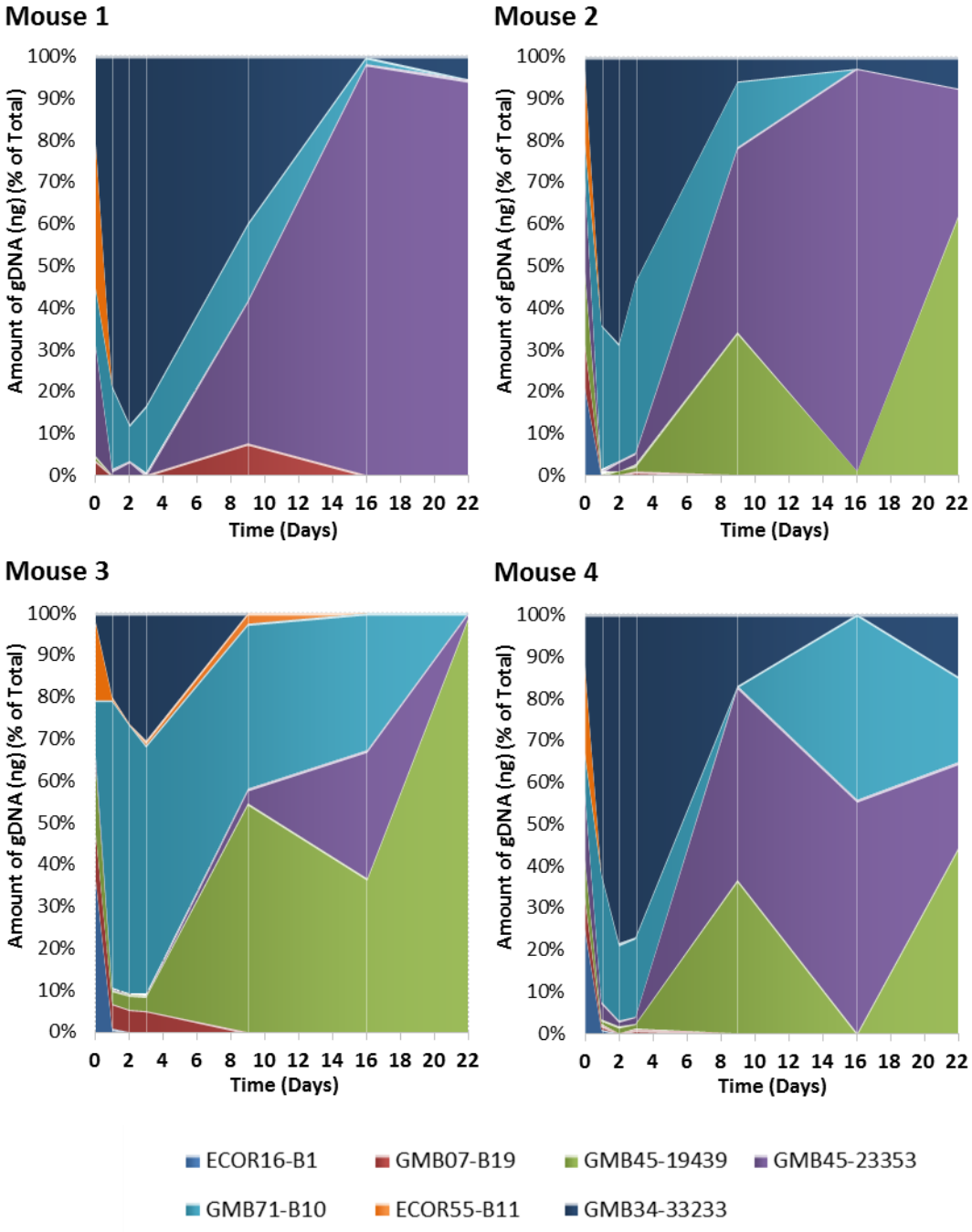


Figure 3.7: Stacked area graphs of the proportions of each barcode strain in mouse faecal samples as a percentage of the whole *E. coli* population for the first *in vivo* experiment. Graphs show proportions for four individual mice calculated from RT-PCR quantification of barcodes. Both ECOR strains, ECOR16-B1 and ECOR55-B11 declined rapidly, becoming undetectable by RT-PCR after three days. There was a visible shift in strain dominance between days three and nine. Samples were taken on days 1, 2, 3, 9, 16 and 22 after oral gavage. Day 0 strain proportions were calculated from the bacterial mixture used to gavage mice.

We also observed variations in the proportion of both GMB45 barcoded versions in each mouse. It is possible that this is as a result of clonal interference, where beneficial mutations arising in GMB45 are competing with one another, leading to fluctuations in which barcode version is dominant as new beneficial mutations arise. Dynamic changes would occur in the distribution of the two barcode versions of GMB45 as mutations become spread through the population and are then replaced by new ones in another clonal line. It was quite unexpected that the levels of one barcoded version of GMB45 fell below the detection limit in a single mouse (mouse one lost GMB45-19439) as the mice were cohoused and exhibit coprophagia, so sharing of their microbiota is expected (Barnes et al., 1963). Both ECOR isolates, ECOR55-B11 and ECOR16-B1, were quickly lost from the host, becoming undetectable after day three.

For this study, faecal rather than intestinal content samples were used to determine gut fitness of *E. coli* strains. We used the assumption that an increased prevalence in faeces correlates with an increased abundance, and thus increased fitness, in the GI-tract. However, differences in strain presence and diversity have been observed between faecal and intestinal samples (Gordon et al., 2015). This limits the ability to interpret dominance in faecal samples, like that displayed by GMB45, as increased fitness in the GI-tract. It is therefore important that the distribution of barcoded strains within the gut is also examined.

3.5.1.1 Barcode strain distribution in gut contents

Intestinal contents from different areas of the gut were collected from the mice on day 27, 5 days following the collection of the final faecal samples. Unfortunately, there was an insufficient amount of faecal material on day 27 to collect for CFUs and RT-PCR, so only bacterial presence in the GI-tract was determined. Only two of the barcoded strains were detectable by RT-PCR, GMB71-B10 and GMB45-23353, from all of the colonic, caecal and ileal samples. These two strains were both detected in all four mice at day 22 at relatively high densities (figure 3.8). The distribution of these two strains was not equal in these three sections of the GI-tract, with one strain typically being detected in

isolation for each sample, except for two caecal samples where the two strains were detected at almost equal levels (figure 3.8). Two ileal samples contained no detectable barcoded *E. coli* strains, which may be a result of lower levels of *E. coli* typically being present in the ileum compared to the colon (Savageau, 1974).

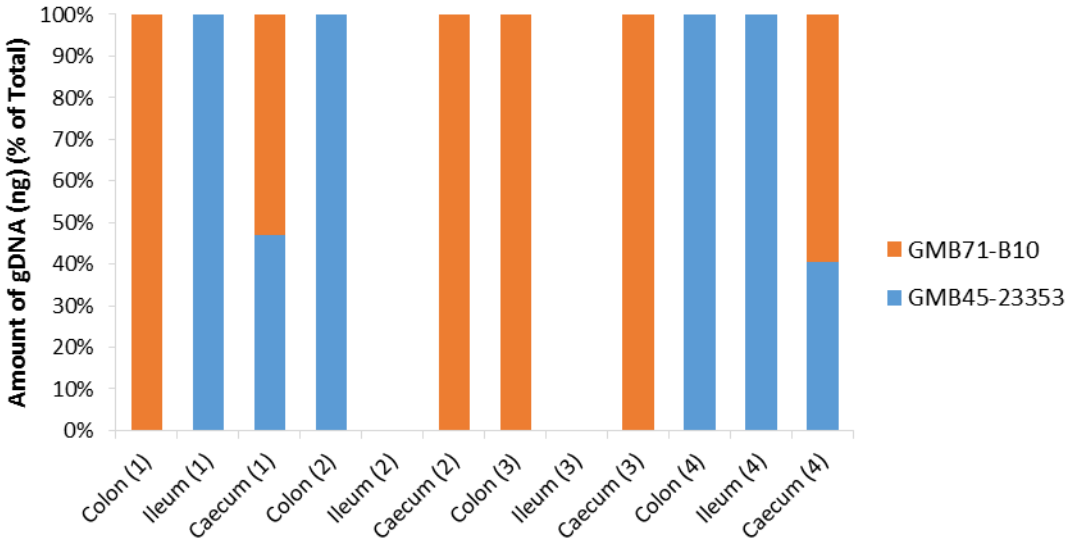


Figure 3.8: Proportion of strains GMB71-B10 and GMB45-23353 in colonic, ileal and caecal samples as a percentage of total *E. coli* detected. Only GMB71-B10 and GMB45-23353 were detectable in intestinal contents. Mouse three had only one strain detected in intestinal contents, whereas in all other mice both GMB71-B10 and GMB45-23353 were detectable. All other mice displayed compartmentalisation, where strains were detected only in certain regions of the GI-tract, rather than being evenly spread throughout.

The detection of these two strains was, however, near the detection limit of the RT-PCR (0.01pg gDNA). As a result, the observed compartmentalisation of GMB71-B10 and GMB45-23353 in the GI-tract may be caused by inaccuracies in detection rather than differences in distribution. This also has implications for whether other strains could still be present in the GI-tract at levels below the RT-PCR detection threshold. It was unexpected that only GMB45-23353 was detected, and not GMB45-19439, as both were detectable at high levels on day 22 in three of the mice. The levels of the strains carrying these two barcodes do, however, vary considerably throughout the experiment, with GMB45-19439 disappearing in mouse 4 at day 16 and then reappearing at day 22 (figure 3.7). This difference between strains in detection may, however, be caused by

variations in primer efficiency at low gDNA concentrations, which can be seen in some of the RT-PCR standard curves (figure 3.4).

3.5.2 Mouse competition experiment 2

For the second *in vivo* competition study, detection and quantification of the barcodes was performed by using high-throughput sequencing, rather than RT-PCR. This enabled us to increase the number of strains to put under direct competition to 12 (see table A.1 for strain information). Indeed, the number of RT-PCR reactions increases exponentially in relation to the number of strains, making the use of this quantification method less practical for large numbers of strains.

Informed by the results of the first experiment where the two ECOR strains used were undetectable after three days, possibly as a result of mutations accumulated during long-term storage, only strains from the GMB collection were used in this second experiment. The strain mixture was given to five cohoused, antibiotic treated, C57BL/6 mice by single oral gavage. As with the first mouse competition study, the barcoded strains were able to colonise and persist within the GI-tract for the duration of the experiment (30 days). CFU counts showed that *E. coli* levels peaked at day two, at 7.18×10^9 CFU/g, and then declined to approximately 10^6 CFU/g after a week, and persisted at those levels for the remainder of the experiment (figure 3.9). Although the levels were about the same at day two between the two mouse experiments, the levels of *E. coli* were about 1 log higher in the second experiment when the *E. coli* population had stabilised. To determine whether high-throughput sequencing could be used as an alternative to RT-PCR for barcode detection, faecal gDNA samples were used to prepare amplicons containing the barcode sequence. These were then run on the Illumina HiSeq sequencing platform to determine the levels of each barcode strain in the faecal samples.

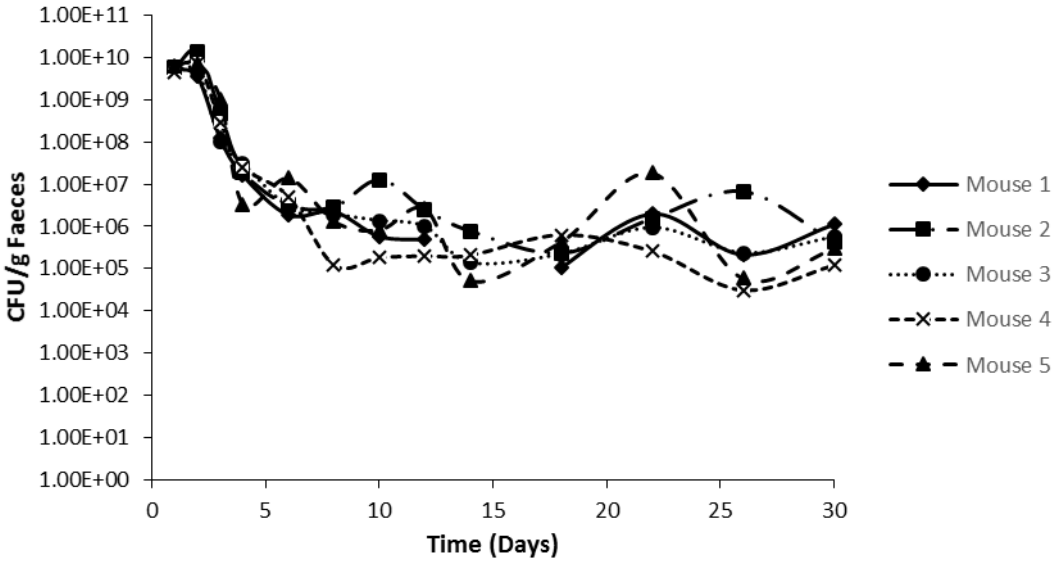


Figure 3.9: CFUs for total *E. coli* from mouse faecal samples. CFU counts peaked at day 2, and then declined to stabilise at approximately 10^6 CFU/g at the end of the experiment. Mice were monitored for 30 days. Results show faecal CFU/g for five mice ± standard error of the mean.

One strain, GMB98-30438, was dominant throughout the experiment, making up at least 60% of the entire population at day one in all mice, which increased to over 80% by day six (figure 3.10). GMB98-30438 was displaced as the dominant strain only in mouse one, where GMB71-B10, GMB72-B8 and GMB54-33301 became more abundant after day 14. The two versions of GMB45 also became more prominent later in the experiment, as was observed in the first competition experiment. In agreement with the results from the first mouse experiment, variation was observed between GMB45-19439 and GMB45-23353, with both appearing at different proportions in each mouse (figure 3.10).

When the distribution of the remaining strains was examined, there were clear patterns emerging that suggested a shift in the prevalence of each strain between day three and 14. During the first few days, GMB32-20239 and GMB34-33233 became the second most prevalent group, making up at least 70% of strains other than GMB98-30438 on day three for all mice (figure 3.11). Compared to the first competition study, strain GMB07-B19 colonised the gut at a much higher level during the first two weeks of the second experiment. It is possible that the presence of other strains not included in the first experiment facilitated the increased persistence of GMB07-B19. Finally, GMB23-23277,

GMB18-B5 and GMB104-29884 colonised the GI-tract poorly for all mice. Importantly, GMB23-23277 was the only strain to become undetectable by day 30 using high-throughput sequencing in four of the mice.

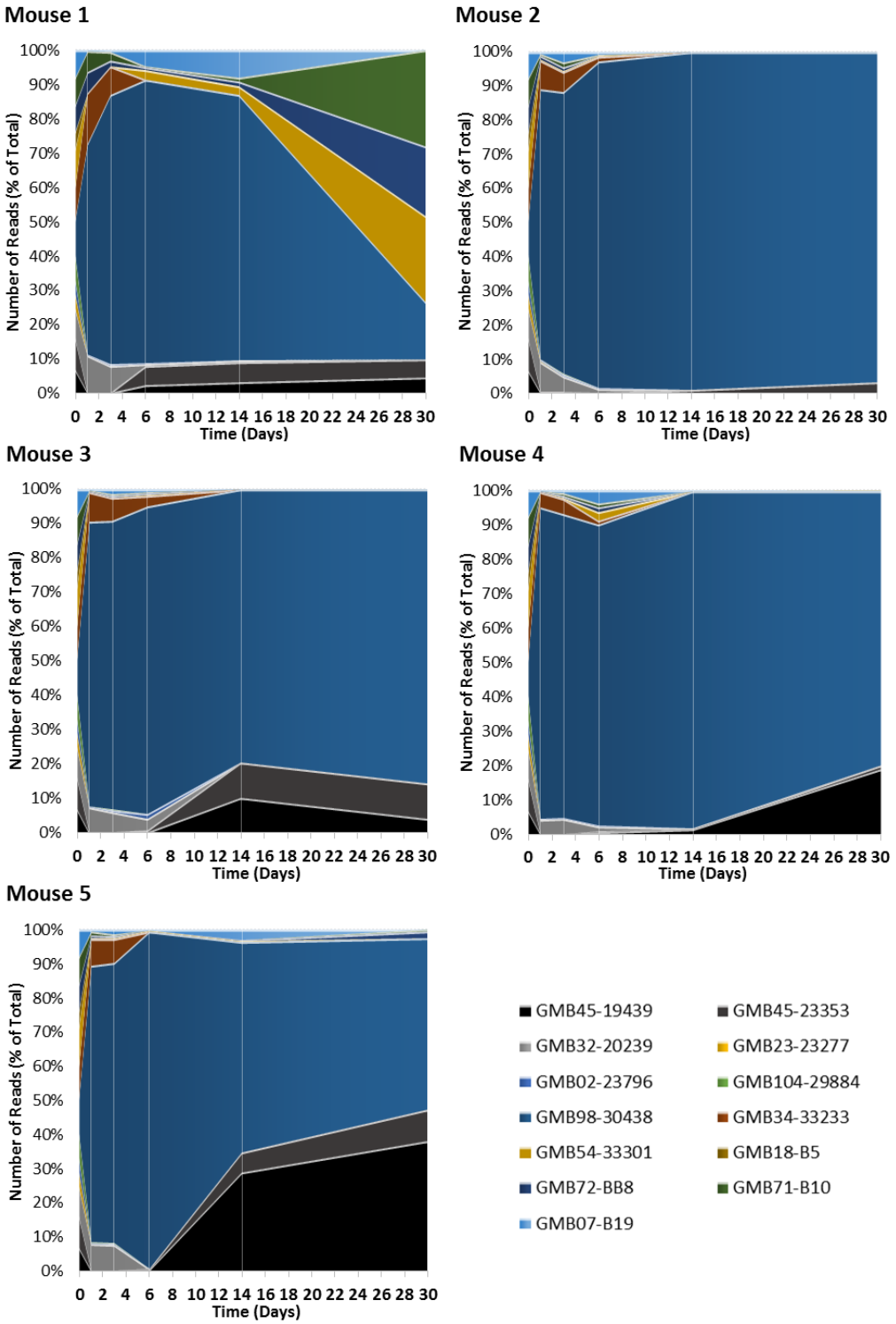


Figure 3.10: Stacked area graphs of the proportions of each barcode strain in mouse faecal samples as a percentage of the whole *E. coli* population for *in vivo* experiment 2. Graphs show proportions of barcoded strains for five individual mice calculated from high-throughput sequencing quantification of barcodes. GMB98 was dominant in all mice throughout the experiment, except for mouse one on day 30. Samples sequenced were taken on days 1, 3, 6, 14 and 30 after oral gavage. Day 0 strain proportions were calculated from the bacterial mixture used to gavage mice.

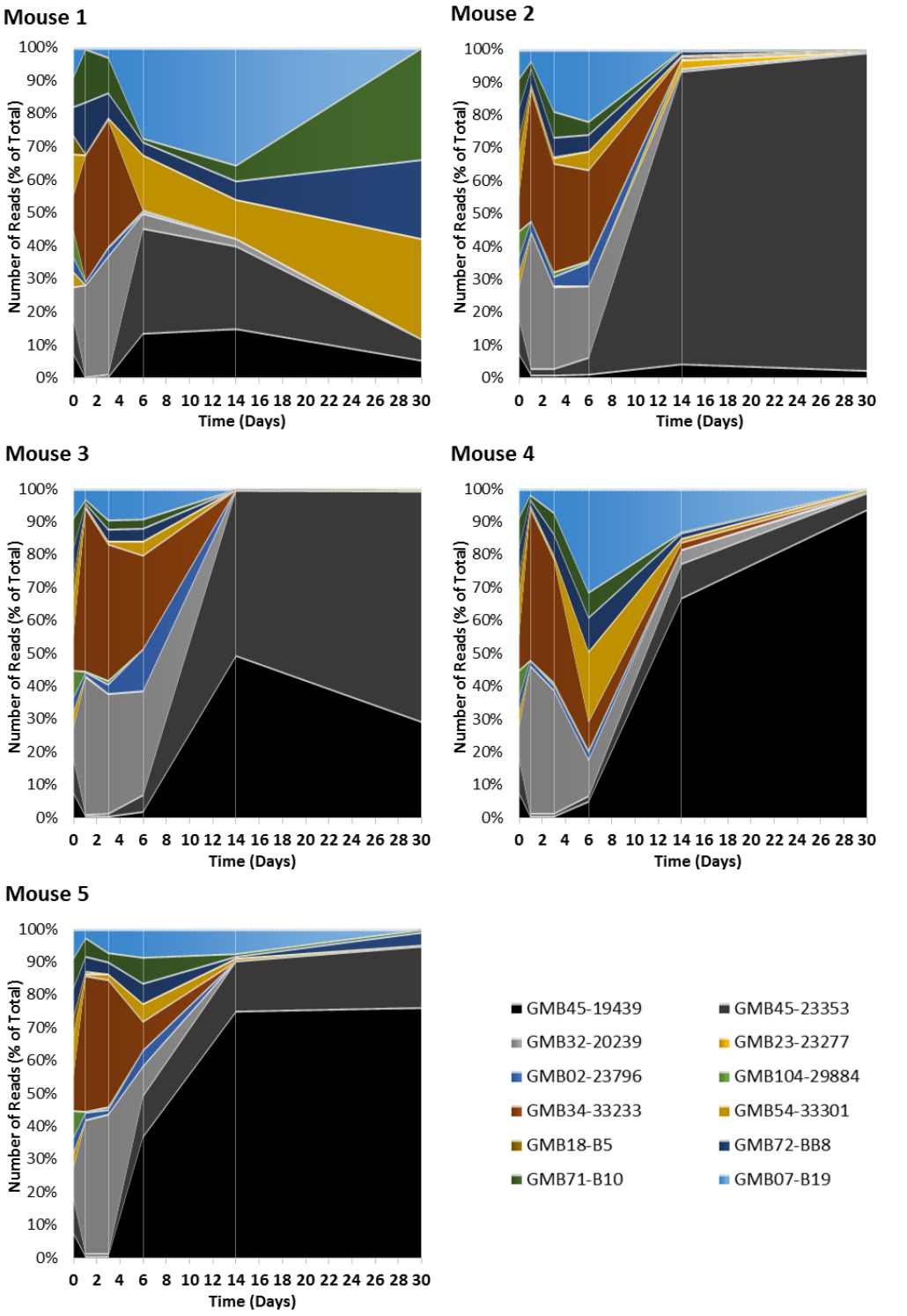


Figure 3.11: Stacked area graphs of the proportions of sub-dominant barcode strains in mouse faecal samples as a percentage of the whole *E. coli* population for *in vivo* experiment 2. Graphs display proportions of barcoded strains for five individual mice calculated from high-throughput sequencing. Dominant strain GMB98-30438 was excluded from the analysis. In sub-dominant strains, there was a clear shift in proportions between the first six days and days 14 and 30. Samples sequenced were taken on days 1, 3, 6, 14 and 30 after oral gavage. Data for day 0 is the proportions of each strain in the mixed-strain preparation given to mice via oral gavage.

3.5.2.1 Validation of high-throughput sequencing using RT-PCR

To compare the use of high-throughput sequencing with RT-PCR and ensure that the method of analysis did not influence the results, three samples (gavage mixture, mouse 2 day 3 and mouse 5 day 14 faecal samples) that were analysed using high throughput sequencing from the second mouse experiment were also analysed using RT-PCR. The results showed similar distributions of each strain for the three samples analysed (figure 3.12). There were minor differences for some of the barcoded strains, with the proportions of GMB104-29884, GMB32-20239 and GMB54-33301 being slightly increased in the RT-PCR. However, the high-throughput sequencing results indicate that it has increased sensitivity compared to the RT-PCR, as two strains that were undetectable in the RT-PCR (GMB02-23796 and GMB34-33233) were present in the sequencing results for the mouse 5 day 14 (M5D14) sample. These results indicate that both RT-PCR and high-throughput sequencing can be used to quantify the amount of each barcoded strain in mixed-strain populations.

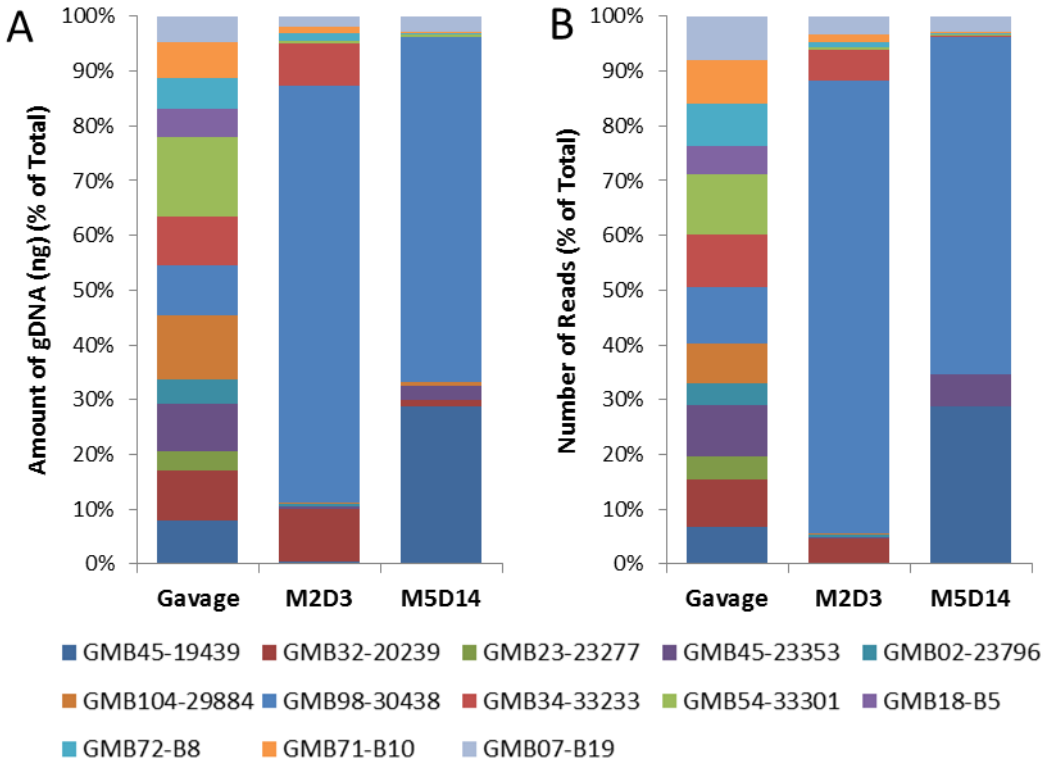


Figure 3.12: Proportion of each strain in the gavage mixture, mouse 2 day 3 (M2D3) sample and mouse 5 day 14 (M5D14) sample. Both (A) RT-PCR and (B) high-throughput sequencing was used for detection of barcodes. GMB104-29884, GMB32-20239 and GMB54-33301 are detected at higher levels in the RT-PCR, and as a result, can only be clearly seen on the RT-PCR graph for the M5D14 sample. GMB02-23796 and GMB34-33233 could only be detected by high-throughput sequencing in the M5D14 sample, but at such low levels that they cannot be visualised on the graph above.

3.6 *E. coli* population dynamics in the external environment using a soil competition study

The next experiment was performed to assess whether the barcoding method could be used for assessing the competitiveness of *E. coli* strains in a complex environment other than the gut. A mixture of 10 barcoded strains (see table A.1 for strain information) was used to inoculate the surface of soil columns (soil collected from a Norfolk farm) at room temperature, followed by incubation in the dark at 20°C over a period of 24 days. Samples were taken at each time point from the top, middle and bottom layers of each soil column. CFU counts showed that there was a loss of viability of *E. coli* used to inoculate soil columns during the first 12 days, with population stabilising at 10^4 CFU/g thereafter

(figure 3.13). gDNA was also extracted from soil samples and used to determine barcoded strain survival and persistence using RT-PCR.

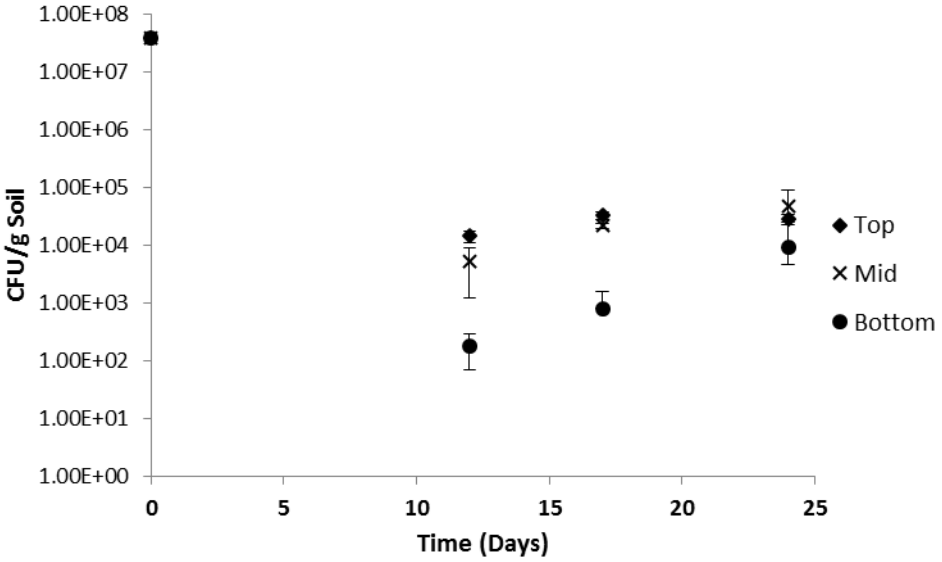


Figure 3.13: CFUs for total *E. coli* from top, middle and bottom layers of soil columns. CFU counts were reduced by three logs after 12 days, compared to the inoculum. CFUs were relatively stable in the top and middle layers of soil throughout the experiment, at about 10^4 CFU/g. CFUs increased in the bottom layer of soil between day 12 and 24. Results show average soil CFU/g for three soil columns ± standard error of the mean.

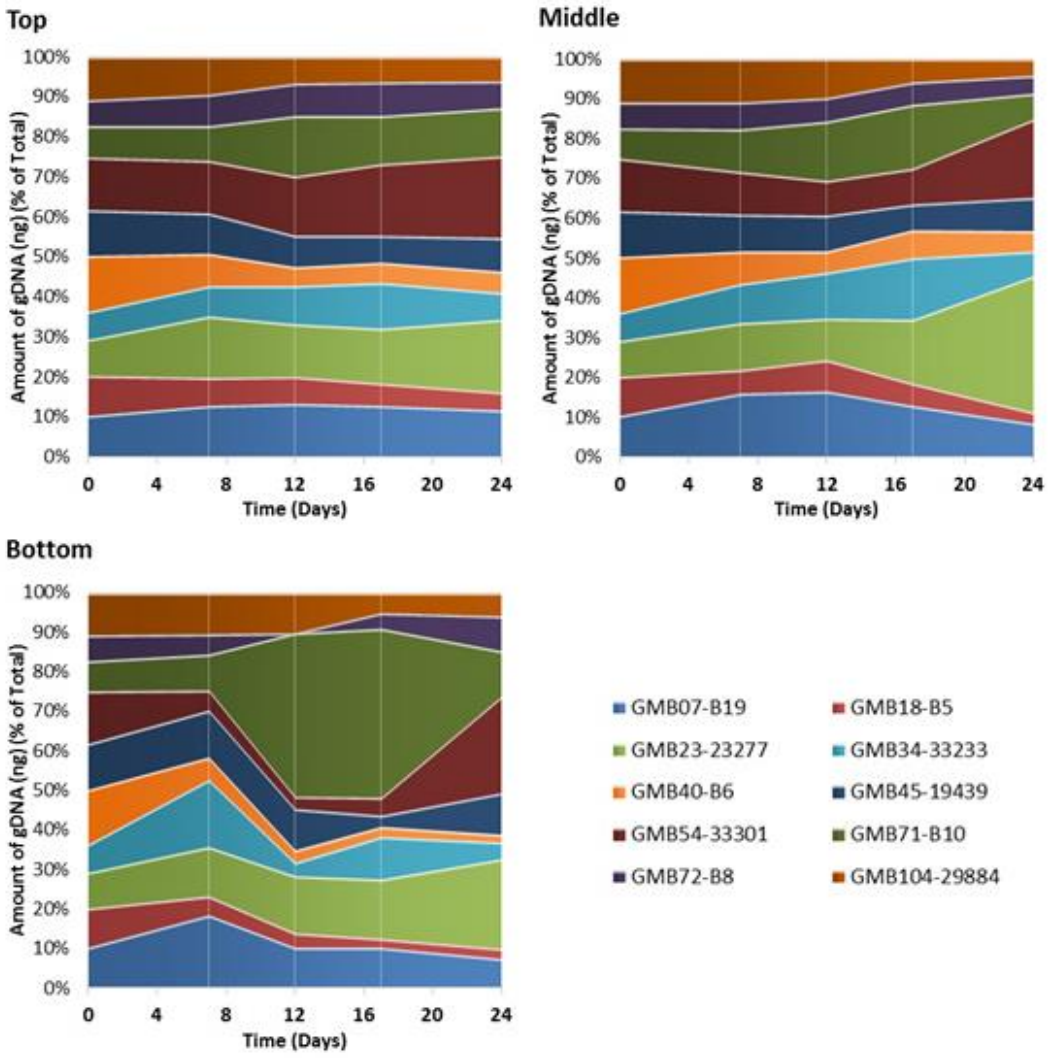


Figure 3.14: Stacked area graphs of the proportions of each barcode strain as a percentage of the entire *E. coli* population in the top, middle and bottom layer of soil columns. Graphs display the average proportions (*n* = 3) of barcoded strains from middle, bottom and top layers of soil columns. Barcoded strains were quantified using RT-PCR. Strain proportions stayed relatively consistent for the duration of the experiment. Data for day 0 represents the proportion of each strain in the mixed-strain preparation used to inoculate the surface of soil columns. Samples were taken on days 7, 12, 17 and 24.

The proportions of each strain remained fairly stable throughout the experiment, except for the bottom layer of soil, where the overall number of bacteria (CFU/g) increased by over 1 log, suggesting that there was growth of bacteria during the experiment (figure 3.14). GMB71-B10 in particular had increased numbers in the bottom layer of soil, which suggests that in comparison to the other strains it was more adapted to the environment at the bottom of the soil column. Alternatively, GMB71-B10 may have increased

motility, allowing it to access the lower layers of soil more readily than other strains. Other trends were beginning to appear on day 24, such as GMB40-B6, GMB104-29884, GMB72-B8 and GMB18-B5 declining, while GMB23-23277 and GMB54-33301 were becoming larger proportions of the whole *E. coli* population. The population was much less dynamic compared to those in the GI-tract, possibly reflecting the poor availability of nutrients in soils, such as organic carbon (Klein and Casida, 1967; Ishii et al., 2006; van Elsas et al., 2011). Sensitivity of the method was very good as bacteria present at levels of 10^3-10^4 CFU/g could be detected and quantified. It might be interesting in the future to extend this study beyond 24 days to determine whether differences in fitness are exacerbated.

3.7 Discussion

The results shown in this chapter indicate that the novel barcoding method used in this study is capable of monitoring individual strains in parallel in mixed-strain *E. coli* populations with high sensitivity and reproducibility. Compared to currently available techniques for identifying and quantifying large numbers of strains in complex environments, this barcoding method allows quick, easy and reliable identification of strains using either RT-PCR or high-throughput sequencing.

The rate of insertion of the barcode sequence was relatively high, with transposition occurring in 77% of transformed strains after two attempts. Of these successful strains, 81% of had no alterations in phenotype, resulting in an overall success rate of 62%. Once barcoded strains were constructed, they could quickly and easily be used in mixed competition studies. For the RT-PCR, accurate detection of barcoded strains was possible down to a concentration of 0.01pg of gDNA. Results from the validation of the high-throughput sequencing suggest that it is even more sensitive than the RT-PCR as it can detect a greater number of barcoded strains. The high level of reproducibility between experiments suggests that we were able to identify strains with increased or decreased fitness. Importantly, it appears that these differences in strain dominance are non-random, and are not determined solely by stochastic

events. Large numbers of strains could potentially be used in mixed population studies; however, it is important to note that populations that experience bottlenecks, or increase in size above the carrying capacity of an environment, may experience stochastic loss of strains. This could lead to strains being incorrectly identified as having reduced fitness in a given environment, and also lead to inconsistencies between replicates. On the other hand, this would provide important information on how bottlenecks affect *E. coli* population structure.

In the LB multi-strain competition study most of the strains were able to coexist even though it was a homogeneous environment. This may be because LB medium is rich in nutrients, giving multiple strains enough carbon sources to survive together. The main carbon sources of LB are amino acids derived from tryptone and yeast extract. BIOLOG carbon utilisation profiles of the seven strains used in the mixed culture competition studies did not indicate any differences in amino acid catabolism that may have resulted in the fitness differences that were observed between the isolates. Bacteria often display preferential use of certain nutrients and only switch to using other nutrients when the preferred one has run out (Stülke and Hillen, 1999). Shifts in dominance in the population may represent when the bacteria have switched to alternative carbon sources, with the strain most efficient at utilising the new carbon source becoming dominant.

Changes in utilisation of carbon sources or other nutrients can be a result of cross-feeding. This is when bacteria are able to catabolise metabolic breakdown products secreted by neighbouring cells (Pfeiffer and Bonhoeffer, 2004). Some *E. coli* have also been shown to utilise nutrients derived from dead cells in batch culture experiments (Farrell and Finkel, 2003). This has been linked to the growth advantage in stationary phase (GASP) phenotype, where strains are able to grow and persist under severe nutrient deprivation (Finkel, 2006). In our mixed culture experiments, GMB71-B10 displays continued growth up to 92h, suggesting that it may have the GASP phenotype. This phenotype has been linked to mutations in the *rpoS* and *lrp* genes (Zambrano et al., 1993; Zinser and

Kolter, 2000) which result in changes in nutrient use and the stress response of cells. Extending the experiment may have resulted in the loss of some strains from the mixed cultures as nutrients became increasingly unavailable. However, in batch cultures there is no wash out, so strains might persist for very long periods without growing.

The LB culture reached its highest OD_{600} reading after approximately four hours, after which it declined, suggesting that nutrient limitation may have started occurring at this point. This coincides with the reduction of ECOR16, which was observed between three and seven hours, indicating that it was unable to adequately compete for declining resources. Importantly, this strain grows as well as the other strains when grown in monoculture (data not shown), suggesting that factors other than the ability to utilise particular nutrients may be causing it to decline so rapidly. Alternatively, ECOR16 could be sensitive to a toxin, for example colicin, produced by another strain used in the mixed culture. Production of bacteriocins is usually induced during the SOS response, which can be caused by nutrient limitation (Majeed et al., 2011). Bacteria are able to interact with each other not only through the secretion of metabolites or bacteriocins, but also by making direct contact. Cross-feeding has been shown to occur between bacteria through the use of intercellular nanotubes, for example, where cells share cytoplasm (Pande et al., 2015). It would be interesting to repeat the culture competition experiment with the strains separated by membranes that allow the passage of secreted molecules but not direct cell-cell contact to see whether the same population dynamics would occur.

In the mouse GI-tract, we observed a shift in dominance between strains from day three and nine in the first mouse experiment, and day three to six in the second experiment. There are several factors that may influence this change, including the restoration of the microbiota, the recovery of the gut following antibiotic administration and changes in nutrient use by strains. Administration of antibiotics to mice has been shown to reduce bacterial load and diversity in the gut (Manichanh et al., 2010; Carvalho et al., 2012). Bacterial load has been

shown to be restored following cessation of antibiotic treatment after nine days in mice (Linninge et al., 2015) and seven days in humans (Panda et al., 2014). We observed a reduction in the bacterial load of *E. coli*, as CFUs decreased from 10^9 CFU/g to 10^4-10^6 CFU/g, which does not occur when antibiotics are given to mice continuously (Pi et al., 2012). This could be due to the microbiota being restored. The antibiotics used in this study, ampicillin and neomycin, are both broad spectrum antibiotics that have activity against several Gram-negative bacteria, including *E. coli*, as well as some Gram-positive species. It has been shown that administration of ampicillin and neomycin to mice reduces the bacterial load of the gut by 90%, as well as enlarging the caecum (Vijay-Kumar et al., 2010). How the microbiota recovers following removal of antibiotics is less well characterised, so it is difficult to determine exactly how the microbiota is affecting the barcoded *E. coli* as it returns. It is possible that increased competition with the microbiota for nutrients or space results in some of the strains that are less fit in the GI-tract being unable to persist.

It has also been shown that treatment of mice with streptomycin induces an inflammatory response by the gut epithelium that favours growth of *E. coli*, especially those that are capable of nitrate respiration (Spees et al., 2013). As a result, withdrawal of antibiotics may result in a reduction in *E. coli* as well as a change in dominance as strains more adapted to the inflamed gut decline. However, it is not known whether ampicillin and neomycin induce inflammation. Alternatively, different strains may adopt different colonisation strategies that result in a variation in persistence within the GI-tract. For example, the shift in dominance may occur as preferred nutrients in the GI-tract run out and strains switch to alternatives. Carbon source utilisation by *E. coli* has been shown to change between initial colonisation and persistence within the gut (Chang et al., 2004). It is possible, therefore, that the strains that become dominant later in the experiment are better able to utilise these new carbon sources.

Looking at the phylogenetic groups of the strains that demonstrated increased prevalence in faecal samples, both GMB98 and GMB45 belong to the B2 group,

which has been associated with increased adaptation to the GI-tract (Nowrouzian et al., 2005). The three strains that displaced GMB98 in mouse one, GMB54, GMB71 and GMB72, all belong to the phylogenetic group D. It has been suggested that the A/B1 groups and B2/D groups have different colonisation niches, with the B2 and D groups showing a higher degree of host specialisation (Escobar-Páramo, Grenet, et al., 2004). It is important to highlight, however, that few strains were used in these experiments, so further studies are needed to confirm these general observations about phylogeny and colonisation ability.

It is not clear why there was a shift in mouse one to the three strains GMB54, GMB71 and GMB72, but possibilities include the acquisition of deleterious mutations in GMB98 resulting in reduced fitness. This, however, is unlikely, as deleterious mutants would be expected to be displaced while the wild-type remained in the GI-tract at high levels. Alternatively, beneficial mutations could have occurred in GMB54, GMB71 and GMB72, though this also seems unlikely to have occurred in all three strains simultaneously. It is possible, however, that a beneficial mutation occurred in one of these strains which altered the gut environment to the benefit of the other two strains, or directly inhibited the growth of GMB98.

Phylogroup A has also been shown to be prevalent in the GI-tract, especially of humans (Duriez et al., 2001; Escobar-Páramo, Grenet, et al., 2004; Pallecchi et al., 2007; Tenaillon et al., 2010). However, in our competition studies, the majority of strains from the A group did not colonise the gut well (GMB02, GMB104 and GMB23), with only GMB32 and GMB34 being present at a relatively high proportion. A study on the distribution of *E. coli* in animals in Australia showed that the B1 and D groups were predominant in wild mouse (*Mus musculus*) faecal samples (Gordon and Cowling, 2003). This may explain why the majority of group A strains used in our studies were only detected at low levels. However, this raises the important question of how accurately a mouse model can represent what occurs in humans. Physiological, behavioural and ecological differences between mice and humans may limit the extent to

which these studies can be used to infer *E. coli* colonisation and persistence in the human GI-tract. However, the Gordon and Cowling (2003) study results contrast our observations that the B2 phylogenetic group is often dominant. Domestication is an important factor in influencing the phylogenetic distribution of *E. coli* in animals (Tenaillon et al., 2010), which may explain the differences between wild and laboratory mice. Also, the wild mice sampled in the Gordon and Cowling (2003) study were constantly exposed to *E. coli*, so direct comparisons to our study where a single dose of bacteria were given are difficult.

The early loss of all ECOR strains from the mouse gut, despite them being host isolates, could indicate that these isolates are host-specific, as ECOR16 and ECOR55 were isolated from leopard (*Panthera pardus*) and human faeces respectively (Ochman and Selander, 1984). However, specialisation to a specific host group or species has not been observed in many *E. coli* and is generally considered as rare (Escobar-Paramo et al., 2006). Moreover, this hypothesis could be applied to the GMB isolates which are considered to be in the external environment following a faecal contamination event. As a result, their primary environment should be in the GI-tract of the host responsible for the contamination event. It is more likely that, as the ECOR collection is over 30 years old compared to the GMB collection isolated in 2008-2009, that the ECOR strains have become laboratory-adapted or damaged, so that they no longer reflect the characteristics of *E. coli* strains found in the GI-tract. As a result of this, ECOR strains were not used in any of the following competition studies. In contrast to the ECOR strains, only one GMB strain was drastically reduced throughout the experiment, which was GMB07-B19. This strain belongs to the B1 phylogenetic group and has been shown to have traits associated with adaptation to the non-host environment and plant colonisation (Méric et al., 2013).

Barcoded strains detected in the gut contents of mice support observations that the major *E. coli* phylogenetic groups are heterogeneously distributed throughout the GI-tract in humans (Gordon et al., 2015), and that *E. coli*

detected in faeces are not necessarily representative of gut contents (Gordon et al., 2015). Gut content gDNA samples were also collected during mouse experiment 2, but are awaiting sequencing. Once analysed, these should give a clearer indication of how *E. coli* are distributed throughout the GI-tract.

Populations of *E. coli* in particular environments outside the host, in this case soil, may be less dynamic due to lack of nutrients or *E. coli* not being as well adapted to these environments as it is to its primary niche. Survival of *E. coli* in soil has been shown to be influenced by temperature, soil texture, moisture levels, nutrient availability and the presence of a soil microbiota (Ishii et al., 2010). Increased amounts of organic matter in soil have been linked with increased levels of growth by *E. coli*, although this increase in population size was only detected at temperatures over 25°C (Ishii et al., 2006). However, growth has been detected at temperatures of 19°C in manure rich soil (Berry and Miller, 2005) suggesting that nutrient rich soils can support growth of *E. coli* at lower temperatures. For our soil competition experiment, soil columns were incubated at 20°C, and from the CFU counts there were no significant changes in the population size, other than in the bottom layer of soil. One factor that may influence fitness and growth in soil is oxygen availability. An oxygen gradient is present in soil, with higher amounts of oxygen available closer to the surface due to shorter diffusion distances. Several soil characteristics, including moisture, density and temperature, are known to influence the ability of oxygen to diffuse through soil from the air (Stępniewski and Stępniewska, 2009). It is possible that GMB71-B10 and other strains that have increased numbers lower down the soil column may be more adapted to environments that have reduced oxygen availability.

Our soil competition experiments indicate that the initial die-off of *E. coli* in soil was comparable for all strains, suggesting that in the short term all strains and phylogenetic groups have a similar fitness level. After 24 days, however, some trends were emerging, notably GMB40, GMB104, GMB72 and GMB18 were declining, while GMB23 and GMB54 were becoming larger proportions of the whole *E. coli* population. Several studies have shown that a large proportion of

strains that are isolated from the external environment are from the B1 phylogenetic group, so it was interesting that GMB18, which is a B1 strain, had one of the lowest levels on day 24. GMB07 was another B1 isolate, but it appeared to persist at a relatively constant proportion of the population throughout the experiment. Of course, as few strains were used in this experiment, it is hard to extrapolate which phylogenetic groups have increased persistence in soils, but future studies using larger numbers of strains may help identify soil adapted *E. coli*.

The barcoding method described in this chapter has the potential to provide a powerful tool for assessing population dynamics in *E. coli* populations, which can be adapted to other bacteria. Importantly, this technique has a high sensitivity that allows for the identification of *E. coli* present in the population at low levels, which may be undetected using currently available methods. As a result, many studies have focussed on the persistence and growth of dominant strains. Potentially, this technique can be used on large numbers of strains, and together with genome-wide association studies, identify adaptive traits associated with the environment.

4. Siderophore gene distribution and production in commensal *E. coli* of faecal and plant origin

The results in this chapter form part of the paper 'Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *E. coli*' by Laura J. Searle, Guillaume Méric, Ida Porcelli, Samuel K. Sheppard and Sacha Lucchini published in the journal *PLoS One* (see appendix D). I carried out all experimental work except the siderophore production assays on chrome azurol S (CAS) agar plates, which were carried out by Guillaume Méric.

4.1 Introduction

In the previous chapter the possibility of determining whether specific strains of *E. coli* had increased fitness in the gut using a novel barcoding system was discussed. Barcoding, alongside genome-wide association studies have the potential to identify specific traits associated with gut colonisation by *E. coli*. One trait that has been identified as being involved in GI-tract colonisation by *E. coli* is siderophore production, which will be discussed in this chapter.

For most bacteria, iron is an essential element that, due to its versatility both as a ferrous (Fe^{2+}) and ferric (Fe^{3+}) ion is incorporated into a wide range of enzymes and proteins that are used in many different cellular processes including DNA replication, respiration and protection from oxidative stress. In the presence of oxygen Fe^{2+} is rapidly oxidised into the less soluble Fe^{3+}, making it far less available to bacteria. To maintain intracellular iron levels of 10^{-7} to 10^{-5}M (Garenaux et al., 2011), bacteria have developed several mechanisms to scavenge iron from the surrounding environment which involve increasing the solubility and availability of Fe^{3+}. These mechanisms include lowering the pH of the surrounding environment, the production of molecules known as siderophores that bind Fe^{3+} with high-affinity and facilitate uptake into the cell, or reducing Fe^{3+} to Fe^{2+} which is more readily internalised into the cell (Guerinot, 1994). The main mechanism through which *E. coli* takes up Fe^{3+} is through siderophores. Most *E. coli* and many other *Enterobacteriaceae* are able

to secrete the siderophore enterobactin, however up to three other siderophore systems have been found in some strains of *E. coli*. These are aerobactin, salmochelin and yersiniabactin.

These three siderophores have all been linked to virulence in pathogenic strains of *E. coli*, especially ExPEC strains that cause disease in areas outside of the GI-tract. However, a significant proportion of commensal *E. coli* strains have these three additional siderophores and the majority of ExPEC strains are thought to live as commensals as part of the healthy microbiota (Köhler and Dobrindt, 2011), raising an important question about the role that siderophores have in the ecology of *E. coli*. This includes a possible role in both the primary host gut environment as well as the secondary external non-host environment, where *E. coli* populations can persist for long periods of time (van Elsas et al., 2011).

E. coli is constantly switching between the host and non-host environments due to repeated faecal deposition followed by ingestion by the host (faecal-oral route transmission). Faecal-oral transmission requires bacteria to survive in these secondary environments as they pass between hosts. To reflect this, *E. coli* have been isolated from many different secondary environments, including soil, surface and groundwaters and from vegetables and salad crops (van Elsas et al., 2011). Long-term persisting strains of *E. coli* have been isolated from secondary environments, including tropical soils (Byappanahalli et al., 2006; Goto and Yan, 2011), water (Bermúdez and Hazen, 1988; Power et al., 2005; Vital et al., 2008; Goto and Yan, 2011), sediments (Solo-Gabriele et al., 2000; Whitman and Nevers, 2003; Ishii et al., 2007) and plants (Solomon et al., 2003; Islam et al., 2004; Ibekwe et al., 2007). It was postulated that tropical soils and waters provided a warm and moist environment that replicated the GI-tract sufficiently to permit growth of *E. coli* in these environments (Winfield and Groisman, 2003). However, the same *E. coli* isolates could be detected over long periods of time in soils from temperate climates (Ishii et al., 2006; Ishii et al., 2007; Texier et al., 2008; Brennan, Abram et al., 2010; Brennan, O'Flaherty et al., 2010), suggesting that *E. coli* is able to form sustainable populations in the

secondary environment as naturalised *E. coli* (Byappanahalli et al., 2006; Ishii et al., 2006).

The role that siderophores have in the gut and in the external non-host environments is unclear; however, there is evidence for the use of siderophores in both. Enterobactin production and receptor *E. coli* K-12 strain MG1655 mutants were outcompeted by wild-type *E. coli* in the murine GI-tract (Pi et al., 2012). Importantly, mutants were still able to colonise the gut when given to mice alone, suggesting that although not essential to survival within the gut, siderophores increase competitiveness. However, Kupz et al. (2013) found no significant reduction in fitness of an *E. coli* strain Nissle 1917 *ΔentC* mutant, but there was a trend towards lower CFUs, with the mutant being undetectable in some mice after 3-7 days, compared to the wild-type. As *E. coli* strain Nissle 1917 has all four siderophore systems, it is possible that other siderophore systems were able to compensate for the loss of enterobactin (and salmochelin) production. The *E. coli* K-12 MG1655 strain used by Pi et al. (2012) only has the enterobactin siderophore, which may be why there was a discrepancy between these two studies. *E. coli* strain Nissle 1917 has been suggested to outcompete pathogenic *S. enterica* in mice by competing for iron through the production of siderophores (Deriu et al., 2013).

Siderophores have also been shown to influence survival in external environments which may be iron-poor (Jurkevitch et al., 1992; Cornelis, 2010; Diallo et al., 2011). Enterobactin biosynthesis genes were shown to improve plant colonisation in *S. enterica* Typhimurium, and both enterobactin and salmochelin genes were upregulated during plant colonisation (Hao et al., 2012). Several phytopathogenic strains of bacteria, including *Erwinia amylovora*, *Erwinia chrysanthemi*, *Pseudomonas syringae* and *Dickeya dadantii*, have been shown to utilise siderophores during infection of host plants (Dellagi et al., 1998; Dellagi et al., 2005; Franza et al. 2005; Taguchi et al., 2010).

This contrasting evidence on the role of individual siderophore systems in *E. coli* and related bacterial species probably indicates that the genomic context is an

important factor. Since the distribution of a phenotype is influenced by the selection pressures of the environments, with traits that increase fitness becoming more prevalent (Bergholz et al., 2011; Méric et al., 2013), a population-level comparative approach should thus provide important complementary information. In this chapter, we build on the initial observation that siderophore production on chrome azurol S (CAS) agar plates was increased in *E. coli* isolates from healthy faecal samples (ECOR-F) compared to strains isolated from plants (GMB) (Méric, 2011).

4.2 Validation of the comparison of siderophore production by host and plant *E. coli* isolates

As ExPEC strains are known to be enriched for virulence genes, including siderophores (Escobar-Páramo, Clermont et al., 2004), the ECOR strains isolated for women either with a urinary tract infection (UTI) or asymptomatic bacteriuria (Ochman and Selander, 1984) were removed to prevent any biases in the data. 11 ECOR strains (ECOR11, ECOR14, ECOR40, ECOR48, ECOR50, ECOR56, ECOR60, ECOR62, ECOR64, ECOR71 and ECOR72) were therefore excluded from any comparative analyses. The remaining ECOR isolates, termed ECOR-F, were used as examples of healthy faecal isolates for the comparison between host and non-host strains. A further four strains (ECOR29, ECOR52, GMB37 and GMB69) were excluded from the siderophore production analysis as they did not grow on CAS agar plates. Growth was not restored in MM9 medium (similar growth conditions to CAS agar) supplemented with 100µM ferrous sulphate ($FeSO_4$), suggesting that auxotrophic mutations prevent these strains from growing under these conditions, rather than reduced iron acquisition (data not shown).

CAS agar plates are able to detect total siderophore production through an iron-bound indicator dye in the agar that changes colour, from blue to orange, in the presence of siderophores (Payne, 1994). The total siderophore production of a colony was calculated using the diameter of the halo that appeared around siderophore producing colonies and standardising to colony diameter (figure

4.1A). The CAS agar assay showed that faecal strains (ECOR-F) had increased siderophore production compared to plant isolates (GMB) (figure 4.1B) (Méric, 2011; Searle et al., 2015). Indeed, a higher proportion of ECOR-F isolates included high siderophore producers (37% vs 11%) (figure 4.1C).

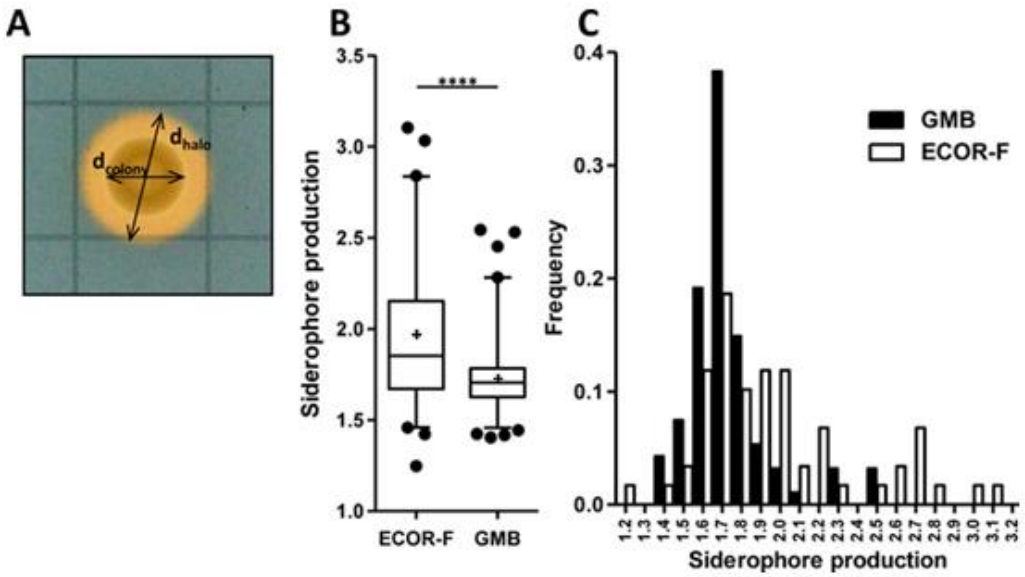


Figure 4.1: Siderophore production is higher in faecal isolates compared to plant-associated *E. coli*. A) Production on CAS agar was calculated by dividing the halo diameter (d_{halo}) by the colony diameter (d_{colony}). B) Box plot showing siderophore production for ECOR-F and GMB isolates ($n_{ECOR\text{-}F}$ = 57, n_{GMB} = 96). The median and interquartile ranges are displayed by the rectangle and the whiskers span the 5-95 percentile. Black circles represent outliers and plus signs represent the mean. Statistical significance was determined using the Student's *t*-test. ****P<0.0001. C) Frequency histogram showing siderophore production for both collections.

To exclude the possibility that the differences in siderophore production were not a result of siderophore-specific differences in diffusion, for example, a siderophore with a higher diffusion may be more prevalent in ECOR-F strains, resulting in a larger halo, liquid chrome azurol S (CAS) assays were performed on a representative subgroup of ECOR and GMB strains (*n*=33). The liquid CAS assay measures siderophore production in liquid culture supernatant, so diffusion rates between siderophores should be similar. The liquid assay confirmed the significantly higher siderophore production by host-associated strains compared to plant-associated strains (figure 4.2) as observed on CAS agar plates.

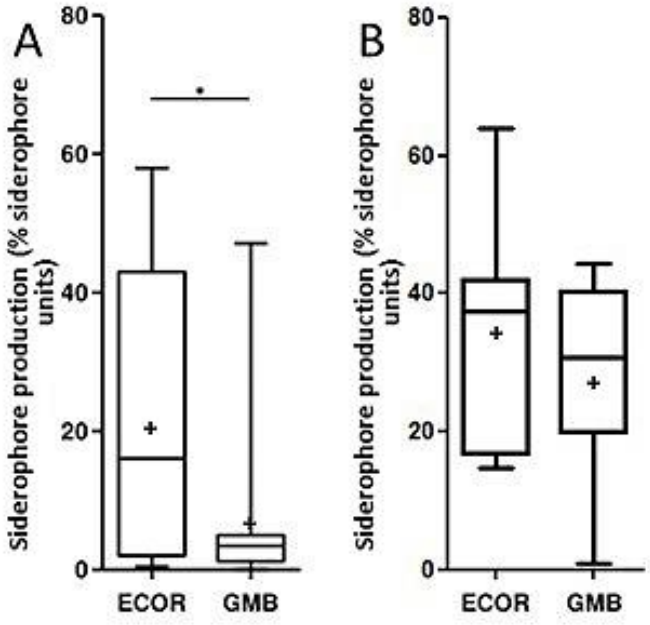


Figure 4.2: Siderophore production is increased in faecal *E. coli* compared to plant isolates when grown on glucose. Boxplots showing siderophore production for a subset of ECOR-F and GMB strains ($n_{ECOR\text{-}F}$ = 18, n_{GMB} = 15) grown on A) glucose and B) glycerol. The median and interquartile ranges are displayed by the rectangle and the whiskers span the 5-95 percentile. Plus signs represent the mean. Statistical significance was determined using the Student's *t*-test. *P<0.05.

Liquid CAS assays were also performed using supernatants from strains grown in MM9 containing glycerol, rather than glucose, as the sole carbon source. Glucose is known to induce carbon catabolite repression (CCR) in *E. coli*, which has been hypothesised as influencing siderophore gene expression (Zhang et al., 2005; Valdebenito et al., 2006). We observed a much greater level of siderophore production in most strains when grown with glycerol (figure 4.2). Interestingly, there was no longer a significant difference between the ECOR and GMB collections when strains were grown on glycerol, highlighting possible regulatory differences between the two collections.

4.3 Multiplex PCR design and validation

Both the liquid and agar CAS assays showed stronger siderophore production by the ECOR-F samples compared to the GMB isolates. However, this does not indicate which specific siderophores are responsible for increased siderophore production in ECOR-F isolates. Here we address the question of whether differences in siderophore production between the GMB and ECOR-F collections

were reflected at the genome level by a different complement of siderophore genes. To investigate this, a multiplex PCR was designed to detect one receptor and four biosynthesis genes for all four siderophore systems: enterobactin, salmochelin, aerobactin and yersiniabactin. As the salmochelin system only has one biosynthesis gene, two degradation genes and an export gene were included in the multiplex PCR alongside the receptor and biosynthesis genes.

4.3.1 Multiple Sequence Alignments

To design highly specific and sensitive primers for the multiplex PCR, publically available *E. coli* genome sequences (both commensal and pathogenic strains) were retrieved from both the EcoCyc (Keseler et al., 2013) and the National Centre for Biotechnology Information (NCBI) collections. The DNA sequences of all the siderophore production loci that could be identified through nBLAST similarity searches were then used to generate ClustalW multiple alignments. These enabled the identification of conserved regions for each target siderophore gene for the PCR. As well as finding conserved regions, the alignments highlighted some of the differences in distribution of each siderophore and the structure about each siderophore locus. The enterobactin locus was detected in every *E. coli* genome analysed, and the entire enterobactin locus appears to be maintained as a whole unit throughout. This supports the idea that siderophores have an important role in the lifestyle of *E. coli*. Of course, gene presence does not equate to production, so it is still possible that even though all these strains have enterobactin, they might not all be capable of producing or using it.

The other siderophore systems were far less prevalent than enterobactin, with yersiniabactin being the second most common. From the Ecocyc alignments, the prevalence for the 34 strains analysed was as follows: 14.7% for salmochelin, 26.4% for aerobactin and 41.2% for yersiniabactin. This was similar to alignments performed using the 54 *E. coli* strains available on NCBI, with prevalences of 14.8% for salmochelin, 25.9% for aerobactin and 33.3% for yersiniabactin. However, the strains available on both EcoCyc and NCBI were predominantly pathogenic strains (52.9% and 61.1% for EcoCyc and NCBI

respectively), which may influence how common each siderophore is, as siderophores have been linked to virulence. In fact, only three commensal strains (14.3%) of *E. coli* possessed a siderophore other than enterobactin. The salmochelin, yersiniabactin and aerobactin loci showed the same level of conservation as a whole unit, suggesting that once obtained, the ability to make these siderophores was retained.

4.3.2 Validation of the multiplex PCR using the ECOR collection

To determine the specificity of the multiplex PCR, results obtained using the ECOR collection were compared to previous studies that monitored siderophore gene presence in this collection of *E. coli* isolates. These studies used PCR-based approaches to identify the genes encoding the siderophore receptors (Johnson, Delavari, Kuskowski and Stell, 2001) and the high pathogenicity island (HPI) containing the yersiniabactin locus (Schubert et al., 2009). A more recent study using multigenome arrays was also included as it monitored the presence of all the genes within the enterobactin, salmochelin and aerobactin loci (Jackson et al., 2011). The results from our multiplex PCR for the enterobactin, aerobactin and salmochelin genes (1,080 genes in total) showed good correlation with the array data, with 96.4% (1,041 genes) results matching the Jackson et al. (2011) dataset. Comparisons with the Johnson et al. (2001) PCR data showed a 99.1% match for the 216 enterobactin, aerobactin and salmochelin receptor genes tested. For the yersiniabactin locus, our multiplex PCR results were different for only two ECOR strains (97.2% match) when compared to PCR results in Schubert et al. (2009) (see appendix C).

To rule out a strain identification error causing differences between our multiplex results and conflicting results from other published datasets, CRISPR (clustered regularly interspaced short palindromic repeats) regions were assessed. These CRISPRs, alongside CRISPR-associated genes (CAS) are part of a prokaryotic acquired immune system offering protection against phages and other mobile genetic elements by integrating short DNA sequences from them that allow identification and removal from the cell (Barrangou et al., 2007).

These CRISPR regions can become hypervariable as a result, and can be used to distinguish between different ECOR strains (Diez-Villasenor et al., 2010). The strains ECOR02 and ECOR67 were both shown in the multiplex PCR to not possess the aerobactin and salmochelin genes respectively, which was in contradiction to published data (Johnson, Delavari, Kuskowski and Stell, 2001; Jackson et al., 2011). However, the CRISPR2.1 and CRISPR2.3 regions DNA sequences were a match to available sequences. Strains ECOR11 and ECOR72 showed differences in yersiniabactin gene presence to previous PCR results (Schubert et al., 2009), but also contained the correct CRISPR sequences, confirming the identity of our strains. These discrepancies reflect that changes have been observed in the phenotype and genotype of *E. coli* strains between laboratories (Johnson, Delavari, Stell et al., 2001) as well as during transport of strains (Spira et al., 2011), highlighting the importance of characterising strains used in independent studies.

4.4 Comparison of siderophore gene distribution in plant- and host-associated *E. coli* isolates

4.4.1 Maintenance of siderophore loci

The distribution of siderophore genes in 96 isolates from the GMB collection was determined using multiplex PCR to compare to the ECOR collection to assess whether the distribution of siderophore genes is influenced by the environment from which strains were isolated. After performing the multiplex PCR on all isolates, it was evident that the siderophore receptor and biosynthesis genes were always detected together (133/133), highlighting not only the sensitivity of the PCR, but also that the siderophore loci are evolutionarily maintained as a complete unit. This agrees with what was observed in the multiple alignments.

The ECOR05 strain does not have a full siderophore operon; however, it also does not possess either the receptor or biosynthesis genes. It only possesses the hydrolysis genes *iroD* and *iroE* from the salmochelin locus. These genes, however, can be used in the hydrolysis of enterobactin as well as the

breakdown of salmochelin (Lin, 2005; Zhu, 2005). It is not possible to determine from the results whether these siderophore genes were acquired separately from the rest of the loci, or whether the strain has lost the rest of the salmochelin genes required for its synthesis and uptake.

4.4.2 ECOR and GMB siderophore gene presence

Comparing the distribution of siderophore genes between the plant- and healthy host-associated *E. coli* isolates, strains isolated from plants on average possessed fewer siderophore systems (figure 4.3), supporting the observed increased siderophore production in the ECOR collection (Méric, 2011; Searle et al., 2015).

Examining the specific siderophore systems, aerobactin and yersiniabactin were found at significantly lower proportions in the GMB collection. No significant difference in prevalence was seen for the salmochelin genes (table 4.1). Enterobactin was present in all strains, however, the proportion of strains that only possessed the enterobactin locus and had no secondary siderophore systems was significantly higher in the GMB collection, at 76% (73/96) compared to 42.6% (26/61) for the ECOR-F group (Fisher's exact test, $p<0.0001$; see appendix C for tables). Interestingly, the B2 phylogenetic group had an increased prevalence of siderophore genes compared to groups A and B1 (figure 4.4), possibly reflecting the association of the B2 group with the host environment (Nowrouzian et al., 2006)

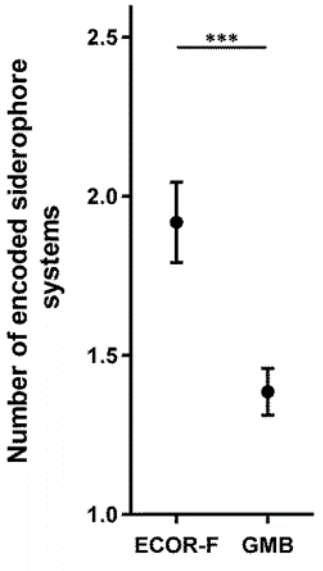


Figure 4.3: Faecal strains encode more siderophore systems compared to plant-associated isolates at the population level. Mean number of siderophore production systems for the GMB (*n* = 96) and ECOR-F (*n* = 61) collections are displayed. Error bars show the standard error of the mean. Statistical significance was determined using the Student's *t*-test. ***P<0.001.

Table 4.1: A greater proportion of faecal isolates possess the aerobactin and yersiniabactin loci compared to plant-associated *E. coli*. The prevalence data from faecal samples used to compare to ECOR-F and GMB results were obtained from recent human studies. The range in the number of isolates for these isolates reflects that not all studies determined the distribution of all four siderophore systems (aerobactin *n* = 1042, yersiniabactin *n* = 618, salmochelin *n* = 808). Significance was determined using the Fisher's exact test. In the case of multiple comparisons, the Benjamini and Hochberg False discovery rate method was used for correction. NS: no statistical significance.

Siderophore	Proportion of detected systems (%)			Statistical significance		
	GMB (n=96)	ECOR-F (n=61)	Faecal (n=618-1042)	GMB vs ECOR-F	GMB vs Faecal	ECOR-F vs Faecal
Aerobactin	5	25	29	<0.001	<0.001	NS
Yersiniabactin	19	48	35	<0.001	<0.01	NS
Salmochelin	16	20	21	NS	NS	NS

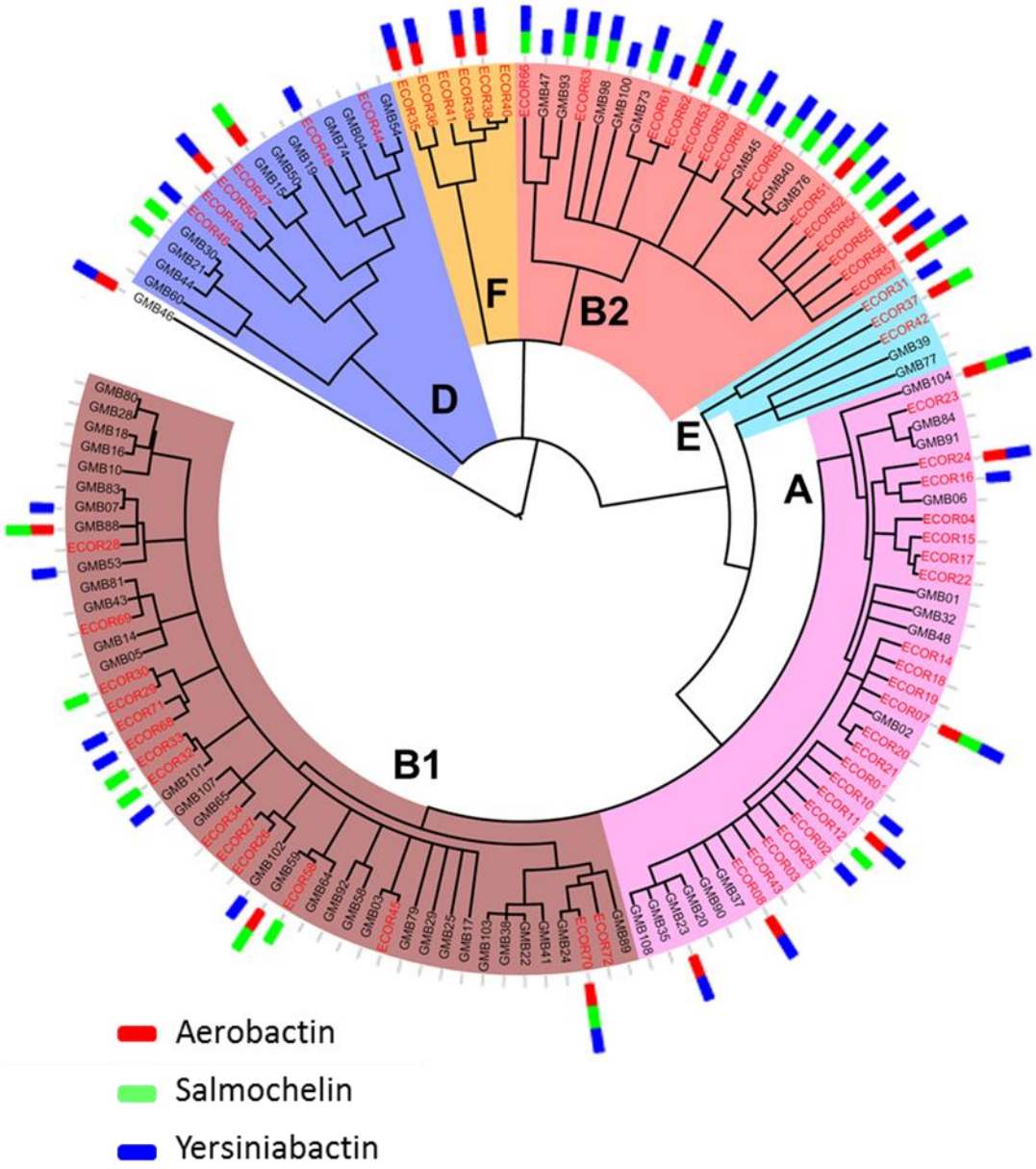


Figure 4.4: ClonalFrame phylogenetic tree showing the presence of three siderophore systems in the major phylogenetic groups of *E. coli*. Phylogenetic groups are based on MLST analysis using 8 housekeeping genes: *dinB*, *icdA*, *pabB*, *polB*, *putP*, *trpA*, *trpB* and *uidA* (Méric et al., 2013). ECOR (*n* = 66) strains are labelled in red and GMB (*n* = 67) strains in black (adapted from Méric et al., 2013).

4.4.3 Gene presence and siderophore production

The results from the multiplex PCR suggest that the differences seen between GMB and ECOR siderophore production levels may be linked to the aerobactin and yersiniabactin loci, which are more prevalent in ECOR-F isolates. Supporting this hypothesis, aerobactin was almost exclusively found in high producing strains (table 4.2). However, there was no link between yersiniabactin gene

presence and siderophore production on CAS agar plates. Aerobactin gene presence did not account for all the observable variation in siderophore production between GMB and ECOR-F isolates. In fact, when strains capable of producing only enterobactin were examined, the variation in siderophore production levels within each strain collection was different, with ECOR-F strains displaying higher siderophore production levels compared to GMB at the population level (unpaired t-test, $p<0.01$) (figure 4.5). This highlights the possibility that the host and non-host environments select for differential regulation of siderophore and iron uptake genes.

Table 4.2: Distribution of siderophore systems relative to siderophore production on CAS agar plates. Significance tests were performed comparing the top and bottom quartiles using the Fisher's exact test (n = 161). The Benjamini and Hochberg False discovery rate method was used to correct for multiple comparisons. *P<0.05, ***P<0.001.

	Top 25%	25-50%	50-75%	Bottom 25%
Aerobactin (%)	42***	5	3	3
Yersiniabactin (%)	42	24	32	31
Salmochelin (%)	26*	21	13	5

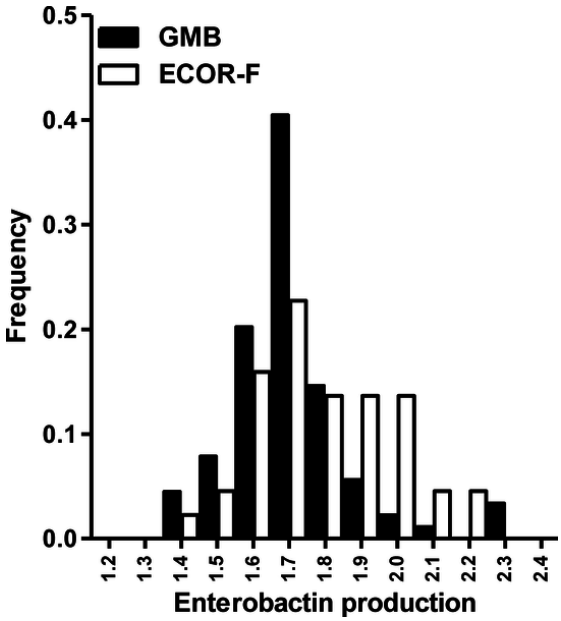


Figure 4.5: Enterobactin production is higher in faecal isolates compared to plant-associated *E. coli*. Frequency histogram comparing siderophore production on CAS plates, calculated by dividing the halo diameter by the colony diameter (d_{halo}/d_{colony}) (Méric, 2011), for GMB and ECOR-F strains only encoding the enterobactin locus. Sample sizes were $n_{GMB} = 69$, $n_{ECOR\text{-}F} = 28$.

4.4.3.1 Liquid CAS assay

To exclude the possibility that the link between aerobactin gene presence and high siderophore production was not a result of siderophore-specific differences in diffusion, we compared the results of the liquid CAS assay for strains with and without the aerobactin locus (figure 4.6). The liquid assay confirmed the clear link between high siderophore production and presence of the aerobactin locus (figure 4.6). Siderophore production of strains possessing the aerobactin locus was high in medium containing either glycerol or glucose. In contrast, other strains only displayed high siderophore production when grown on glycerol (figure 4.7).

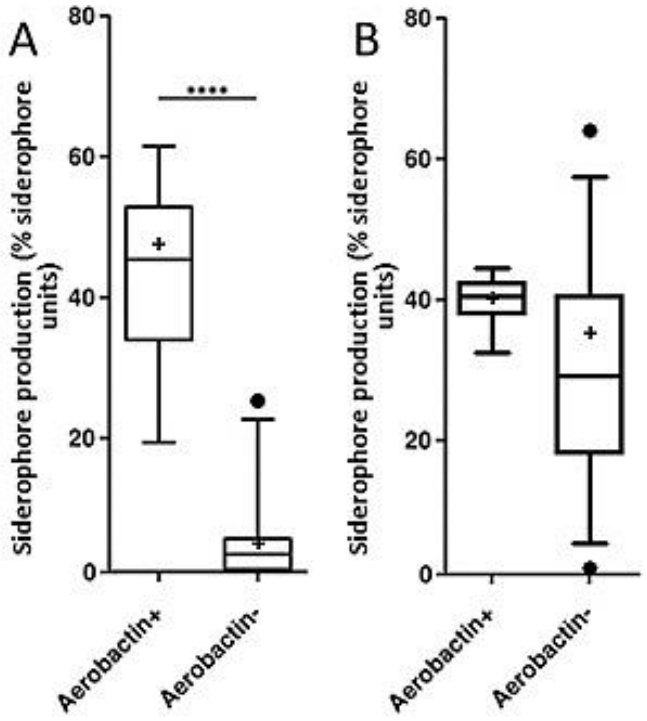


Figure 4.6: Siderophore production is higher in aerobactin producing strains. Boxplots showing siderophore production for a subset of ECOR-F and GMB strains with or without the aerobactin locus (n_{iuc+} = 10, n_{iuc-} = 23) grown on A) glucose and B) glycerol. The median and interquartile ranges are displayed by the rectangle and the whiskers span the 5-95 percentile. Black circles represent outliers and plus signs represent the mean. Statistical significance was determined using the Student's *t*-test. ****P<0.0001.

4.4 Discussion

Iron acquisition is essential to bacterial growth and survival during gut colonisation. The siderophore enterobactin has been shown to be important in

the mouse GI-tract for *E. coli* colonisation (Pi et al., 2012). However, the role that other siderophores produced by *E. coli* have in the gut is not clear.

The comparison of the distribution of siderophore genes in the ECOR-F and GMB collections suggests that the environment of isolation influences which siderophore systems are present, and that the gut appears to select for both increased siderophore production and the ability to produce a wider variety of siderophores. It was also observed that the B2 phylogenetic group, in particular, encoded a higher number of siderophore systems. This group has been widely associated with host commensalism, especially in humans where it is frequently the most abundant group, as well as the dominant group (Zhang et al., 2002; Escobar-Páramo, Grenet et al., 2004; Nowrouzian et al., 2005; Gordon et al., 2015). Our competition studies in mice (Chapter 3, section 3.5) also demonstrated that the B2 phylogenetic group can become dominant within the gut environment. We therefore hypothesised that isolates from the gut had higher siderophore production because the ability to produce siderophores confers a fitness advantage in the healthy gut environment. In contrast, the non-host environment seems to elicit a weaker selection pressure on *E. coli* to maintain a diverse set of siderophore systems.

These differences may reflect environmental differences in iron availability. Ferrichrome and ferric citrate have both been found in the rhizosphere and phyllosphere (Reid et al., 1984; De Vos et al., 1986; Crowley, 2006) and could act as alternative iron sources. In the GI-tract, bacteria acquire iron based on the host's diet, so the form of iron available will vary. In humans, most ingested iron (up to 90%) is not absorbed and is available for the microbiota (Hurrell and Egli, 2010). Ingested iron can be broadly divided into haem and non-haem iron, with haem being derived from meat and animal products and non-haem mainly from plants (Monsen et al., 1978). The form that the iron acquires when it is in the colon is unclear as it will largely depend on digestion and the composition of the meal the host has consumed. It is, however, likely that iron is available in a variety of forms.

Having multiple siderophores may provide a benefit to bacteria due to the fact that different siderophores have different characteristics and properties based on pH and carbon source (Valdebenito et al., 2006). In the heterogeneous environment such as the gut, where pH and carbon source availability varies considerably, having multiple siderophores may confer an advantage. In the *E. coli* Nissle 1917 strain, for example, it was shown that the binding potential of aerobactin for iron is higher than that of enterobactin when the pH is low, potentially leading to aerobactin being most useful under acidic conditions (Valdebenito et al., 2006). Another potential advantage for producing multiple siderophores is that it might allow bacteria to be better competitors for iron compared to other strains. Co-culture experiments have shown that under low iron conditions when strains are under direct competition, strains that can produce siderophores with higher affinity sequester iron away from competitors, thus inhibiting their growth (Weaver and Kolter, 2004; Joshi et al., 2006).

The liquid CAS assays also highlighted that siderophore production, except for aerobactin, is decreased when *E. coli* is grown in the presence of glucose compared to glycerol. There are two possible explanations for this difference. Glycerol and glucose are metabolised differently by *E. coli*, with glycerol requiring aerobic respiration, whereas glucose can be metabolised through fermentative pathways. The enzymes that are required for aerobic respiration include some that contain both Fe-S clusters and haem groups (Py and Barras, 2010), which would require more iron compared to the enzymes involved in fermentation. This could result in increased iron uptake by *E. coli* grown on glycerol. Alternatively, siderophore production could be under carbon catabolite repression (CCR). *E. coli* strain BW25113 *Δcrp* mutants, which cannot make cyclic AMP receptor protein (CRP), a regulator which is involved in carbon source utilisation and CCR, were shown to have decreased expression of enterobactin (Zhang et al., 2005), suggesting that siderophore production is under CCR. However, the exact mechanism through which CRP regulates siderophore production is not clear. CRP has been shown to regulate the

expression of *fur*, but only by a relatively small amount (Zhang et al., 2005). There is, however, a possibility that CRP is able to directly regulate siderophore genes, and a putative CRP binding domain has been found upstream of the *entC* operon (Zhang et al., 2005), but no evidence has been recorded of direct binding and regulation by CRP.

From the multiple alignments as well as the multiplex PCR results, it was clear that siderophore loci are evolutionarily maintained as a complete unit. This has implications for the possibility of cheating in natural populations of *E. coli*. The potential for cheating to arise is linked to the fact that siderophores are "social goods" that are secreted by bacteria and can be utilised by neighbouring bacteria with the required receptor, and that siderophore production incurs a metabolic cost (Lv and Henderson, 2011). Cheating for siderophores has been shown to occur for a wide range of bacterial populations, from *P. aeruginosa* in the cystic fibrosis lung (De Vos et al., 2001; Buckling et al., 2007) to marine bacteria (D'Onofrio et al., 2010; Cordero et al., 2012). As all of the strains from the GMB and ECOR collections contained both biosynthesis and receptor genes, this suggests that cheating is uncommon and selected against in natural populations of *E. coli*.

One possible explanation is that the mucus layer where *E. coli* is thought to reside in the gut is highly viscous, which would limit the ability of secreted siderophores to diffuse to neighbouring cells or microcolonies (Kümmerli et al., 2009). This is supported by the observation that a hypersecretor mutant (receptor knockout mutant) is iron starved in the mouse GI-tract (Pi et al., 2012). Siderophores secreted by this mutant would not diffuse away from the bacteria, but would bind and trap any surrounding iron, making it unavailable to any other iron uptake systems. However, *P. aeruginosa* siderophore (pyoverdine) mutants in cystic fibrosis patients are found in the thick mucus that fills the lungs. This would suggest that siderophores would be able to diffuse through the mucus layer covering the gut epithelium. Recent studies have, however, suggested that *P. aeruginosa* strains in the cystic fibrosis lung are adapting towards alternative iron uptake mechanisms, such as haem uptake

(Ross-Gillespie et al., 2015) and the use of an alternative siderophore pyochelin (Nguyen et al., 2014; Ross-Gillespie et al., 2015), rather than utilising pyoverdine produced by neighbouring bacteria. Culture experiments using cheaters have also highlighted that cheating only tends to arise in populations where there is a very strong selection for siderophore production (Kümmerli et al., 2010). In populations where siderophores can be recycled (Lehmann, 2007; Kümmerli et al., 2010) or where low levels of siderophores are required, the cost associated with production is not high enough to outweigh the benefit of cheating. It is likely that iron availability in different environments in varied, which would mean that retaining siderophore producing genes is important. As the receptor and biosynthesis genes are under different promoters for most siderophore systems, the only exception being enterobactin (Garenaux et al., 2011), it is a possibility that under certain conditions *E. coli* may be able to "cheat" by switching off biosynthesis genes while ensuring the receptor genes are still expressed.

One possible limitation to this study is that the two collections being compared were isolated from different geographic locations at different times. The GMB collection was primarily collected from the UK, whereas the ECOR collection was from a mixture of European and American samples. However, it has been shown that approximately 2% of genetic diversity between ECOR strains from Europe and America was attributable to geographic location of isolation (Miller and Hartl, 1986). To account for the over 30 year difference in the ECOR and GMB collections, more recent studies (Hilali et al., 2000; Nowrouzian et al., 2003; Johnson et al., 2005; Johnson et al., 2008; Lee et al., 2010; Unno et al., 2011; Vollmerhausen et al., 2011; White et al., 2011; Kudinha et al., 2012; Mao et al., 2012) that identified siderophore gene presence in *E. coli* isolated from healthy faecal samples were compared to the results from the ECOR and GMB isolates. The prevalence of each siderophore system was not significantly different between these faecal samples and the ECOR-F group, supporting our observations that siderophores are more prevalent in faecal *E. coli* strains compared to plant isolates (table 4.1). It would be interesting to investigate if

the observed differences seen here between the GMB and ECOR collection could be extended to *E. coli* strains isolated from other plants or secondary environments.

In summary, the results of the multiplex PCRs suggest that yersiniabactin and aerobactin production is associated with gut colonisation by *E. coli*. CAS assays on strains that only possess enterobactin highlight the possibility that strain-specific differences in siderophore production may therefore impact on the ability to colonise the GI-tract. In the next chapter, we will investigate these differences in siderophore production and whether siderophore production affects fitness in the gut.

5. *In vitro* and *in vivo* siderophore gene expression, characterisation and production

The *in vitro* siderophore biosynthesis gene expression in MM9 medium results in this chapter form part of the paper 'Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *E. coli*' by Laura J. Searle, Guillaume Méric, Ida Porcelli, Samuel K. Sheppard and Sacha Lucchini published in the journal *PLoS One* (see appendix D). I designed and performed these experiments as well as analysing the data.

5.1 Introduction

In the previous chapter, siderophore production and gene distribution was shown to be increased at the population level in host-associated *E. coli* compared to plant-associated isolates, leading to the hypothesis that siderophore production confers an advantage in the GI-tract. In particular, higher production could be linked to the presence of the aerobactin and salmochelin loci. However, there was a significant amount of variation that was not explained by differences in gene distribution. Indeed, strains that only possessed the enterobactin gene showed a wide range of siderophore production levels, suggesting that there may be differences in the regulation of siderophore production between individual *E. coli* strains.

All four siderophore systems in *E. coli* have been shown to be regulated by the global regulator Fur (ferric uptake regulator). Fur acts with its co-factor Fe^{2+} to repress the expression of siderophore genes, along with other genes involved in ferric iron uptake. When intracellular concentrations of Fe^{2+} are low, Fur is unable to bind to DNA and expression of siderophore genes increases (Hantke, 2001). The expression of siderophore genes or Fur itself may be modified by additional regulators, such as CRP and OxyRS, possibly resulting in unique expression profiles for each siderophore depending on environmental conditions. The heterogeneous environment in the GI-tract, therefore, may select for multiple siderophore systems, with factors such as pH, carbon source

availability and host secreted molecule lipocalin-2 likely to influence siderophore production.

In this chapter, we build on the observed differences in siderophore production between individual *E. coli* isolates from the previous chapter and assess the role of each individual siderophore system in iron uptake *in vitro* and the interactions between carbon source utilisation and iron homeostasis using siderophore biosynthesis and uptake mutants. Finally, following the observation that siderophore production was increased in host-associated strains, and to build on work that showed enterobactin to be important during GI-tract colonisation (Pi et al., 2012), siderophore gene expression and siderophore mutants were assessed in the mouse intestine to elucidate a possible role for additional siderophores in commensal *E. coli* within the host environment.

5.2 Strain-specific diversity in siderophore gene expression under low iron conditions

To evaluate changes in gene expression and regulation as a possible explanation for siderophore production variation observed during liquid and agar CAS assays (Chapter 4, section 4.2), the expression of one biosynthesis gene for each siderophore system was determined in eight GMB isolates under iron limiting conditions in MM9 medium. The isolates analysed were GMB23, GMB30, GMB40, GMB53, GMB88, GMB91, GMB100 and GMB104 (see table A.1 for strain information), and were selected to represent a wide range of siderophore production levels.

Analysis of RT-PCR results showed that enterobactin and aerobactin were the most highly expressed siderophore systems, displaying a 4 and 400-fold greater expression level compared to salmochelin and yersiniabactin respectively (figure 5.1A). Aerobactin and yersiniabactin both showed a narrow range of expression for all eight strains, always being expressed at high or very low levels respectively. This supports the association of the aerobactin locus with high siderophore production (Chapter 4, section 4.4.3). The consistent low expression levels of the yersiniabactin biosynthesis gene may also explain the

absence of a correlation between yersiniabactin locus presence and siderophore production. Importantly, enterobactin and salmochelin displayed a wide range of expression levels between the eight strains (figure 5.1A). These expression levels showed a strong correlation to siderophore production measured on CAS agar plates (figure 5.1B), suggesting strain-specific diversity in the regulation of these two siderophore systems.

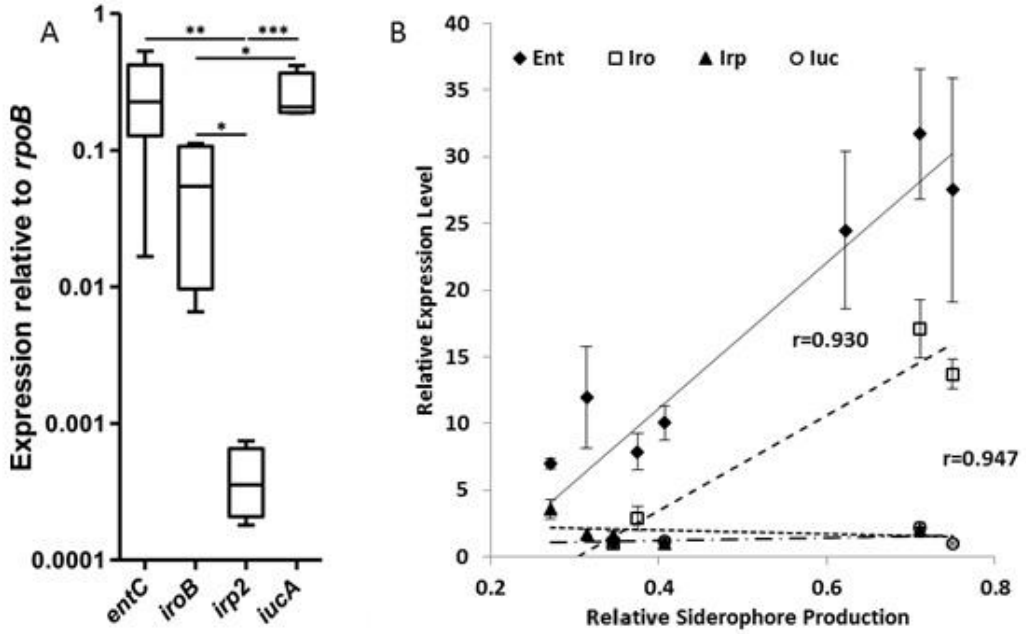


Figure 5.1: Gene expression level of siderophore biosynthesis genes in plant-associated *E. coli*. The expression of one biosynthesis gene for each siderophore system present in eight GMB isolates was determined; *entC* (enterobactin), *iroB* (salmochelin), *irp2* (yersiniabactin), *iucA* (aerobactin). The strains analysed were GMB23 (*entC*, *irp2*, *iucA*), GMB30 (*entC*, *iroB*), GMB40 (*entC*, *iroB*, *irp2*), GMB53 (*entC*, *irp2*), GMB88 (*entC*, *iroB*, *iucA*), GMB91 (*entC*), GMB100 (*entC*, *irp2*) and GMB104 (*entC*, *iroB*, *irp2*, *iucA*). A) Box plot showing the gene expression levels of *entC*, *iroB*, *irp2* and *iucA* relative to the internal reference rpoB. The centre rectangle of the plot spans the interquartile range (IQR). The segment inside the rectangle shows the median, while the bars above and below show the maximum and minimum values respectively. Statistical significance was determined using the Student's *t*-test. In case of multiple tests, the significance of individual t-tests was determined using the Benjamini and Hochberg False discovery method; *p<0.05, **p<0.01, ***p<0.001. B) To visualise the link between the gene expression level of each siderophore and total production, mRNA levels were normalised to the lowest corresponding value and plotted against relative siderophore production on CAS plates. Relative siderophore production was calculated by dividing the halo diameter by the colony diameter (d_{halo}/d_{colony}) and normalising to the highest producer.

5.3 Siderophore gene expression during severe iron depletion

As each siderophore system displayed varying levels of expression in MM9 medium, the expression of all four siderophores was measured in nutrient broth with varying amounts of iron chelator 2,2'-dipyridyl (DIP). In particular, we were interested in whether yersiniabactin would be induced by low iron levels, as its expression was significantly lower than that of the other three siderophore systems in MM9 medium (figure 5.1). Compared to expression in MM9 medium, expression in nutrient broth (NB) was 24 and 4-fold lower for enterobactin and salmochelin respectively (figure 5.2A), indicating iron availability was higher in NB medium. However, expression of yersiniabactin and aerobactin was not significantly increased in MM9 medium (figure 5.2A).

There was, however, a significant induction of biosynthesis genes for all four siderophore systems on addition of DIP to NB medium (figure 5.2B). In particular, enterobactin and salmochelin expression was greatly increased in 200µM DIP, both showing a 290-fold increase compared to expression in nutrient broth (figure 5.2B). This increase was lower at 500µM DIP, however, possibly reflecting the reduced growth observed at this concentration (data not shown). It has been previously observed that bacteria that are grown with severe iron limitation stop growing and producing siderophores (Merrell et al., 2003; Valdebenito et al., 2006). Compared to enterobactin and salmochelin, aerobactin and yersiniabactin showed a lower level of induction in the presence of 200µM DIP, with a 28-fold and 18-fold induction respectively. Their expression, however, was slightly increased at 500µM DIP, to 49-fold and 28-fold respectively. Although aerobactin displayed a reduced level of induction compared to enterobactin and salmochelin in NB with 200µM DIP, these three siderophores were expressed at the same level relative to the housekeeping gene *rpoB* (figure 5.2A). In the case of aerobactin, the lower level of induction might be linked to its already high expression in nutrient broth, which is 10-fold higher than enterobactin. The fact that aerobactin is also expressed in MM9 and NB media at a similar level, suggests that the aerobactin locus is not repressed

to the same extent as enterobactin and salmochelin in the presence of iron. Yersiniabactin expression was increased in the presence of DIP, however, it was still expressed at lower levels than the three other siderophore systems in NB (figure 5.2A). This suggests that yersiniabactin may require factors other than low iron to induce high expression levels.

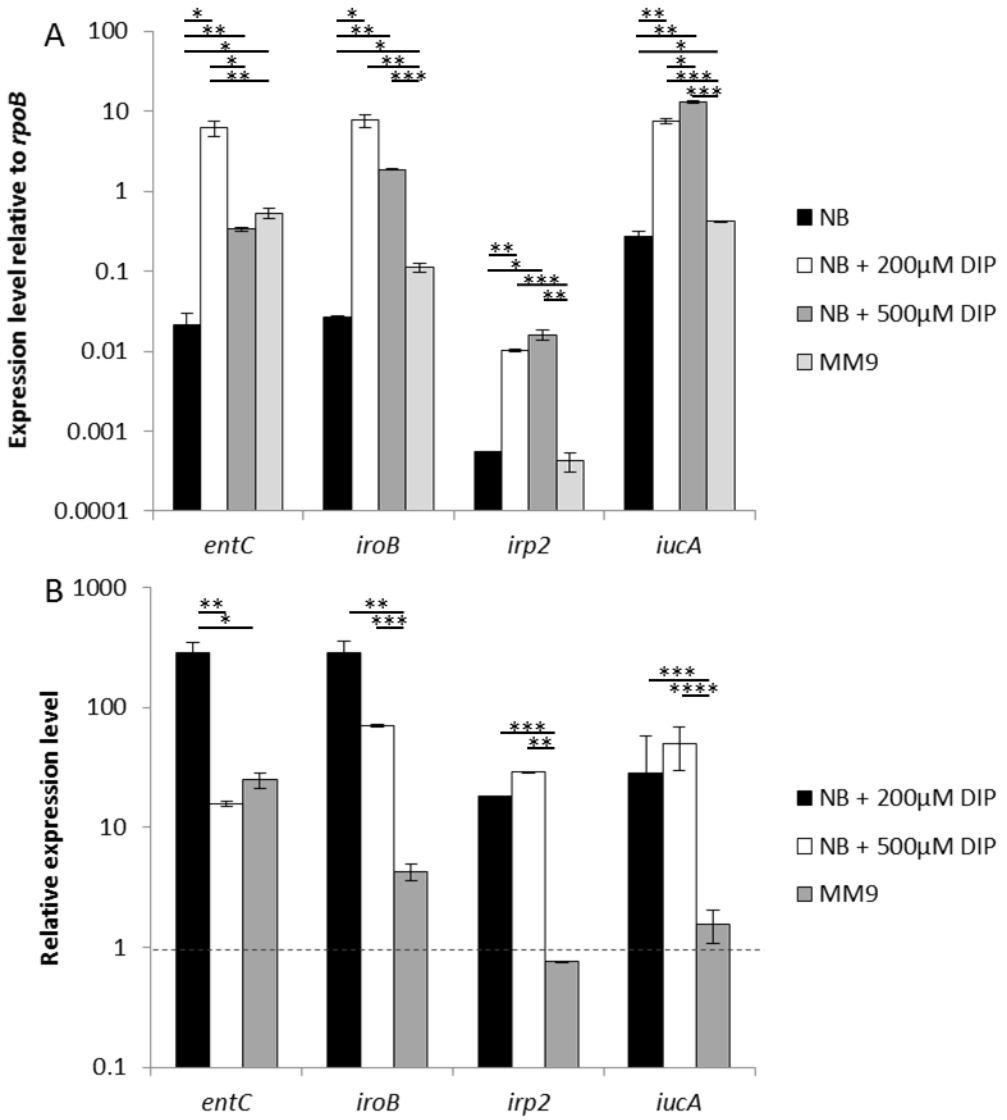


Figure 5.2: Gene expression of siderophore biosynthesis genes in isolate GMB104 under iron replete and limiting conditions. Expression levels of *entC* (enterobactin), *iroB* (salmochelin), *irp2* (yersiniabactin) and *iucA* (aerobactin) for strain GMB104 in MM9 medium, nutrient broth (NB), NB with 200µM DIP (2,2'-dipyridyl) and NB with 500µM DIP. A) Expression relative to the housekeeping gene rpoB. B) Expression relative to expression in NB (iron replete conditions). The dotted line signifies NB expression levels for all four siderophore systems. Results show the mean ± standard error (*n* = 3). Statistical significance was determined using the Student's *t*-test. In case of multiple tests, the significance of individual t-tests was determined using the Benjamini and Hochberg False discovery method; *p<0.05, **p<0.01, ***p<0.001.

5.4 Characterisation of siderophore biosynthesis mutants

5.4.1 Biosynthesis mutant growth curves

To assess the contribution of each siderophore system to iron uptake during iron limited growth, siderophore biosynthesis and *ΔtonB* mutants were generated (see table 2.3 for a list of mutants). Siderophore mutants were generated in a strain that encodes all the four siderophores (GMB104) and a strain only possessing the ability to generate enterobactin (GMB91). Growth curves were established in MM9 medium, and MM9 medium supplemented with either 100μM DIP or ferrous sulphate ($FeSO_4$). Only *ΔtonB* mutants showed reduced growth in MM9 medium compared to the wild-type (figure 5.3). Growth was restored by addition of 100μM $FeSO_4$, suggesting that the decrease in growth was as a result of a reduced ability to take up ferric iron (figure 5.3C and D).

Supplementation of MM9 with 100μM DIP resulted in slight reductions in growth for GMB104es, GMB104a, GMB104esy and GMB104asy, with larger reductions in growth rate observed for GMB91e, GMB104eas and GMB104easy (figure 5.3). This indicates that aerobactin or enterobactin production is required to maintain growth equivalent to the wild-type under iron limitation, supporting the observation that enterobactin and aerobactin are expressed at a higher level compared to salmochelin and yersiniabactin in MM9 medium. The GMB104eas mutant, which can only secrete yersiniabactin, grew poorly compared to the other triple knockout mutants GMB104esy and GMB104asy. This is consistent with the expression results for yersiniabactin, which was expressed 600-fold lower than enterobactin and aerobactin. The salmochelin deletion mutant appeared to have no change in fitness compared to the wild-type. This is most likely due to the fact that salmochelin is synthesised by modification of enterobactin. Therefore, if salmochelin is not produced, the bacteria can still obtain iron through unmodified enterobactin *in vitro*.

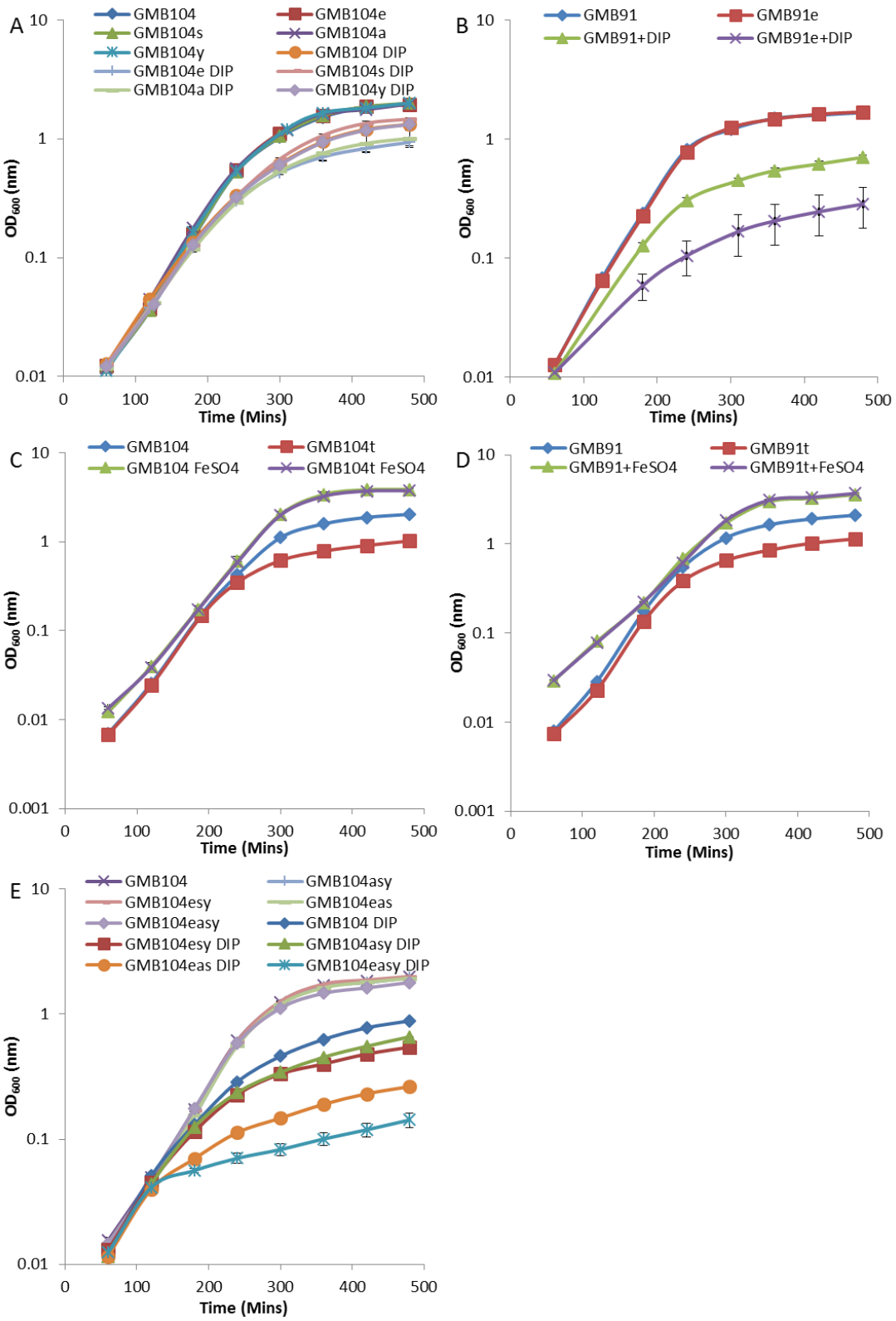


Figure 5.3: Growth curves of strains GMB104 and GMB91 and their siderophore biosynthesis and tonB mutants in MM9 medium and MM9 medium supplemented with 100µM 2,2'-dipyridyl (DIP) or FeSO4 at 37°C and pH 7.0. A) GMB104 single deletion mutants GMB104e, GMB104a, GMB104s and GMB104y B) GMB91e mutant C) GMB104t mutant D) GMB91t mutant E) GMB104 triple and quadruple deletion mutants GMB104asy, GMB104esy, GMB104eas and GMB104easy. Results show the mean ± standard error (*n* = 3). See table 2.3 for full details of mutants.

5.4.2 Csàky and Arnow assays

To assess production of individual siderophores by biosynthesis and *ΔtonB* mutants, Csàky and Arnow assays were performed that can measure hydroxamate (aerobactin) and catecholate-type (enterobactin and salmochelin) siderophores respectively. Production of aerobactin was not detected in GMB104 mutants lacking the *iucABCD* locus (GMB104a and GMB104easy) (figure 5.4A). The Arnow assay showed that production of enterobactin and/or salmochelin was reduced in GMB91e, GMB104es, GMB104eas and GMB104easy (figure 5.4B), showing gene deletions successfully removed siderophore synthesis activity. The salmochelin deletion mutant GMB104s did not show a reduction in the amount of catecholate-type siderophores, due to enterobactin production. Interestingly, enterobactin production appeared to be increased in the GMB104asy mutant, which may be as a result of compensation for the loss of aerobactin synthesis. However, the GMB91 and GMB104 *ΔtonB* mutants did not display hyperproduction of enterobactin and/or salmochelin compared to the wild-type, so it is unclear whether reduced siderophore uptake does result in increased enterobactin production. The GMB104 *ΔtonB* mutant did, however, display increased aerobactin production compared to the wild-type.

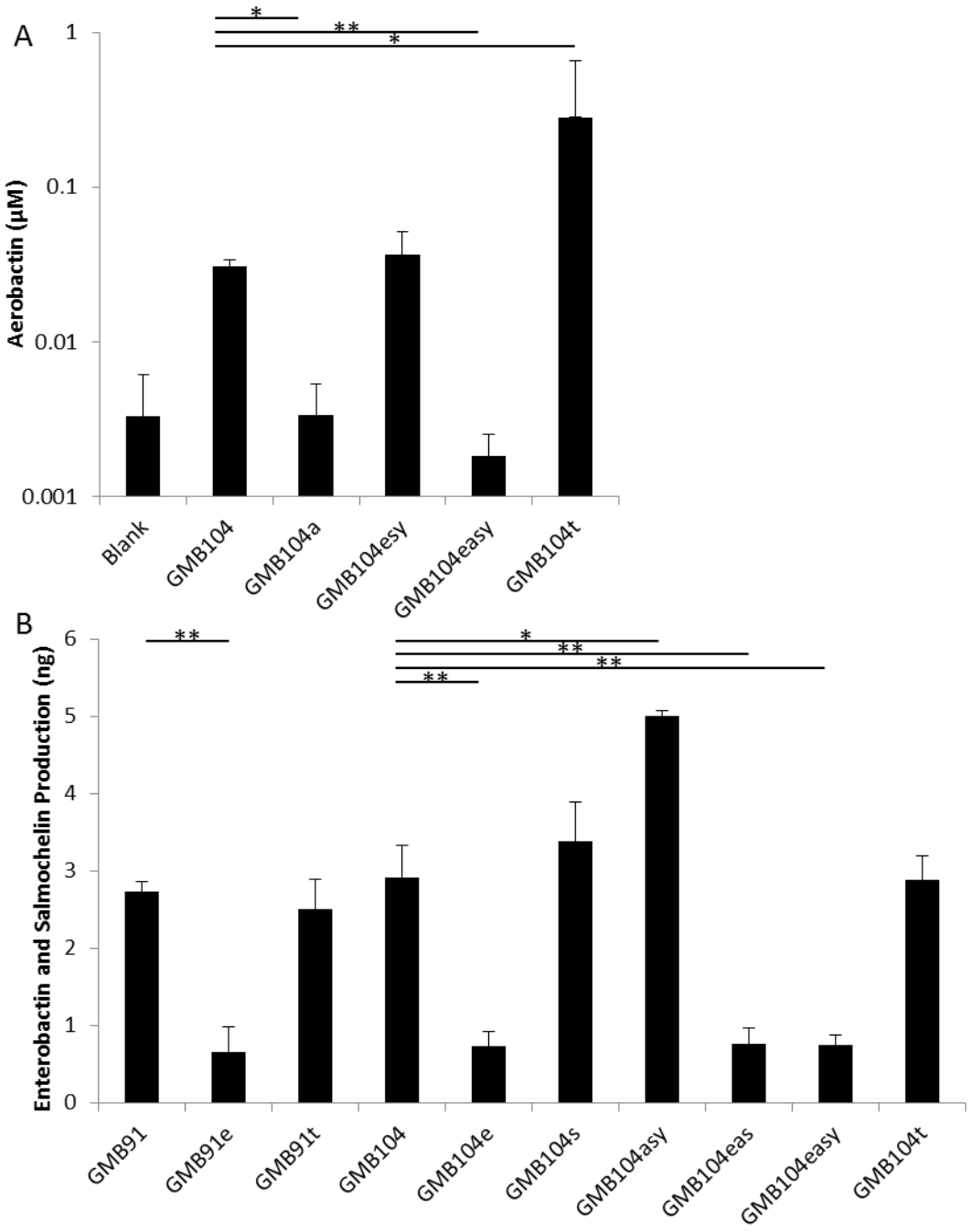


Figure 5.4: Csàky and Arnow assays to measure production of hydroxamate and catecholate siderophores respectively. A) Hydroxamate (aerobactin) and B) catecholate (enterobactin and salmochelin) production was measured in MM9 medium at pH 7.0 with glucose at 37°C. No production was detected in mutants lacking the required biosynthesis genes. Results show the mean ± standard error (*n* = 3). Statistical significance was determined using the Student's *t*-test; *p<0.05, **p<0.01

5.4.3 Liquid CAS assay

Total siderophore production by mutant and wild-type strains was assessed using the liquid CAS assay with glucose or glycerol as the sole carbon source.

GMB104 mutants that produced aerobactin (GMB104es, GMB104s, GMB104y and GMB104esy) had similar total siderophore production levels to the wild-type when grown on both glycerol and glucose (figure 5.5). This supports the link between high siderophore production and the presence of the aerobactin locus (Chapter 4, section 4.4.3). GMB104a and GMB104asy, which are unable to produce aerobactin, but are able to synthesise enterobactin, showed reduced siderophore production in the presence of glucose but not glycerol. This reduction of siderophore production in the presence of glucose is also observable in the GMB91 wild-type, which only possesses enterobactin. These results support observations that siderophore production is only maintained in *E. coli* isolates that possess the aerobactin locus when grown on glucose (figure 4.7). This suggests that under the growth conditions used in this study, only enterobactin, and, by extension, salmochelin, are under carbon catabolite repression by glucose. Unfortunately, as salmochelin is produced from enterobactin, it is not possible to construct a mutant that is only able to synthesise salmochelin, so the direct effect of carbon source on salmochelin production cannot be determined.

Three mutants, GMB104eas, GMB104easy and GMB91e, all showed very low siderophore production levels in the presence of glucose or glycerol (figure 5.5). GMB91e and GMB104easy mutants should both be unable to produce any siderophores. However, GMB104eas is still capable of synthesising yersiniabactin, suggesting that environmental factors other than iron starvation are needed to induce yersiniabactin expression in this strain of *E. coli*, supporting the observation of low yersiniabactin expression in MM9 medium. The *ΔtonB* mutants for both GMB91 and GMB104 strains, which can produce but not internalise siderophores, showed hypersecretion of siderophores, with increases in siderophore production of approximately 100% for both GMB91t and GMB104t in glucose, and increases of 53.5% and 400.2% in glycerol for GMB104t and GMB91t respectively (figure 5.5). This is slightly in contrast to the Arnow assay (figure 5.4), which showed no increase in catecholate siderophore production for GMB91t.

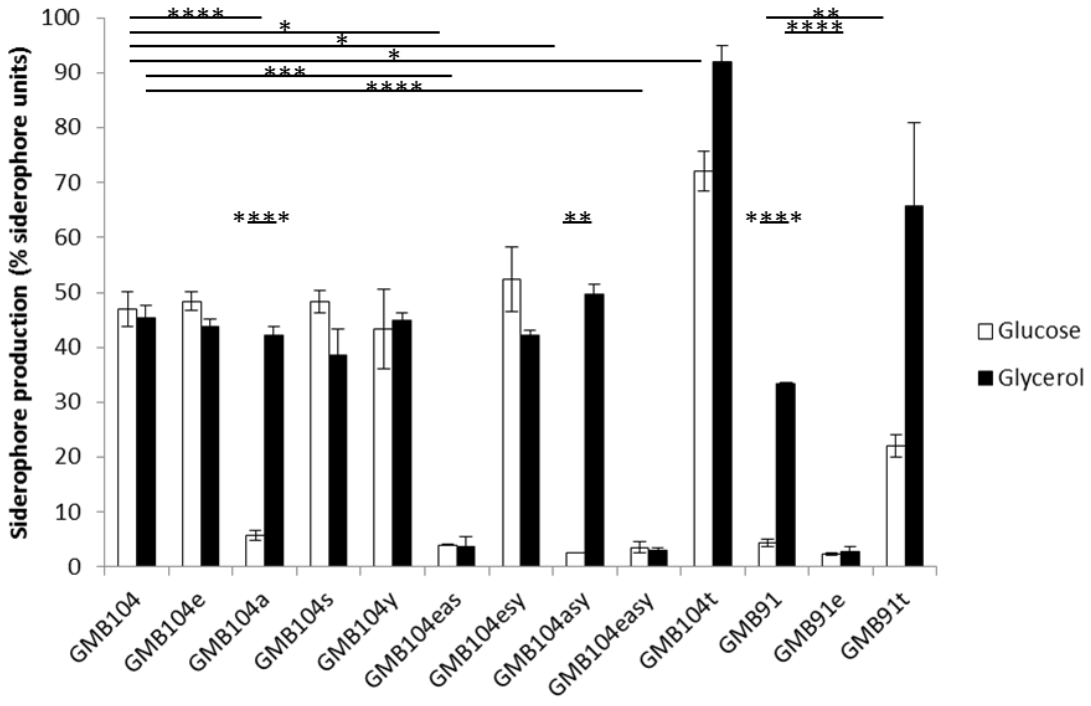


Figure 5.5: Total siderophore production by siderophore biosynthesis and *ΔtonB* mutants of *E. coli* isolates GMB104 and GMB91. Production was measured in MM9 medium with either glucose or glycerol as the sole carbon source at pH 7.0 and 37°C. Bar graph shows percent siderophore units, calculated using $[(A_{reference} - A_{sample})/A_{reference}]$, where the reference is MM9 medium mixed with CAS assay solution. Results show the mean ± standard error (*n* = 3). Statistical significance was determined using the Student's *t*-test; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

5.4.4 Carbon source utilisation by siderophore biosynthesis mutants

GN2 BIOLOG plates were used to identify the carbon utilisation ability for each siderophore mutant at 37°C. Carbon utilisation profiles were used to calculate correlations between mutants and carbon sources to construct PCA (principal component analysis) plots. We analysed both the variation in carbon source utilisation between mutants and the wild-type, as well as differences between each carbon source. Plots show the spread of the data based on the two principal components (variables) that explain most of this variation. Investigating the variation between mutants, there was a clear separation of the two genetic backgrounds based on carbon source utilisation, as only GMB91 was able to metabolise D-galactonic acid lactone (figure 5.6A). There was also a clear separation of mutants that had a significantly reduced ability to take up

iron (GMB91e, GMB91t, GMB104eas, GMB104easy and GMB104t) from their respective wild-type. This could be as a result of mutants being unable to utilise certain carbon sources due to low intracellular iron levels.

The analysis of the utilisation profile of individual carbon sources identified a group that are associated with lower growth in *ΔtonB* and GMB91e, GMB104eas and GMB104eas mutants (figure 5.6B). Both succinic acid and bromosuccinic acid appear to be unable to support growth of these five mutants, supporting observations of *Δfur* and *ryhB* hyperexpression mutants having poor growth on succinate (Hantke, 1987; Wassarman et al., 2001) due to inhibition of the TCA cycle (Massé and Gottesman, 2002; Seo et al., 2014). These five mutants also showed no growth after 48h on propionic acid, which is metabolised through the activity of the *prpECDB* operon, which has recently been shown to have altered expression in an *E. coli* K-12 strain MG1655 *Δfur* mutant (Seo et al., 2014). The amino acids D-alanine and L-proline were also unable to support the growth of several mutants (GMB91t, GMB104easy and GMB104t). The *dadX* and *dadA* genes involved in D-alanine metabolism and the *putA* gene involved in L-proline degradation have been shown to have altered expression in a *Δfur* mutant (Seo et al., 2014). Proline metabolism also yields fumarate, which is metabolised through the TCA cycle. Two carbon sources, p-hydroxylacetic acid and D-saccharic acid, were unable to support growth of GMB104 strain mutants (GMB104eas, GMB104easy and GMB104t) after 48h, but GMB91 mutants could grow, albeit to a reduced degree compared to the wild-type. This suggests possible differences in the metabolism of these two carbon sources between strains. Alternatively, GMB104 mutants may be more iron deprived due to the presence or absence of alternative iron uptake systems in GMB91 and GMB104.

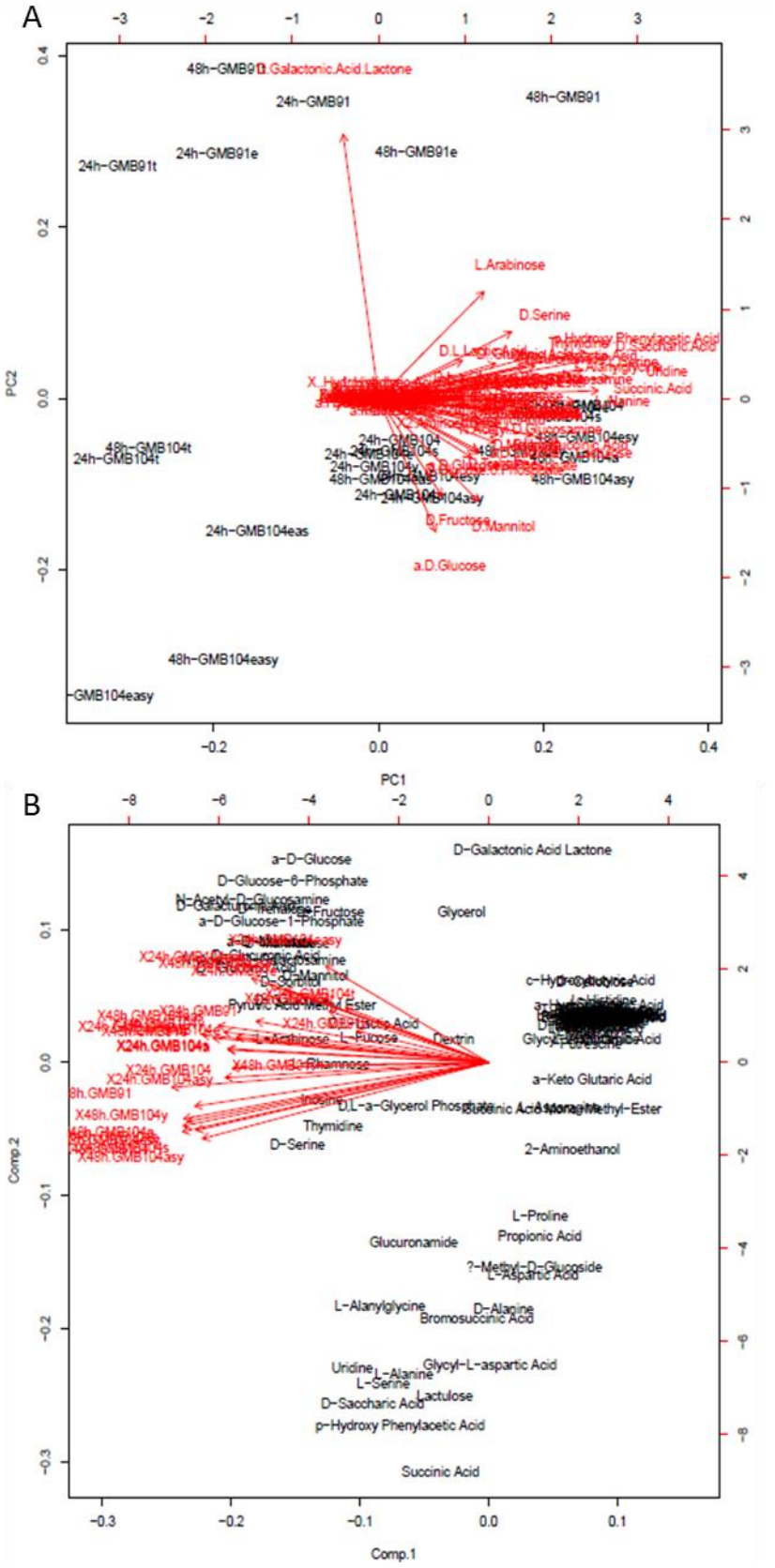


Figure 5.6: Principal component analysis (PCA) diagrams from carbon source utilisation profiles on GN2 BIOLOG plates of GMB91 and GMB104 siderophore biosynthesis and *ΔtonB* mutants. A) PCA biplot displaying the variance of the carbon utilisation among the strains tested (*n* = 13). B) PCA biplot displaying the variance of utilisation among carbon sources (*n* = 96). The red arrows indicate variable PCA scores in the direction of increasing value for each variable.

Many of the mutants displayed reduced growth after 48h on several carbon sources. One possible reason for this reduction in growth could be a switch from oxidative phosphorylation (TCA cycle) to fermentative pathways as proposed in other studies (O'Brien et al., 2013; Seo et al., 2014), which could result in slower growth, and a trend for reduced growth in severely iron deprived mutants for most carbon sources. However, there are a few carbon sources, D-glucose, L-arabinose, D-fructose and α-D-glucose-1-phosphate, which supported higher growth in siderophore knockout mutants, but not in the *ΔtonB* mutants. This suggests that siderophore production has a metabolic cost during growth on these sugars.

5.5 Siderophore expression *in vivo*

To assess a possible role for additional siderophore systems to enterobactin in the GI-tract, the expression of one biosynthesis gene for all four siderophores was measured in the colon contents of mice orally gavaged with strain GMB104. Analysis of semi-quantitative results showed that siderophore biosynthesis gene expression was lower in the colon compared to *in vitro* (figure 5.7B). However, expression of enterobactin, salmochelin and aerobactin was detected within the colon (figure 5.7A). Expression of salmochelin was 2.4 and 2.1-fold higher than enterobactin and aerobactin respectively. Yersiniabactin gene expression was not detectable, suggesting it does not play a role in the colon.

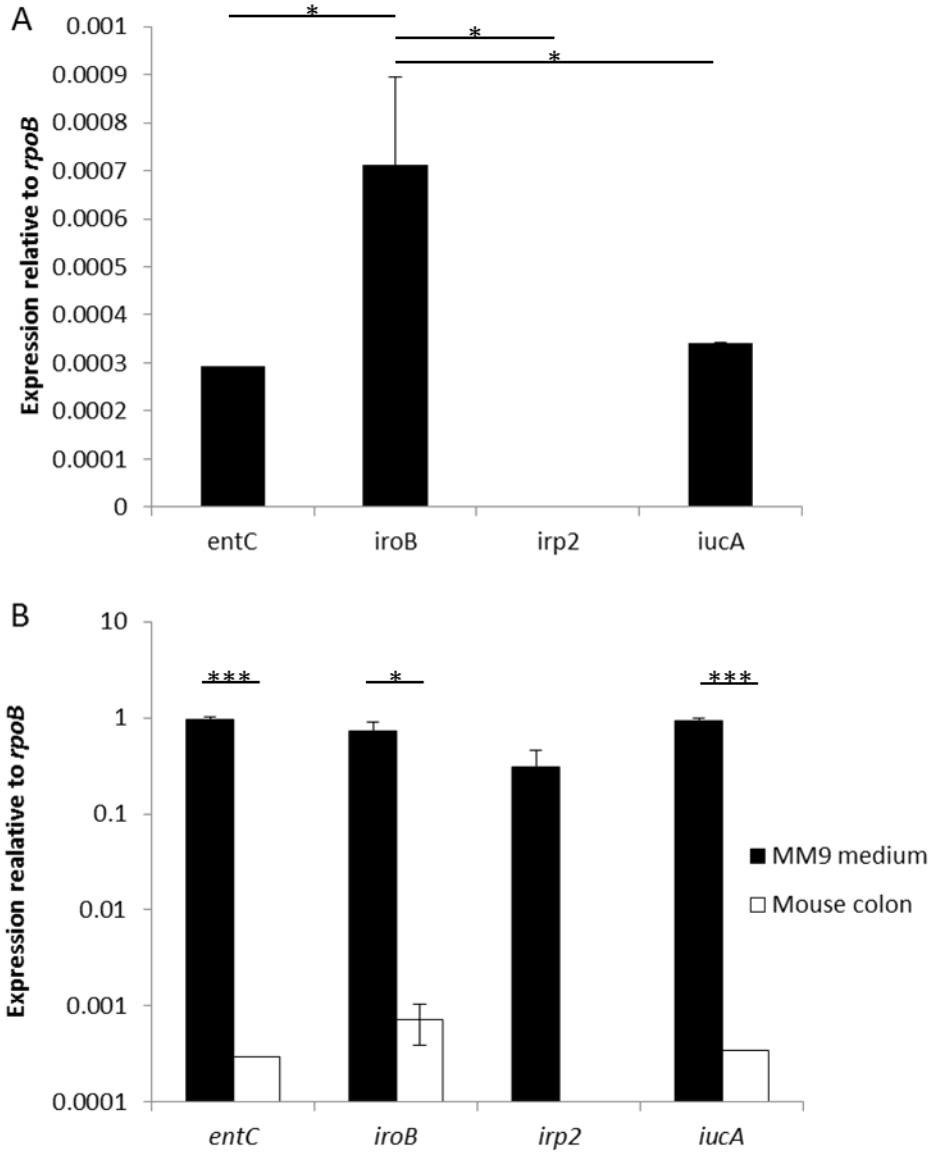


Figure 5.7: Gene expression levels of siderophore biosynthesis genes in the mouse colon. The expression of one biosynthesis gene for each siderophore system present in GMB104 was determined; *entC* (enterobactin), *iroB* (salmochelin), *irp2* (salmochelin), *iucA* (aerobactin). A) Expression of siderophore biosynthesis genes relative to *rpoB*. B) Expression in the mouse colon compared to expression in MM9 medium. Results show the mean ± standard error (n_{mice} = 4, n_{MM9} = 3). Statistical significance was determined using the Student's *t*-test. In case of multiple tests, the significance of individual t-tests was determined using the Benjamini and Hochberg False discovery method; *p<0.05, **p<0.01, ***p<0.001.

5.6 Siderophore mutant competitive index assays *in vivo*

To determine whether siderophore production or uptake conferred a fitness advantage within the GI-tract, competitive index assays were performed in mice using 1:1 mixtures of the wild-type and a mutant (GMB104t, GMB91t or

GMB91e). CFU counts detected in faecal samples were lower by approximately 2.5 logs by the end of the experiment (28 days) for both GMB104t and GMB91t compared to wild-type strains (figure 5.8), suggesting that they had reduced fitness in the gut and ferric iron is important for GI-tract colonisation. GMB91e, in contrast, did not show any change in colonisation ability compared to the wild-type (figure 5.8). Together these results indicate that although ferric iron uptake is important in the GI-tract, siderophore production may not be required to obtain the necessary amount. However, it is important to highlight that the GMB91e mutant retains the enterobactin receptor, so it may be able to utilise siderophores secreted by the wild-type and mask any reduction in fitness.

By extension, the wild-type may also be able to cheat in the *ΔtonB* competition assays. *ΔtonB* mutants hypersecrete siderophores, as shown by the liquid CAS assay (figure 5.5), which would be available for the wild-type. It is interesting then, that the CFU levels for the wild-type are increased at the end of the experiment, at 2.28×10^7 and 5.96×10^8 CFU/g faeces for GMB104 and GMB91 respectively. This is quite high compared to the 9.3×10^5 CFU/g faeces measured for GMB91 in the GMB91e competitive index assay.

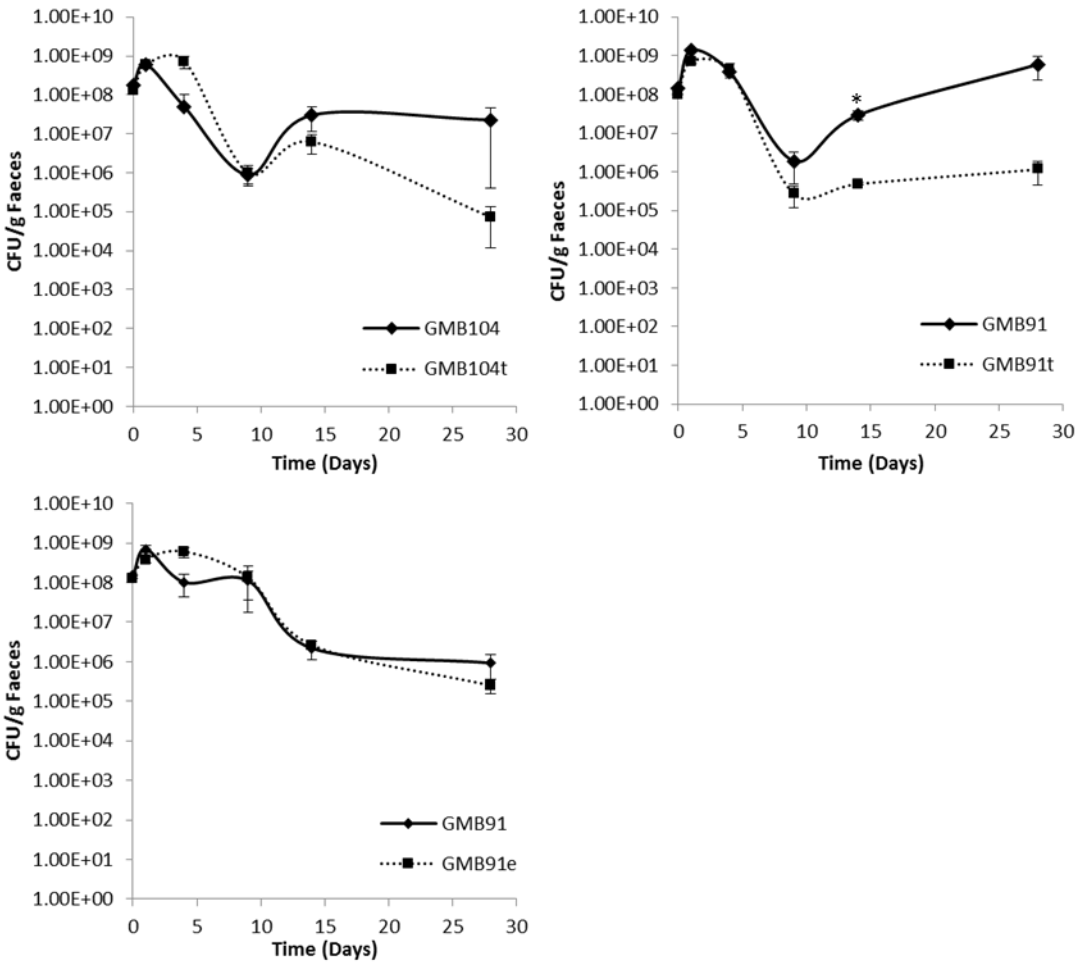


Figure 5.8: *ΔtonB* mutants show reduced colonisation of the GI-tract compared to wild-type strains. No significant difference in colonisation was seen for GMB91 Δ*entC* mutant. In each group, mice were orally gavaged with a 1:1 mixture of a mutant strain and the wild-type parent strain (day 0). If the gene of interest confers a competitive advantage in the GI-tract, the wild-type is expected to outcompete the mutant and become dominant. If the gene has no competitive advantage, the mutant and wild-type should colonise the gut at similar levels. Mice were monitored for 4 weeks. Results show mean faecal CFU/g ± standard error (*n* = 5). Statistical significance was determined using the Student's *t*-test; *p<0.05.

5.7 Discussion

Observed differences in siderophore production between the ECOR-F and GMB collections were not fully explained by differences in siderophore system presence, as discussed in Chapter 4. Investigating the expression of siderophore biosynthesis genes confirmed that the four siderophore systems had differing expression, which could indicate variations in gene regulation. Aerobactin and enterobactin were the most highly expressed. This was supported by siderophore mutants that were unable to produce both of these siderophores

having the greatest growth reduction under low iron conditions. The variation observed in siderophore production on CAS agar plates correlated with the variation in expression of enterobactin and salmochelin.

Interestingly, yersiniabactin was not highly expressed in any of the eight GMB strains studied. Induction of yersiniabactin was achieved in strain GMB104 on the addition of 2,2'-dipyridyl to NB medium, however, expression levels were still lower than those observed for all other siderophores when grown in NB with sufficient iron availability. It is possible that low levels of yersiniabactin are sufficient to mediate iron uptake, as each siderophore is able to bind and transport iron differently. Aerobactin, for example, has been shown to support growth of *E. coli* at lower concentrations than enterobactin (Williams and Carbonetti, 1986). However, the fact that the GMB104eas mutant displays a much greater reduction in growth under iron limitation compared to the other triple mutants GMB104esy and GMB104asy, suggests that insufficient levels of yersiniabactin are being produced. This is in contrast to siderophore production data for *E. coli* strain Nissle 1917 (Valdebenito et al., 2006) and *E. coli* ABU and UTI strains (Watts et al., 2012) which displayed high yersiniabactin production levels. However, it has been shown that faecal isolates have reduced production of yersiniabactin and salmochelin in comparison to UTI strains (Henderson et al., 2009). It is possible that the GMB strains used in this study do not have the correct genetic background, or are unable to express yersiniabactin at high levels under the conditions tested.

The yersiniabactin locus encodes an AraC-type regulator, *ybtA*, which has been shown to induce expression of yersiniabactin biosynthesis genes (Fetherston et al., 1996). However, the activity of YbtA is dependent on yersiniabactin-Fe^{3+} as a co-factor, resulting in yersiniabactin production positively regulating itself (Anisimov et al., 2005b). None of the other siderophore systems have been shown to have a similar feedback mechanism, so it is possible that this may be responsible for the different expression of yersiniabactin.

It has been proposed that there is a metabolic cost associated with siderophore production, which has been used as an explanation for the appearance of cheaters that are able to internalise siderophores produced by neighbouring cells (Griffin et al., 2004; Henderson et al., 2009; Lv et al., 2014). This metabolic cost may explain why the siderophore biosynthesis mutants displayed higher growth on some sugars in the BIOLOG assay. *E. coli* siderophore biosynthesis mutants have been shown to have different metabolic profiles compared to wild-type strains when grown in monoculture (Lv et al., 2014). It is possible that changes in iron uptake, as a result of removing siderophore activity, results in altered metabolism by changing iron availability within the cell. Alternatively, the increased availability of compounds used to make siderophores may alter metabolic pathways. Although a link between carbon source utilisation and iron uptake has been observed (Zhang et al., 2005), further work is needed to understand all the metabolic pathways influenced by siderophore production. It is possible that individual siderophores have a specific cost that influences nutrient uptake and growth. The fact that aerobactin was expressed at a consistently high level, even during growth in high iron conditions, could therefore be associated with an elevated metabolic cost. Therefore, there might be a weak selective pressure for the maintenance of this locus in *E. coli* strains that do not require aerobactin for iron acquisition. Interestingly, aerobactin prevalence is very low in environmental *E. coli* compared to faecal isolates, suggesting a more important role of aerobactin in the GI-tract (figure 4.4). In the pathogenic *E. coli* strain UTI89, salmochelin production in particular was linked to significant changes in the metabolome, possibly because salmochelin synthesis consumes UDP-glucose from the gluconeogenesis pathway (Lv et al., 2014). However, there were no clear differences in carbon utilisation profile on BIOLOG plates for the salmochelin mutants used in this study.

It is also important to take into consideration the ability of a cell to re-use siderophores. Iron release from enterobactin and salmochelin requires these siderophores to be hydrolysed due to their high affinity for Fe^{3+} (Langman et al., 1972; Ratledge and Dover, 2000). This produces dihydroxybenzoic acid (DHBA)

in the case of enterobactin, which can also act as a siderophore, although with a much reduced binding affinity for iron (Hantke, 1990). Glucosyl-DHBA produced from breakdown of salmochelin has been hypothesised as not being re-usable by *E. coli*, rendering salmochelin production more costly to a bacterium (Lv et al., 2014). Aerobactin is not hydrolysed during iron release, and can be recycled by cells (Braun et al., 1984). No hydrolases have been described for yersiniabactin (Garenaux et al., 2011), so it is possible that yersiniabactin can also be recycled.

Several carbon sources on the BIOLOG plates were influenced by intracellular iron availability. The mutants that displayed the biggest differences in carbon source utilisation compared to the wild-type were those most hampered in their ability to take up iron, and thus having the lowest intracellular iron levels. Intracellular iron levels may also modulate carbon source utilisation, possibly as a result of alterations in iron incorporation into enzymes involved in metabolism. *Δfur* mutants have been shown to be unable to grow on succinate (Hantke, 1987) and strains that overexpress RyhB grow poorly (Wassarman et al., 2001), most likely as a result of the succinate dehydrogenase (SdhCDAB) enzyme being inhibited by RyhB (Massé and Gottesman, 2002; Seo et al., 2014). Another enzyme in the TCA cycle, aconitase (AcnA), has also been shown to be regulated by Fur and RyhB, resulting in reduced expression during iron starvation (Seo et al., 2014). Regulation of the TCA cycle has been suggested to enable cells to switch from oxidative phosphorylation to fermentation pathways in response to reduced iron availability (O'Brien et al., 2013; Seo et al., 2014), most likely due to several of the enzymes involved in the TCA cycle requiring iron (Py and Barras, 2010).

Despite the availability of information about siderophore-mediated iron acquisition, it is not clear how having higher production and a more diverse repertoire of siderophores may benefit gut commensal *E. coli*. The fact that the four siderophore systems appear to be differentially regulated and are expressed and produced at different levels depending on environmental conditions, such as carbon source and pH (Valdebenito et al., 2006; Watts et al.,

2012), suggests that multiple siderophore systems may be beneficial in the temporally heterogeneous environment found in the GI-tract. Strains possessing several siderophore systems could also have an advantage in the densely populated gut environment when in competition with bacteria that utilise a narrower range of siderophores. Siderophore biosynthesis gene expression was detected *in vivo* for enterobactin, salmochelin and aerobactin, suggesting that multiple siderophores are utilised in the GI-tract. This expression was much lower than that observed *in vitro*, suggesting that *E. coli* is not severely iron-starved in the healthy gut environment, and that iron uptake via siderophores is limited.

No expression of yersiniabactin could be detected, suggesting that it is not used in the gut. This, along with the observed low *in vitro* expression, raises the question of when yersiniabactin is utilised by commensal *E. coli* strains. Its rapid spread through *E. coli* strains suggests that it has a role in the commensal lifestyle (Johnson, Delavari, Kuskowski and Stell, 2001; Schubert et al., 2009; van Elsas et al., 2011), which was supported by our earlier observations that yersiniabactin is more prevalent in host-associated *E. coli* isolates (figure 4.4). The salmochelin biosynthesis gene *iroB* had a higher level of expression in the colon compared to *entB* (enterobactin). This may be because enterobactin has a reduced ability to bind iron within the GI-tract due to binding of host molecules, such as lipocalin-2, which are secreted into the intestinal lumen (Raffatellu et al., 2009; Chassaing et al., 2012). Also, salmochelin expression and production has been shown to be induced under slightly alkaline conditions (Foster et al., 1994), such as those found in the lower intestine (Evans et al., 1988; Fallingborg, 1999; Engevik et al., 2013).

Competitive index assays showed that both *ΔtonB* mutants had a 2.5 log reduction in colonisation compared to the wild-type, suggesting an important role for ferric iron uptake in GI-tract colonisation. *ΔtonB* mutants hypersecrete siderophores, as shown by liquid CAS assays, which may result in them becoming iron starved in the gut. As *E. coli* is thought to reside in the mucus layer in the gut, siderophore diffusion may be limited due to the viscosity of the

mucus (Kümmerli et al., 2009). As a result, *ΔtonB* mutants secreting large amounts of siderophores trap any iron surrounding the cell, preventing it from being internalised by other ferric uptake systems. Catecholate uptake *E. coli* K-12 strain MG1655 mutants have been shown to hypersecrete enterobactin and also have significant reduction in colonisation compared to the wild-type (Pi et al., 2012). However, in both our study and in Pi et al. (2012) the wild-type in these competition assays was detected at much higher densities than when competed against mutants that do not hypersecrete siderophores. This suggests that the wild-type may be cheating by utilising siderophores produced by *ΔtonB* mutants, enabling higher levels of growth. As a result, *ΔtonB* mutants may not have reduced fitness or colonisation ability in the GI-tract. This is supported by the observation that *E. coli* strain Nissle 1917 *ΔtonB* mutants are able to colonise the gut as well as the wild-type when they colonise separately (Deriu et al., 2013). In Pi et al. (2012) *ΔtonB* mutants showed very poor colonisation of the GI-tract, becoming undetectable after 5 days in competitive index assays. However, our competitive index assays were carried out in mice pre-treated with ampicillin and neomycin and then the microbiota was allowed to recover during the assay. In contrast, the assays by Pi et al. (2012) used streptomycin-treated mice that received antibiotics for the duration of the study. Streptomycin is known to increase inflammation in the gut (Spees et al., 2013), during which lipocalin-2 production is increased (Chassaing et al., 2012) and oxygen levels may also increase (Rigottier-Gois, 2013), making siderophore production and Fe^{3+} uptake more important. This could also explain why Pi et al. (2012) observed reduced colonisation by enterobactin synthesis mutants, while in our studies there was no difference in colonisation between GMB91e and the wild-type.

Our analyses have highlighted the heterogeneity of regulation and production of siderophores in *E. coli*. This probably reflects large strain-dependent differences in the requirement for siderophores and the diversity of environments that *E. coli* can adapt to. While our epidemiological data indicates an important role for siderophore biosynthesis in the GI-tract, we did not observe significant

differences in the competitiveness of a siderophore biosynthesis mutant (GMB91e) in a mouse model. Whether this is linked to the ability of mutants to cheat and use siderophores generated by the wild-type strain or to obtain iron through other ways remains to be determined. Although this initial competitive index assay result suggests that siderophore production does not influence *E. coli* fitness in the GI-tract, further studies are needed to investigate whether this is the case for other *E. coli* strains.

6. Conclusions

The aim of this project was to use comparative approaches to investigate traits associated with colonisation of the GI-tract. We developed a new technique for monitoring individual strains of *E. coli* in mixed populations to be able to assess the fitness of multiple strains in parallel in different environments. This barcoding method can potentially be used on large numbers of strains, and may provide greater sensitivity compared to currently available techniques, many of which rely on sampling colonies isolated from a population (Lautenbach et al., 2008). This technique was also relatively efficient, providing a high rate of insertion (77%) and a low rate of phenotypic changes (19%) when barcoding strains. The RT-PCR used for barcode detection showed a large dynamic range and good sensitivity, which was increased when using high-throughput sequencing for the quantification of barcoded strains. Importantly, this sensitivity allowed for the identification of sub-dominant strains which may not be detected using currently available alternative techniques that are limited by sampling size (Lautenbach et al., 2008; Smati et al., 2013). We showed barcoding can be used to reliably characterise populations in different complex environments, including the GI-tract and soil. Barcoding has the potential to be a powerful tool for assessing population dynamics in *E. coli* populations which can be adapted to other bacteria. Together with genome-wide association studies, barcode competition studies may be able to identify adaptive traits associated with certain environments.

For this project, competition studies were performed using several ECOR and GMB strains to identify which strains have increased abundance, and by extension, fitness in certain environments. Using the assumption that the environment shapes the associated *E. coli* population and influences gene distribution (Bergholz et al., 2011; Méric et al., 2013), we hypothesised that *E. coli* isolated from the faeces of healthy hosts would be enriched with genes involved in gut adaptation. Therefore ECOR-F strains should display increased fitness in the GI-tract compared to strains isolated from non-host environments, such as plants (GMB). However, ECOR strains surprisingly showed poor

abundance in the first competition study performed in mice, both in faecal samples and final gut content (figure 3.7). This may be as a result of the ECOR collection being approximately 30 years older than the GMB collection, possibly becoming compromised due to long term storage as has been previously observed (Lang and Malik, 1996; Johnson, Delavari, Stell, et al., 2001; Prakash et al., 2012). As a result, a full GMB and ECOR comparison was not performed in our competition studies, with only GMB strains being used for subsequent experiments, the majority of which were isolated from spinach (see table A.1). If long term storage is responsible for the reduced colonisation ability of ECOR strains in the gut, more recently isolated faecal *E. coli* strains could be used to perform further competition studies.

However, our mouse competition study results suggest that there may be differences in prevalence between strains based on phylogenetic group, as has been observed previously in several different hosts (reviewed in Tenallion et al., 2010). The fact that the phylogenetic population structure is different between faecal and non-host isolates suggests that phylogroup is important in environmental fitness (Bergholz et al., 2011; Méric et al., 2013). Our results showed that B2 group strains were dominant in most mouse faecal samples analysed. Strains in the B2 group have been suggested to be host-specialists, being limited to endothermic vertebrates (Gordon and Cowling, 2003), which includes mice. However, we determined fitness by measuring the abundance of individual strains in faecal samples, which has been shown to underrepresent the number of strains present in the whole GI-tract (Zoetendal et al., 2002; Dixit et al., 2004; Schierack et al., 2009; Abraham et al., 2012; Gordon et al., 2015). It has been observed that different phylogroups of *E. coli* have increased prevalence in different regions of the porcine GI-tract, suggesting that they are adapted to different niches within the gut (Dixit et al., 2004). This may result in a reduced prevalence of certain groups or strains in faecal samples, which may more closely represent the populations present in the rectum and distal colon (Dixit et al., 2004). This may explain the differences we observed in phylogroup prevalence in the mouse competition studies. Analysis of the intestinal content

samples from the second mouse experiment should help determine whether there were differences in the distribution of phylogenetic groups between the distal colon, proximal colon, caecum and ileum in our study. It is important to highlight, however, that few strains were used in our competition experiments, so further studies are needed to confirm these general observations about phylogeny and colonisation ability.

As *E. coli* is considered to have a biphasic lifestyle, switching from internal (host) to external (non-host) environments (van Elsas et al., 2011), it is not clear how easy it is to distinguish between "host associated" and "non-host associated" strains. The difference in phylogenetic population structure between isolates from host and non-host environments suggests that these two populations are unique and may possess genes and phenotypes that are beneficial for their respective environments, as has been previously observed (Méric et al., 2013). However, the original source of the non-host isolates used in this study (the GMB collection) is unknown and may have arisen from a recent faecal contamination event. For this reason, it is unclear to what extent the GMB and ECOR collections are different. Although our soil competition study showed a relatively similar initial die-off for different strains of *E. coli* (figure 3.14), only GMB isolates were used, so it is not possible to determine whether host and non-host strains may have differences in environmental persistence. Further studies are required to investigate how host and non-host populations of *E. coli* are formed and interact with one another.

A population-level comparison between the ECOR-F and GMB collections did, however, show a difference in siderophore production levels with ECOR strains producing a higher amount of siderophores. It is possible that higher siderophore production and the possession of multiple siderophore systems increase fitness in the GI-tract to a greater extent than in the external environment. Although siderophore genes (salmochelin and enterobactin) were shown to have increased expression in *S. enterica* Typhimurium in alfalfa root exudates (Hao et al., 2012), it is not clear to what extent *E. coli* requires siderophores for survival and growth on plants. Ferric iron availability is not

homogeneous on leaves (Joyner and Lindow, 2000; Leveau and Lindow, 2001) and can be reduced by the presence of polyphenolic compounds or oxalic acid that can sequester iron (Karamanoli et al., 2011), as can be found in spinach and rocket. However, the requirement for high production levels of siderophores may also depend on the microbiota found on plants and how they compete for iron. The density of bacteria present on leaves has been measured as up to 10^8 CFU/g (Hirano and Upper, 2000; Lindow and Brandl, 2003), whereas the microbiota in the GI-tract of both humans and mice can reach densities of 10^{11}-10^{12} CFU/g (Berg, 1996; Lee et al., 2013), leading to potentially higher levels of competition for iron in the gut. Which siderophores surrounding bacteria are using is also important, as competing bacteria that utilise siderophores with higher affinities for Fe^{3+} may inhibit the growth of *E. coli* (Weaver and Kolter, 2004; Joshi et al., 2006). Also, within the host, lipocalin-2 is secreted into the lumen of the GI-tract, which binds to enterobactin and prevents its function (Chassaing et al., 2014). This may result in a selective pressure in the gut for *E. coli* to possess multiple siderophore systems.

Aerobactin production was associated with high total siderophore production and was more frequently found in ECOR-F strains compared to GMB isolates. Aerobactin has been shown previously to be enriched in intestinal *E. coli* isolates, supporting a possible association with the GI-tract environment (Nowrouzian et al. 2001a; 2001b). The aerobactin system can be found either within the chromosome or on ColV or ColBM plasmids (De Lorenzo et al., 1986; Gao et al., 2014), which may influence its spread through the *E. coli* population. It is possible that the close proximity of bacteria and environmental conditions within the GI-tract support a higher level of HGT compared to on plants. However, the salmochelin system can also be carried on ColV or ColBM plasmids (Gao et al., 2014), but does not show an increased prevalence in ECOR-F isolates. The multiplex PCRs used in this study were unable to distinguish between chromosomal and plasmid aerobactin and salmochelin genes, so it is not clear to what extent plasmid carriage is responsible for the differences in prevalence of these two systems within the GMB and ECOR-F collections.

Neither of the strains possessing aerobactin (GMB104 and GMB23) did well in mouse competition study 2; however, both of these strains were from phylogroup A, which was in general detected at low levels in faecal samples. This suggests that aerobactin production alone is not a strong indicator of increased fitness in the gut. Also, initial competitive index assays did not support our hypothesis that siderophores were important in *E. coli* colonisation and persistence in the GI-tract. Earlier observations showed decreased colonisation of *E. coli* K-12 strain MG1655 *ΔentA* mutants (Pi et al., 2012). However, our competitive index assay results may contradict those by Pi et al. (2012) as a result of experimental design. The antibiotic treatment was different, with antibiotics being administered throughout the Pi et al. (2012) study, rather than being removed before gavage. Removal of antibiotics may have resulted in changes in the microbiota (Manichanh et al., 2010; Panda et al., 2014; Linninge et al., 2015), or a reduction in inflammation (Vijay-Kumar et al., 2010; Spees et al., 2013), which may have altered the requirement for siderophores. Further competitive index assays in a mouse inflammation model or during prolonged antibiotic treatment may determine whether siderophore use is altered. In agreement with Pi et al. (2012), we observed that the *tonB* mutants were reduced in comparison to the wild-type, but it was not possible to determine whether cheating was occurring. As a result, we were unable to confirm whether ferric iron uptake was required for increased fitness in the gut. Also, as only a small number of strains were used in the competition studies and competitive index assays, further experiments are required to determine whether siderophore production has a role in GI-tract colonisation by *E. coli*.

It is possible that differences on the distribution of phylogroups between environments may contribute to observed differences in siderophore production and gene presence. We observed that more siderophore systems were present on average in ECOR-F isolates compared to GMB strains (figure 4.3); however, a larger proportion of the ECOR collection belongs to the B2 phylogenetic group. Our multiplex PCRs showed that the B2 group was enriched for siderophore systems compared to other phylogenetic groups, particularly

yersiniabactin and salmochelin. However, no significant differences were observed in siderophore production on CAS plates between phylogenetic groups (data not shown). As a result, although phylogroup may be a strong indicator of the possession of multiple siderophore systems, it may not be correlated with strain-specific differences in siderophore gene regulation. It is important to note, however, that salmochelin and yersiniabactin did not influence total siderophore production on CAS plates, as yersiniabactin was not produced under the conditions used in our studies and salmochelin would replace enterobactin. This may be why phylogroup is linked to siderophore gene presence, but not siderophore production.

The observed differences between ECOR-F and GMB in siderophore production seem largely explained by aerobactin gene presence or absence. However, only 12.7% of the isolates used in this study possessed the aerobactin system and significant differences were observed in strains that only possessed enterobactin (figure 4.5), suggesting that expression and regulation of siderophore genes explains a large amount of variation in siderophore production between individual strains. We observed differences in expression for the four siderophore systems in *E. coli* when grown under iron limitation, with a strong correlation between enterobactin/salmochelin expression and overall siderophore production (figure 5.1). Regulation of siderophore production, through the Fur protein, is tightly controlled alongside other systems, such as iron storage and the redox stress response, to maintain an optimal level of intracellular iron (Andrews et al., 2003). Factors, both within the bacterial cell that alter iron requirement and outside the cell in the environment that influence iron availability, can potentially alter the regulation and expression of siderophore systems.

The liquid CAS, Csàky and Arnow assays of wild-type strains and siderophore knockout mutants confirmed changes in siderophore production depending on carbon source that have been previously observed in *E. coli* strain Nissle 1917 (Valdebenito et al., 2006). As different metabolic pathways require different enzymes, changes in carbon source can affect the iron requirement of a cell.

For example, growth on glycerol requires respiration and involves the TCA cycle, which requires iron, whereas iron demand may be reduced during growth on glucose, which can be used during fermentation (Valdebenito et al., 2006; Seo et al., 2014). As a result, siderophore production may be under carbon catabolite repression, possibly through the activity of CRP (Zhan et al., 2005). We further examined the link between carbon source and iron homeostasis in *E. coli* using GN2 BIOLOG plates and observed altered metabolism in several of our siderophore knockout mutants. Our results support observations of changes in metabolic pathways towards fermentation to reduce the requirement for iron when iron availability is low (O'Brien et al., 2013; Seo et al., 2014). As carbon source availability is different depending on environment, the requirement for iron may change between host and non-host environments. Alternatively, the production of siderophores may alter carbon source use and metabolism in cells as has been observed in *E. coli* siderophore biosynthesis mutants (Lv et al., 2014). It is highly likely that, as both iron and carbon are essential components of bacterial cells, alterations in the acquisition or utilisation of either of these nutrients will greatly affect growth and gene expression in bacteria.

Other factors have also been shown to affect siderophore use, including pH (Valdebenito et al., 2006), competing bacteria in the microbiota (Deriu et al., 2013) and oxygen levels (Zheng et al., 1999). Further studies of individual isolates should allow for more detailed characterisation and comparisons and help determine how siderophore regulation and production is altered based on these environmental factors. The growth conditions used in this study did not induce yersiniabactin expression, so further studies may also be able to determine what environmental conditions are required for yersiniabactin production. However, high levels of yersiniabactin production have been observed in for *E. coli* strain Nissle 1917 (Valdebenito et al., 2006) and *E. coli* ABU and UTI strains (Watts et al., 2012) under growth conditions similar to those used in our study. It is possible that the strains used in this study do not have the correct genetic background, or are unable to express yersiniabactin at high levels under the conditions tested.

In conclusion, this work has added to the evidence that phylogenetic groups in *E. coli* can influence environmental fitness. This may be through differences in gene content between groups, as we observed with siderophore genes. Although we did not show any clear benefits for siderophore production within the GI-tract, further studies should help elucidate more detailed information about regulation of siderophores and how this is influenced by feedback from the host or non-host environment. Our newly developed barcoding method, alongside genome-wide association studies, should allow for the identification of other genes or traits that are associated with increased fitness in different environments. This knowledge is essential to assess the relationship between the environment, the associated *E. coli* populations and the potential risk they represent to human health.

References

Abdul-Tehrani, H., Hudson, A.J., Chang, Y.-S., Timms, A.R., Hawkins, C., Williams, J.M., Harrison, P.M., Guest J.R., and Andrews, S.C. (1999) 'Ferritin mutants of *Escherichia coli* are iron deficient and growth impaired, and *fur* mutants are iron deficient', *Journal of Bacteriology*, **181**(5): 1415-1428.

Abraham, S., Gordon, D.M., Chin, J., Brouwers, H.J., Njuguna, P., Groves, M.D., Zhang, R., and Chapman, T.A. (2012) 'Molecular characterization of commensal *Escherichia coli* adapted to different compartments of the porcine gastrointestinal tract', *Applied and Environmental Microbiology* **78**(19): 6799-6803.

Abramoff, M.D., Magalhães, P.J., and Ram, S.J. (2004) 'Image processing with ImageJ', *Biophotonics international*, **11**(7): 36-42.

Achard, M.E., Chen, K.W., Sweet, M.J., Watts, R.E., Schroder, K., Schembri, M.A., and McEwan, A.G. (2013) 'An antioxidant role for catecholate siderophores in *Salmonella*', *Biochemical Journal* **454**(3): 543-549.

Adler, C., Corbalan, N.S., Peralta, D.R., Pomares, M.F., de Cristóbal R.E., and Vincent, P.A. (2014) 'The alternative role of enterobactin as an oxidative stress protector allows *Escherichia coli* colony development', *PLoS One* **9**(1): e84734.

Adlerberth, I., Hanson, L.A., Svanborg, C., Svennerholm, A.M., Nordgren S., and Wold A.E. (1995) 'Adhesins of *Escherichia coli* associated with extra-intestinal pathogenicity confer binding to colonic epithelial cells', *Microbial Pathogenesis* **18**(6): 373-385.

Adlerberth, I., Jalil, F., Carlsson, B., Mellander, L., Hanson, L.A., Larsson, P., Khalil, K., and Wold, A.E. (1998) 'High turnover rate of *Escherichia coli* strains in the intestinal flora of infants in Pakistan', *Epidemiology and Infection* **121**(03): 587-598.

Aktan, I., La Ragione, R.M., and Woodward, M.J. (2007) 'Colonization, persistence, and tissue tropism of *Escherichia coli* O26 in conventionally reared weaned lambs', *Applied and Environmental Microbiology* **73**(3): 691-698.

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) 'The niche of *Escherichia coli*', in Walk, S.T., and Peng, P.C. (eds.) *Population Genetics of Bacteria: A Tribute to Thomas S. Whittam*. Washington, DC, ASM Press: 69-89.

Alpert, C., Scheel, J., Engst, W., Loh G., and Blaut M. (2009) 'Adaptation of protein expression by *Escherichia coli* in the gastrointestinal tract of gnotobiotic mice', *Environmental Microbiology* **11**(4): 751-761.

Altier, C. (2005) 'Genetic and environmental control of *Salmonella* invasion', *Journal of Microbiology* **43**(S): 85-92.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones F. (2003) 'Bacterial iron homeostasis', *FEMS Microbiology Reviews* **27**(2-3): 215-237.

Anisimov, R., Brem, D., Heesemann, J., and Rakin, A. (2005a) 'Molecular mechanism of YbtA-mediated transcriptional regulation of divergent overlapping promoters *ybtA* and *irp6* of *Yersinia enterocolitica*', *FEMS Microbiology Letters* **250**(1): 27-32.

Anisimov, R., Brem, D., Heesemann, J. and Rakin, A. (2005b) 'Transcriptional regulation of high pathogenicity island iron uptake genes by YbtA', *International Journal of Medical Microbiology* **295**(1): 19-28.

Apperloorenkema, H.Z., Van der Waaij, B.D., and Van der Waaij D. (1990) 'Determination of colonization resistance of the digestive-tract by biotyping of *Enterobacteriaceae*', *Epidemiology and Infection* **105**(2): 355-361.

Archibald, F. (1983) '*Lactobacillus plantarum*, an organism not requiring iron', *FEMS Microbiology Letters* **19**(1): 29-32.

Archibald, F.S., and Duong, M.N. (1984) 'Manganese acquisition by *Lactobacillus plantarum*', *Journal of Bacteriology* **158**(1): 1-8.

Arnold, K.W., and Kaspar, C.W. (1995) 'Starvation- and stationary-phase-induced acid tolerance in *Escherichia coli* O157:H7', *Applied and Environmental Microbiology* **61**(5): 2037-2039.

Arnow, L.E. (1937) 'Colorimetric determination of the components of 3,4-dihydroxyphenylalaninetyrosine mixtures', *Journal of Biological Chemistry* **118**(2): 531-537.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J.J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal,E.G., Wang, J., Guarner, F., Pedersen, O., De Vos, W.M., Brunak, S., Dore, J., Weissenbach, J., Ehrlich, S.D., and Bork, P. (2011) 'Enterotypes of the human gut microbiome' *Nature* **473**(7346): 174-180.

Aslam, M., and Service, C. (2006) 'Antimicrobial resistance and genetic profiling of *Escherichia coli* from a commercial beef packing plant', *Journal of Food Protection* **69**(7): 1508-1513.

Atuma, C., Strugala, V., Allen, A., and Holm, L. (2001) 'The adherent gastrointestinal mucus gel layer: thickness and physical state *in vivo*', *American Journal of Physiology – Gastrointestinal and Liver Physiology* **280**(5): G922-G929.

Autieri, S.M., Lins, J.J., Leatham, M.P., Laux, D.C., Conway, T., and Cohen, P.S. (2007) 'L-fucose stimulates utilization of D-ribose by *Escherichia coli* MG1655 *ΔfucAO* and *E. coli* Nissle 1917 *ΔfucAO* mutants in the mouse intestine and in M9 minimal medium', *Infection and Immunity* **75**(11): 5465-5475.

Bajka, B.H., Rigby, N.M., Cross, K.L., Macierzanka, A., and Mackie, A.R. (2015) 'The influence of small intestinal mucus structure on particle transport *ex vivo*', *Colloids and Surfaces B: Biointerfaces* **135**: 73-80.

Barak, J.D., Jahn, C.E., Gibson, D.L., and Charkowski, A.O. (2007) 'The role of cellulose and O-antigen capsule in the colonization of plants by *Salmonella enterica*', *Molecular Plant-Microbe Interactions Journal* **20**(9): 1083-1091.

Barelmann, I., Taraz, K., Budzikiewicz, H., Geoffroy, V., and Meyer, J.M. (2002) 'The structures of the pyoverdins from two *Pseudomonas fluorescens* strains accepted mutually by their respective producers', *Zeitschrift fur Naturforschung C: A Journal of Biosciences* **57**(1-2): 9-16.

Barnes, R.H., Fiala, G., and Kwong, E. (1963) 'Decreased growth rate resulting from prevention of coprophagy', *Federation Proceedings* **22**: 125-128.

Barnich, N., and Darfeuille-Michaud, A. (2007) 'Adherent-invasive *Escherichia coli* and Crohn's disease', *Current Opinion Gastroenterology* **23**(1): 16-20.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science* **315**(5819): 1709-1712.

Bartholomew, B., and Hill, M.J. (1984) 'The pharmacology of dietary nitrate and the origin of urinary nitrate', *Food and Chemical Toxicology* **22**(10): 789-795.

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

Baur, B., Hanselmann, K., Schlimme, W., and Jenni, B. (1996) 'Genetic transformation in freshwater: *Escherichia coli* is able to develop natural competence', *Applied and Environmental Microbiology* **62**(10): 3673-3678.

Beiko, R.G., Harlow, T.J., and Ragan, M.A. (2005) 'Highways of gene sharing in prokaryotes', *Proceedings of the National Academy of Sciences of the United States of America* **102**(40): 14332-14337.

Beinert, H., Holm, R.H., and Munck, E. (1997) 'Iron-sulfur clusters: nature's modular, multipurpose structures', *Science* **277**(5326): 653-659.

Benjamin, W.H. Jr., Turnbough, C.L. Jr., Posey, B.S., and Briles, D.E. (1985) 'The ability of *Salmonella typhimurium* to produce the siderophore enterobactin is not a virulence factor in mouse typhoid', *Infection and Immunity* **50**(2): 392-397.

Benjamini, Y., and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society Series B* **57**(1): 289-300.

Berg, R.D. (1996) 'The indigenous gastrointestinal microflora', *Trends in Microbiology* **4**(11): 430-435.

Berger, C.N., Shaw, R.K., Brown, D.J., Clare, S., Dougan, G., Pallen, M.J., and Frankel, G. (2009a) 'Interaction of *Salmonella enterica* with basil and other salad leaves', *The ISME Journal* **3**(2): 261-265.

Berger, C.N., Shaw, R.K., Ruiz-Perez, F., Nataro, J.P., Henderson, I.R., Pallen, M.J., and Frankel, G. (2009b) 'Interaction of enteroaggregative *Escherichia coli* with salad leaves', *Environmental Microbiology Reports* **1**(4): 234-239.

Bergholz, P.W., Noar, J.D., and Buckley, D.H. (2011) 'Environmental patterns are imposed on the population structure of *Escherichia coli* after faecal deposition', *Applied and Environmental Microbiology* **77**(1): 211-219.

Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2007) 'Ohno's dilemma: evolution of new genes under continuous selection', *Proceedings of the National Academy of Sciences of the United States of America* **104**(43): 17004-17009.

Bergthorsson, U., and Ochman, H. (1998) 'Distribution of chromosome length variation in natural isolates of *Escherichia coli*', *Molecular Biology and Evolution* **15**(1): 6-16

Bermúdez, M., and Hazen, T.C. (1988) 'Phenotypic and genotypic comparison of *Escherichia coli* from pristine tropical waters', *Applied and Environmental Microbiology* **54**(4): 979-983.

Berry, E.D., and Miller, D.N. (2005) 'Cattle feedlot soil moisture and manure content: II. Impact on *Escherichia coli* O157', *Journal of Environmental Quality* **34**(2): 656-663.

Bétis, F., Brest, P., Hofman, V., Guignot, J., Kansau, I., Rossi, B., Servin, A., and Hofman, P. (2003) 'Afa/Dr diffusely adhering *Escherichia coli* infection in T84 cell monolayers induces increased neutrophil transepithelial migration, which in turn promotes cytokine-dependent upregulation of decay-accelerating factor (CD55), the receptor for Afa/Dr adhesins', *Infection and Immunity* **71**(4): 1774-1783.

Bingle, L.E., Bailey, C.M., and Pallen, M.J. (2008) 'Type VI secretion: a beginner's guide', *Current Opinion in Microbiology* **11**(1): 3-8.

Blanchard, J.L., Wholey, W.Y., Conlon, E.M., and Pomposiello, P.J. (2007) 'Rapid changes in gene expression dynamics in response to superoxide reveal SoxRS-dependent and independent transcriptional networks', *PLoS One* **2**(11): e1186.

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997) 'The complete genome sequence of *Escherichia coli* K-12', *Science* **277**(5331): 1453-1462.

Bleuel, C., Große, C., Taudte, N., Scherer, J., Wesenberg, D., Krauß, G.J., Nies, D.H., and Grass, G. (2005) 'TolC is involved in enterobactin efflux across the outer membrane of *Escherichia coli*', *Journal of Bacteriology* **187**(19): 6701-6707.

Bobrov, A.G., Kirillina, O., Fetherston, J.D., Miller, M.C., Burlison, J.A., and Perry, R.D. (2014) 'The *Yersinia pestis* siderophore, yersiniabactin, and the ZnuABC system both contribute to zinc acquisition and the development of lethal septicaemic plague in mice', *Molecular Microbiology* **93**(4): 759-775.

Bochner, B.R. (2009) 'Global phenotypic characterization of bacteria', *FEMS Microbiology Reviews* **33**(1): 191-205.

Bochner, B.R., and Savageau, M.A. (1977) 'Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms', *Applied and Environmental Microbiology* **33**(2): 434-444.

Bosshard, P.P., Zbinden, R., Abels, S., Boddinghaus, B., Altwegg, M., and Bottger, E.C. (2006) '16S rRNA gene sequencing versus the API 20 NE system and the VITEK 2 ID-GNB card for identification of nonfermenting Gram-negative bacteria in the clinical laboratory', *Journal of Clinical Microbiology* **44**(4): 1359-1366.

Boucher, D.H. (1985) 'The idea of mutualism, past and future', in Boucher, D.H. (eds.) *The Biology of Mutualism: Ecology and Evolution*. New York, Oxford University Press: 1-28.

Boukhalfa, H., and Crumbliss, A.L. (2002) 'Chemical aspects of siderophore mediated iron transport', *BioMetals* **15**: 325-339.

Boyd, E.F., and Hartl, D.L. (1998) 'Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution', *Journal of Bacteriology* **180**(5): 1159-1165.

Brandl, M.T. (2008) 'Plant lesions promote the rapid multiplication of *Escherichia coli* O157:H7 on postharvest lettuce', *Applied and Environmental Microbiology* **74**(17): 5285-5289.

Brandl, M.T., and Amundson, R. (2008) 'Leaf age as a risk factor in contamination of lettuce with *Escherichia coli* O157:H7 and *Salmonella enterica*', *Applied and Environmental Microbiology* **74**(8): 2298-2306.

Braun, V., and Braun, M. (2002) 'Iron transport and signalling in *Escherichia coli*', *FEBS Letters* **529**(1): 78-85.

Braun, V., Braun, M., and Killmann, H. (2004) 'Ferrichrome- and citrate-mediated iron transport', in Crosa, J., Mey, A., and Payne, S. (eds.) *Iron Transport in Bacteria*. Washington, DC, ASM Press: 158-177.

Braun, V., Brazel-Faisst, C., and Schneider, R. (1984) 'Growth stimulation of *Escherichia coli* in serum by iron(III) aerobactin: recycling of aerobactin', *FEMS Microbiology Letters* **21**(1): 99-103.

Braun, V., Pramanik, A., Gwinner, T., Koberle, M., and Bohn, E. (2009) 'Sideromycins: tools and antibiotics', *BioMetals* **22**(1): 3-13.

Brem, D., Pelludat, C., Rakin, A., Jacobi, C.A., and Heesemann, J. (2001) 'Functional analysis of yersiniabactin transport genes of *Yersinia enterocolitica*', *Microbiology* **147**(5): 1115-1127.

Brennan, F.P., Abram, F., Chinalia, F.A., Richards, K.G., and O'Flaherty, V. (2010) 'Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil', *Applied and Environmental Microbiology* **76**(7): 2175-2180.

Brennan, F.P., O'Flaherty, V., Kramers, G., Grant, J., and Richards, K.G. (2010) 'Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils', *Applied and Environmental Microbiology* **76**(5): 1449-1455.

Brumbaugh, A.R., Smith, S.N., Subashchandrabose, S., Himpsl, S.D., Hazen, T.H., Rasko, D.A., and Mobley, H.L. (2015) 'Blocking yersiniabactin import attenuates extraintestinal pathogenic *Escherichia coli* in cystitis and pyelonephritis and represents a novel target to prevent urinary tract infection', *Infection and Immunity* **83**(4): 1443-1450.

Buckling, A., Harrison, F., Vos, M., Brockhurst, M.A., Gardner, A., West, S.A., and Griffin, A. (2007) 'Siderophore-mediated cooperation and virulence in *Pseudomonas aeruginosa*', *FEMS Microbiology Ecology* **62**(2): 135-141.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Sadowsky, M.J., and Ishii, S. (2006) 'Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed', *Environmental Microbiology* **8**(3): 504-513.

Cao, J., Woodhall, M.R., Alvarez, J., Cartron, M.L., and Andrews, S.C. (2007) 'EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe^{2+} transporter that is cryptic in *Escherichia coli* K-12 but functional in *E. coli* O157:H7', *Molecular Microbiology* **65**(4): 857-875.

Carvalho, B.M., Guadagnini, D., Tsukumo, D.M., Schenka, A.A., Latuf-Filho, P., Vassallo, J., Dias, J.C., Kutoba, L.T., Carvalheira, J.B., and Saad, M.J. (2012) 'Modulation of gut microbiota by antibiotics improves insulin signalling in high-fat fed mice', *Diabetologia* **55**(10): 2823-2834.

Caugant, D.A., Levin, B.R., and Selander R.K. (1981) 'Genetic diversity and temporal variation in the *Escherichia coli* population of a human host', *Genetics* **98**(3): 467-490.

Cavaletti, L., Monciardini, P., Bamonte, R., Schumann, P., Rohde, M., Sosio, M., and Donadio, S. (2006) 'New lineage of filamentous, spore-forming, Gram-positive bacteria from soil', *Applied and Environmental Microbiology* **72**(6): 4360-4369.

Caza, M., Lépine, F., and Dozois, C.M. (2011) 'Secretion, but not overall synthesis, of catecholate siderophores contributes to virulence of extraintestinal pathogenic *Escherichia coli*', *Molecular Microbiology* **80**(1): 266-282.

Caza, M., Lépine, F., Milot, S., and Dozois, C.M. (2008) 'Specific roles of the *iroBCDEN* genes in virulence of an avian pathogenic *Escherichia coli* O78 strain and in production of salmochelins', *Infection and Immunity* **76**(8): 3539-3549.

Cescau, S., Cwerman, H., Létoffé, S., Delepelaire, P., Wandersman, C., and Biville, F. (2007) 'Heme acquisition by hemophores', *BioMetals* **20**(3): 603-613.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., and Conway, T. (2004) 'Carbon nutrition of *Escherichia coli* in the mouse intestine', *Proceedings of the National Academy of Sciences of the United States of America* **101**(19): 7427-7432.

Chang, Y.J., Land, M., Hauser, L., Chertkov, O., Del Rio, T.G., Nolan, M., Copeland, A., Tice, H., Cheng, J.F., Lucas, S., Han, C., Goodwin, L., Pitluck, S., Ivanova, N., Ovchinikova, G., Pati, A., Chen, A., Palaniappan, K., Mavromatis, K., Liolios, K., Brettin, T., Fiebig, A., Rohde, M., Abt, B., Goker, M., Detter, J.C., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.P., and Lapidus, A. (2011) 'Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21)', *Standards in Genomic Sciences* **5**(1): 97-111.

Charpentier, X., Polard, P., and Claverys, J.P. (2012) 'Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS?', *Current Opinion in Microbiology* **15**(5): 570-576.

Chassaing, B., Ley, R.E., and Gewirtz, A.T. (2014) 'Intestinal epithelial cell toll-like receptor 5 regulates the intestinal microbiota to prevent low-grade inflammation and metabolic syndrome in mice', *Gastroenterology* **147**(6): 1363-1377.

Chassaing, B., Srinivasan, G., Delgado, M.A., Young, A.N., Gewirtz, A.T., and Vijay-Kumar, M. (2012) 'Fecal lipocalin 2, a sensitive and broadly dynamic non-invasive biomarker for intestinal inflammation', *PLoS One* **7**(9): e44328.

Chaturvedi, K.S., and Henderson, J.P. (2014) 'Pathogenic adaptations to host-derived antibacterial copper', *Frontiers in Cellular and Infection Microbiology* **4**: 3.

Chaturvedi, K.S., Hung, C.S., Crowley, J.R., Stapleton, A.E., and Henderson, J.P. (2012) 'The siderophore yersiniabactin binds copper to protect pathogens during infection', *Nature Chemical Biology* **8**(8): 731-736.

Chaturvedi, K.S., Hung, C.S., Giblin, D.E., Urushidani, S., Austin, A.M., Dinauer, M.C., and Henderson, J.P. (2014) 'Cupric yersiniabactin is a virulence-associated superoxide dismutase mimic', *ACS Chemical Biology* **9**(2): 551-561.

Chekabab, S.M., Paquin-Veillette, J., Dozois, C.M., and Harel, J. (2013) 'The ecological habitat and transmission of *Escherichia coli* O157:H7', *FEMS Microbiology Letters* **341**(1): 1-12.

Chenault, S.S., and Earhart, C.F. (1991) 'Organization of genes encoding membrane proteins of the *Escherichia coli* ferrienterobactin permease', *Molecular Microbiology* **5**(6): 1405-1413.

Cherepanov, P.P. and Wackernagel, W. (1995) 'Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant', *Gene* **158**(1): 9-14.

Chiang, S.M., Dong, T., Edge, T.A., and Schellhorn, H.E. (2011) 'Phenotypic diversity caused by differential RpoS activity among environmental *Escherichia coli* isolates', *Applied and Environmental Microbiology* **77**(22): 7915-7923.

Ciacci-Woolwine, F., Blomfield, I.C., Richardson, S.H., and Mizel, S.B. (1998) '*Salmonella* flagellin induces tumor necrosis factor alpha in human promonocytic cell lines', *Infection and Immunity* **66**(3): 1127-1134.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013) 'The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups', *Environmental Microbiology Reports* **5**(1): 58-65.

Clermont, O., Lescat, M., O'Brien, C.L., Gordon, D.M., Tenaillon, O., and Denamur, E. (2008) 'Evidence for a human-specific *Escherichia coli* clone', *Environmental Microbiology* **10**(4): 1000-1006.

Coenye, T., and Vandamme, P. (2003) 'Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes', *FEMS Microbiology Letters* **228**(1): 45-49.

Cooke, E.M., Hettiaratchy, I.G., and Buck, A.C. (1972) 'Fate of ingested *Escherichia coli* in normal persons', *Journal of Medical Microbiology* **5**(3): 361-369.

Cooley, M.B., Miller, W.G., and Mandrell, R.E. (2003) 'Colonization of *Arabidopsis thaliana* with *Salmonella enterica* and enterohemorrhagic *Escherichia coli* O157:H7 and competition by *Enterobacter asburiae*', *Applied and Environmental Microbiology* **69**(8): 4915-4926.

Cordero, O.X., Ventouras, L.A., DeLong, E.F., and Polz, M.F. (2012) 'Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations', *Proceedings of the National Academy of Sciences of the United States of America* **109**(49): 20059-20064.

Cornelis, P. (2010) 'Iron uptake and metabolism in pseudomonads', *Applied Microbiology and Biotechnology* **86**(6): 1637-1645.

Corpe, W.A., and Rheem, S. (1989) 'Ecology of the methylotrophic bacteria on living leaf surfaces', *FEMS Microbiology Letters* **62**(4): 243-249.

Cowart, R.E. (2002) 'Reduction of iron by extracellular iron reductases: implications for microbial iron acquisition', *Archives of Biochemistry and Biophysics* **400**(2): 273-281.

Craig, S.A., Carpenter, C.D., Mey, A.R., Wyckoff, E.E., and Payne, S.M. (2011) 'Positive regulation of the *Vibrio cholerae* porin OmpT by iron and Fur', *Journal of Bacteriology* **193**(23): 6505-6511.

Crosa, J.H., and Walsh, C.T. (2002) 'Genetics and assembly line enzymology of siderophore biosynthesis in bacteria', *Microbiology and Molecular Biology Reviews* **66**(2): 223-249.

Crouch, M. L., M. Castor, J. E. Karlinsey, T. Kalhorn and F. C. Fang (2008) 'Biosynthesis and IroC-dependent export of the siderophore salmochelin are essential for virulence of *Salmonella enterica* serovar Typhimurium', *Molecular Microbiology* **67**(5): 971-983.

Crowley, D. (2006) 'Microbial siderophores in the plant rhizosphere', in Barton, L., and Abadia, J. (eds.) *Iron Nutrition in Plants and Rhizospheric Microorganisms*. Netherlands, Springer: 169-198.

Croxen, M.A., and Finlay, B.B. (2010) 'Molecular mechanisms of *Escherichia coli* pathogenicity', *Nature Reviews Microbiology* **8**(1): 26-38.

Csàky, T.Z. (1948) 'On the estimation of bound hydroxylamine in biological materials', *Acta Chemica Scandinavica* **2**: 450-454.

Cullender, T.C., Chassaing, B., Janzon, A., Kumar, K., Muller, C.E., Werner, J.J., Angenent, L.T., Bell, M.E., Hay, A.G., Peterson, D.A., Walter, J., Vijay-Kumar, M., Gewirtz, A.T., and Ley, R.E. (2013) 'Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut', *Cell Host & Microbe* **14**(5): 571-581.

D'Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., and Lewis, K. (2010) 'Siderophores from neighbouring organisms promote the growth of uncultured bacteria', *Chemistry & Biology* **17**(3): 254-264.

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

De Chial, M., Ghysels, B., Beatson, S.A., Geoffroy, V., Meyer, J.M., Pattery, T., Baysse, C., Chablain, P., Parsons, Y.N., Winstanley, C., Cordwell, S.J., and Cornelis, P. (2003) 'Identification of type II and type III pyoverdine receptors from *Pseudomonas aeruginosa*', *Microbiology* **149**(4): 821-831.

De Lorenzo, V., Bindereif, A., Paw, B.H., and Neilands, J.B. (1986) 'Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12', *Journal of Bacteriology* **165**(2): 570-578.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F., and Cerf-Bensussan, N. (2011) 'Trade-off between bile resistance and nutritional competence drives *Escherichia coli* diversification in the mouse gut', *PLoS Genetics* **7**(6): e1002107.

De Vos, C.R., Lubberding, H.J., and Bienfait, H.F. (1986) 'Rhizosphere acidification as a response to iron deficiency in bean plants', *Plant Physiology* **81**(3): 842-846.

De Vos, D., De Chial, M., Cochez, C., Jansen, S., Tümmler, B., Meyer, J.M., and Cornelis, P. (2001) 'Study of pyoverdine type and production by *Pseudomonas aeruginosa* isolated from cystic fibrosis patients: prevalence of type II pyoverdine isolates and accumulation of pyoverdine-negative mutations', *Archives of Microbiology* **175**(5): 384-388.

DeBoy, R.T., and Craig, N.L. (2000) 'Target site selection by Tn*7*:*att*Tn*7* transcription and target activity', *Journal of Bacteriology* **182**(11): 3310-3313.

Deering, A.J., Mauer, L.J., and Pruitt, R.E. (2012) 'Internalization of *E. coli* O157:H7 and *Salmonella* in plants: a review', *Food Research International* **45**(2): 567-575.

Deering, A.J., Pruitt, R.E., Mauer, L.J., and Reuhs, B.L. (2011) 'Identification of the cellular location of internalized *Escherichia coli* O157:H7 in mung bean, *Vigna radiata*, by immunocytochemical techniques', *Journal of Food Protection* **74**(8): 1224-1230.

Delany, I., Spohn, G., Rappuoli, R., and Scarlato, V. (2001) 'The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*', *Molecular Microbiology* **42**(5): 1297-1309.

Dellagi, A., Brisset, M.N., Paulin, J.P., and Expert, D. (1998) 'Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity', *Molecular Plant-Microbe Interactions* **11**(8): 734-742.

Dellagi, A., Rigault, M., Segond, D., Roux, C., Kraepiel, Y., Cellier, F., Briat, J.F., Gaymard F., and Expert, D. (2005) 'Siderophore-mediated upregulation of *Arabidopsis* ferritin expression in response to *Erwinia chrysanthemi* infection', *Plant Journal* **43**(2): 262-272.

Deriu, E., Liu, J.Z., Pezeshki, M., Edwards, R.A., Ochoa, R.J., Contreras, H., Libby, S.J., Fang, F.C., and Raffatellu, M. (2013) 'Probiotic bacteria reduce *Salmonella* Typhimurium intestinal colonization by competing for iron', *Cell Host & Microbe* **14**(1): 26-37.

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. (2002) 'Influence of soil on faecal indicator organisms in a tidally influenced subtropical environment', *Applied and Environmental Microbiology* **68**(3): 1165-1172.

Diallo, S., Crépin, A., Barbey, C., Orange, N., Burini, J.F., and Latour, X. (2011) 'Mechanisms and recent advances in biological control mediated through the potato rhizosphere', *FEMS Microbiology ecology* **75**(3): 351-364

Diard, M., Garry, L., Selva, M., Mosser, T., Denamur, E., and Matic, I. (2010) 'Pathogenicity-Associated Islands in extraintestinal pathogenic

Escherichia coli are fitness elements involved in intestinal colonization', *Journal of Bacteriology* **192**(19): 4885-4893.

Didelot, X., and Maiden, M.C. (2010) 'Impact of recombination on bacterial evolution', *Trends in Microbiology* **18**(7): 315-322.

Diez-Villasenor, C., Almendros, C., Garcia-Martinez J., and Mojica, F.J. (2010) 'Diversity of CRISPR loci in *Escherichia coli*', *Microbiology* **156**(5): 1351-1361.

Dixit, S.M., Gordon, D.M., Wu, X., Chapman, T., Kailasapathy, K., and Chin, J.J. (2004) 'Diversity analysis of commensal porcine *Escherichia coli* – associations between genotypes and habitat in the porcine gastrointestinal tract', *Microbiology* **150**: 1735-1740.

Dobbin, H.S., Hovde, C.J., Williams, C.J., and Minnich, S.A. (2006) 'The *Escherichia coli* O157 flagellar regulatory gene *flhC* and not the flagellin gene *fliC* impacts colonization of cattle', *Infection and Immunity* **74**(5): 2894-2905.

Dobrindt, U., Chowdary, M.G., Krumbholz, G., and Hacker, J. (2010) 'Genome dynamics and its impact on evolution of *Escherichia coli*', *Medical Microbiology Immunology* **199**(3): 145-154.

Dong, T.G., Ho, B.T., Yoder-Himes, D.R., and Mekalanos, J.J. (2013) 'Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*', *Proceedings of the National Academy of Sciences of the United States of America* **110**(7): 2623-2628.

Dozois, C.M., Daigle, F., and Curtiss, R. (2003) 'Identification of pathogen-specific and conserved genes expressed *in vivo* by an avian pathogenic *Escherichia coli* strain', *Proceedings of the National Academy of Sciences of the United States of America* **100**(1): 247-252.

Dozois, C.M., Fairbrother, J.M., Harel, J., and Bosse, M. (1992) '*pap*- and *pil*-related DNA sequences and other virulence determinants associated with *Escherichia coli* isolated from septicemic chickens and turkeys', *Infection and Immunity* **60**(7): 2648-2656.

Drechsel, H., Stephan, H., Lotz, R., Haag, H., Zähner, H., Hantke, K., and Jung, G. (1995) 'Structure elucidation of yersiniabactin, a siderophore from highly virulent *Yersinia* strains', *Liebigs Annalen* **1995**(10): 1727-1733.

Duncan, S.H., Louis, P., Thomson, J.M., and Flint, H.J. (2009) 'The role of pH in determining the species composition of the human colonic microbiota', *Environmental Microbiology* **11**(8): 2112-2122.

Duriez, P., Clermont, O., Bonacorsi, S., Bingen, E., Chaventré, A., Elion, J., Picard, B., and Denamur, E. (2001) 'Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations', *Microbiology* **147**(6): 1671-1676.

Durrett, R., and Levin, S. (1997) 'Allelopathy in spatially distributed populations', *Journal of Theoretical Biology* **185**(2): 165-171.

Durso, L.M., Smith, D., and Hutkins, R.W. (2004) 'Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains', *Applied and Environmental Microbiology* **70**(11): 6466-6472.

Dykhuizen, D.E., and Green, L. (1991) 'Recombination in *Escherichia coli* and the definition of biological species', *Journal of Bacteriology* **173**(22): 7257-7268.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005) 'Diversity of the human intestinal microbial flora', *Science* **308**(5728): 1635-1638.

Engevik, M.A., Faletti, C.J., Paulmichl, M., and Worrell, R.T. (2013) 'Prebiotic properties of galursan HF 7K on mouse gut microbiota', *Cell Physiology and Biochemistry* **32**(7): 96-110.

Enocksson, A., Lundberg, J., Weitzberg, E., Norrby-Teglund, A., and Svenungsson, B. (2004) 'Rectal nitric oxide gas and stool cytokine levels during the course of infectious gastroenteritis', *Clinical and Diagnostic Laboratory Immunology* **11**(2): 250-254.

Erdem, A.L., Avelino, F., Xicohtencatl-Cortes, J., Girón, J.A. (2007) 'Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*', *Journal of Bacteriology* **189**(20): 7426-7435.

Ermund, A., Gustafsson, J.K., Hansson, G.C., and Keita, A.V. (2013) 'Mucus properties and goblet cell quantification in mouse, rat and human ileal Peyer's patches', *PLoS One* **8**(12): e83688.

Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.B., Bui, H., Le Bouguénec C., and Denamur, E. (2004) 'A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*', *Molecular Biology and Evolution* **21**(6): 1085-1094.

Escobar-Páramo, P., Grenet, K., Le Menac'h, A., Rode, L., Salgado, E., Amorin, C., Gouriou, S., Picard, B., Rahimy, M.C., Andremont, A., Denamur E., and Ruimy R. (2004) 'Large-scale population structure of human commensal *Escherichia coli* isolates', *Applied and Environmental Microbiology* **70**(9): 5698-5700.

Escobar-Paramo, P., Le Menac'h, A., Le Gall, T., Amorin, C., Gouriou, S., Picard, B., Skurnik, D., and Denamur, E. (2006) 'Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates', *Environmental Microbiology* **8**(11): 1975-1984.

Escobar-Páramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vaury, C., Denamur, E., and Lecointre, G. (2004) 'Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study', *Molecular Phylogenetics and Evolution* **30**(1): 243-250.

Esnault, E., Valens, M., Espeli, O., and Boccard, F. (2007) 'Chromosome structuring limits genome plasticity in *Escherichia coli*', *PLoS Genetics* **3**(12): e226.

Espey, M.G. (2013) 'Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota', *Free Radical Biology & Medicine* **55**: 130-140.

Evans, D.F., Pye, G., Bramley, R., Clark, A.G., Dyson, T.J., and Hardcastle, J.D. (1988) 'Measurement of gastrointestinal pH profiles in normal ambulant human subjects', *Gut* **29**(8): 1035-1041.

Eydallin, G., Ryall, B., Maharjan, R., and Ferenci, T. (2013) 'The nature of laboratory domestication changes in freshly isolated *Escherichia coli* strains', *Environmental Microbiology* **16**(3): 813-828.

Fabich, A.J., Jones, S.A., Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J.W., Hightower, G.A., Smith, J.T., Autieri, S.M., Leatham, M.P., Lins, J.J., Allen, R.L., Laux, D.C., Cohen, P.S., and Conway, T. (2008) 'Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine', *Infection and Immunity* **76**(3): 1143-1152.

Fallingborg, J. (1999) 'Intraluminal pH of the human gastrointestinal tract', *Danish Medical Bulletin* **46**(3): 183-196.

Falsenstein, J. (1978) 'Cases in which parsimony or compatibility methods will be positively misleading', *Systematic Zoology* **27**: 401-410.

Farrell, M.J., and Finkel, S.E. (2003) 'The growth advantage in stationary-phase phenotype conferred by *rpoS* mutations is dependent on the pH and nutrient environment', *Journal of Bacteriology* **185**(24): 7044-7052.

Faulkner, M.J., Ma, Z., Fuangthong, M., and Helmann, J.D. (2012) 'Derepression of the *Bacillus subtilis* PerR peroxide stress response leads to iron deficiency', *Journal of Bacteriology* **194**(5): 1226-1235.

Feldmann, F., Sorsa, L.J., Hildinger, K., and Schubert, S. (2007) 'The salmochelin siderophore receptor IroN contributes to invasion of urothelial cells by extraintestinal pathogenic *Escherichia coli in vitro*', *Infection and Immunity* **75**(6): 3183-3187.

Ferenci, T. (2005) 'Maintaining a healthy SPANC balance through regulatory and mutational adaptation', *Molecular Microbiology* **57**(1): 1-8.

Ferenci, T., and Spira, B. (2007) 'Variation in stress responses within a bacterial species and the indirect cost of stress resistance', *Annals of the New York Academy of Sciences* **1113**: 105-113.

Fetherston, J.D., Bearden, S.W., and Perry, R.D. (1996) 'YbtA, an AraC-type regulator of the *Yersinia pestis* pesticin/yersiniabactin receptor', *Molecular Microbiology* **22**(2): 315-325.

Fetherston, J.D., Bertolino, V.J., and Perry, R.D. (1999) 'YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*', *Molecular Microbiology* **32**(2): 289-299.

Finkel, S.E. (2006) 'Long-term survival during stationary phase: evolution and the GASP phenotype', *Nature Reviews Microbiology* **4**(2): 113-120.

Finkel, S.E., and Kolter, R. (2001) 'DNA as a nutrient: novel role for bacterial competence gene homologs' *Journal of Bacteriology* **183**(21): 6288-6293.

Fischbach, M.A., Lin, H., Zhou, L., Yu, Y., Abergel, R.J., Liu, D.R., Raymond, K.N., Wanner, B.L., Strong, R.K., Walsh, C.T., Aderem, A., and Smith, K.D. (2006) 'The pathogen-associated *iroA* gene cluster mediates bacterial evasion of lipocalin 2', *Proceedings of the National Academy of Sciences of the United States of America* **103**(44): 16502-16507.

Fogle, C.A., Nagle, J.N., and Desai, M.M. (2008) 'Clonal interference, multiple mutations and adaptation in large asexual populations', *Genetics* **180**(4): 2163-2173.

Foster, J.W. (2004) '*Escherichia coli* acid resistance: tales of an amateur acidophile', *Nature Reviews Microbiology* **2**(11): 898-907.

Foster, J.W., Park, Y.K., Bang, I.S., Karem, K., Betts, H., Hall, H.K., and Shaw, E. (1994) 'Regulatory circuits involved with pH-regulated gene expression in *Salmonella typhimurium*', *Microbiology* **140**(2): 341-352.

Franklund, C.V., Baron, S.F., and Hylemon, P.B. (1993) 'Characterization of the *baiH* gene encoding a bile acid-inducible NADH:flavin oxidoreductase from *Eubacterium* sp. strain VPI 12708', *Journal of Bacteriology* **175**(10): 3002-3012.

Franz, E., van Hoek, A.H., Bouw, E., and Aarts, H.J. (2011) 'Variability of *Escherichia coli* O157 strain survival in manure-amended soil in relation to strain origin, virulence profile and carbon nutrition profile', *Applied and Environmental Microbiology* 77(22): 8088-8096.

Franza, T., Mahe, B., and Expert, D. (2005) '*Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection', *Molecular Microbiology* **55**(1): 261-275.

Freter, R., Brickner, H., Fekete, J., Vickerman, M.M., and Carey, K.E. (1983) 'Survival and implantation of *Escherichia coli* in the intestinal tract', *Infection and Immunity* **39**(2): 686-703.

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J.M., Topping, D.L., Suzuki, T., Taylor, T.D., Itoh, K., Kikuchi, J., Morita, H., Hattori, M., and Ohno, H. (2011) 'Bifidobacteria can protect from enteropathogenic infection through production of acetate', *Nature* **469**(7331): 543-547.

Furman, M., Fica, A., Saxena, M., Di Fabio, J.L., and Cabello, F.C. (1994) '*Salmonella typhi* iron uptake mutants are attenuated in mice', *Infection and Immunity* **62**(9): 4091-4094.

Furrer, J.L., Sanders, D.N., Hook-Barnard, I.G., and McIntosh, M.A. (2002) 'Export of the siderophore enterobactin in *Escherichia coli*: involvement of a 43 kDa membrane exporter', *Molecular Microbiology* **44**(5): 1225-1234.

Gantois, I., Ducatelle, R., Pasmans, F., Haesebrouck, F., Hautefort, I., Thompson, A., Hinton, J.C., and Van Immerseel, F. (2006) 'Butyrate specifically down-regulates *Salmonella* pathogenicity island 1 gene expression', *Applied and Environmental Microbiology* **72**(1): 946-949.

Gao, Q., Wang, X., Xu, H., Xu, Y., Ling, J., Zhang, D., Gao, S., and Liu, X. (2012) 'Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model', *BMC Microbiology* **12**: 143.

Gao, Y.D., Zhao, Y.Q., and Huang, J.F. (2014) 'Metabolic modeling of common *Escherichia coli* strains in human gut microbiome', *BioMed Research International* **2014**: 694967.

Garcia, E.C., Brumbaugh, A.R., and Mobley, H.L. (2011) 'Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection', *Infection and Immunity* **79**(3): 1225-1235.

Garenaux, A., Caza, M., and Dozois, C.M. (2011) 'The ins and outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*', *Veterinary Microbiology* **153**(1-2): 89-98.

Garmendia, J., Frankel, G., and Crepin, V.F. (2005) 'Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation', *Infection and Immunity* **73**(5): 2573-2585.

Gauger, E.J., Leatham, M.P., Mercado-Lubo, R., Laux, D.C., Conway, T., and Cohen, P.S. (2007) 'Role of motility and the *flhDC* Operon in *Escherichia coli* MG1655 colonization of the mouse intestine', *Infection and Immunity* **75**(7): 3315-3324.

Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009) 'Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung', *Proceedings of the National Academy of Sciences of the United States of America* **106**(38): 16422-16427.

Gehring, A.M., Bradley, K.A., and Walsh, C.T. (1997) 'Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate', *Biochemistry* **36**(28): 8495-8503.

Gevers, D., Vandepoele, K., Simillon, C., and Van de Peer, Y. (2004) 'Gene duplication and biased functional retention of paralogs in bacterial genomes', *Trends in Microbiology* **12**(4): 148-154.

Ghysels, B., Dieu, B.T., Beatson, S.A., Pirnay, J.P., Ochsner, U.A., Vasil, M.L., and Cornelis, P. (2004) 'FpvB, an alternative type I ferripyoverdine receptor of *Pseudomonas aeruginosa*', *Microbiology* **150**(6): 1671-1680.

Giel, J.L., Rodionov, D., Liu, M., Blattner, F.R., and Kiley, P.J. (2006) 'IscR-dependent gene expression links iron-sulphur cluster assembly to the

control of O_2-regulated genes in *Escherichia coli*', *Molecular Microbiology* **60**(4): 1058-1075.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I., Somerfield, P., Fuhrman, J.A., and Field, D. (2012) 'Defining seasonal marine microbial community dynamics', *The ISME Journal* **6**(2): 298-308.

Giraud, A., Arous, S., De Paepe, M., Gaboriau-Routhiau, V., Bambou, J.C., Rakotobe, S., Lindner, A.B., Taddei, F., and Cerf-Bensussan, N. (2008) 'Dissecting the genetic components of adaptation of *Escherichia coli* to the mouse gut', *PLoS Genetics* **4**(1): e2.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009) 'Identifying genetic determinants needed to establish a human gut symbiont in its habitat', *Cell Host & Microbe* **6**(3): 279-289.

Gorden, J., and Small, P.L. (1993) 'Acid resistance in enteric bacteria', *Infection and Immunity* **61**(1): 364-367.

Gordon, D.M., Bauer, S., and Johnson, J.R. (2002) 'The genetic structure of *Escherichia coli* populations in primary and secondary habitats', *Microbiology* **148**(5): 1513-1522.

Gordon, D.M., Clermont, O., Tolley, H., and Denamur, E. (2008) 'Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method', *Environmental Microbiology* **10**(10): 2484-2496.

Gordon, D.M. and Cowling, A. (2003) 'The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects', *Microbiology* **149**(12): 3575-3586.

Gordon, D.M., O'Brien, C.L., and Pavli, P. (2015) '*Escherichia coli* diversity in the lower intestinal tract of humans', *Environmental Microbiology Reports* **7**(4): 642-648.

Goto, D.K., and Yan, T. (2011) 'Genotypic diversity of *Escherichia coli* in the water and soil of tropical watersheds in Hawaii', *Applied and Environmental Microbiology* **77**(12): 3988-3997.

Gouyer, V., Dubuquoy, L., Robbe-Masselot, C., Neut, C., Singer, E., Plet, S., Geboes, K., Desreumaux, P., Gottrand, F., and Desseyn, J.L. (2015) 'Delivery of a mucin domain enriched in cysteine residues strengthens the intestinal mucous barrier', *Scientific Reports* **5**: 9577.

Gransden, W.R., Eykyn, S.J., Phillips, I., and Rowe, B. (1990) 'Bacteremia due to *Escherichia coli*: a study of 861 episodes', *Reviews of Infectious Diseases* **12**(6): 1008-1018.

Grass, G. (2006) 'Iron transport in *Escherichia coli*: all has not been said and done', *BioMetals* **19**(2): 159-172.

Grass, G., Franke, S., Taudte, N., Nies, D.H., Kucharski, L.M., Maguire, M.E., and Rensing, C. (2005) 'The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum', *Journal of Bacteriology* **187**(5): 1604-1611.

Gray, J.P., and Herwig, R.P. (1996) 'Phylogenetic analysis of the bacterial communities in marine sediments', *Applied and Environmental Microbiology* **62**(11): 4049-4059.

Griffin, A.S., West, S.A., and Buckling, A. (2004) 'Cooperation and competition in pathogenic bacteria', *Nature* **430**(7003): 1024-1027.

Guerinot, M.L. (1994) 'Microbial iron transport', *Annual Review of Microbiology* **48**(1): 743-772.

Gulati, A.S., Shanahan, M.T., Arthur, J.C., Grossniklaus, E., von Furstenberg, R.J., Kreuk, L., Henning, S.J., Jobin, C., and Sartor, R.B. (2012) 'Mouse background strain profoundly influences Paneth cell function and intestinal microbial composition', *PLoS One* **7**(2): e32403.

Gupta, V., and Carroll, K.S. (2014) 'Sulfenic acid chemistry, detection and cellular lifetime', *Biochimica et Biophysica Acta – General Subjects* **1840**(2): 847-875.

Hacker, J., and Carniel, E. (2001) 'Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes', *EMBO Reports* **2**(5): 376-381.

Hafez, M., Hayes, K., Goldrick, M., Warhurst, G., Grencis, R., and Roberts, I.S. (2009) 'The K5 capsule of *Escherichia coli* strain Nissle 1917 is important in mediating interactions with intestinal epithelial cells and chemokine induction', *Infection and Immunity* **77**(7): 2995-3003.

Hahn, M.W. (2009) 'Distinguishing among evolutionary models for the maintenance of gene duplicates', *Journal of Heredity* **100**(5): 605-617.

Haiko, J., and Westerlund-Wikstrom, B. (2013) 'The role of the bacterial flagellum in adhesion and virulence', *Biology* **2**(4): 1242-1267.

Hall, T.A. (1999) 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucleic Acids Symposium Series* **41**: 95-98.

Hammar, M., Bian, Z., and Normark, S. (1996) 'Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **93**(13): 6562-6566.

Hancock, V., Seshasayee, A.S., Ussery, D.W., Luscombe, N.M., and Klemm, P. (2008) 'Transcriptomics and adaptive genomics of the asymptomatic bacteriuria *Escherichia coli* strain 83972', *Molecular Genetics and Genomics* **279**(5): 523-534.

Hansson, G.C. (2012) 'Role of mucus layers in gut infection and inflammation', *Current Opinion in Microbiology* **15**(1): 57-62.

Hantke, K. (1987) 'Selection procedure for deregulated iron transport mutants (*fur*) in *Escherichia coli* K 12: *fur* not only affects iron metabolism', *Molecular and General Genetics* **210**(1): 135-139.

Hantke, K. (1990) 'Dihydroxybenzoylserine – a siderophore for *E. coli*', *FEMS Microbiology Letters* **55**(1-2): 5-8.

Hantke, K. (2001) 'Iron and metal regulation in bacteria', *Current Opinion in Microbiology* **4**(2): 172-177.

Hantke, K. (2005) 'Bacterial zinc uptake and regulators', *Current Opinion in Microbiology* **8**(2): 196-202.

Hao, L.Y., Willis, D.K., Andrews-Polymenis, H., McClelland, M., and Barak, J.D. (2012) 'Requirement of siderophore biosynthesis for plant colonization by *Salmonella enterica*', *Applied and Environmental Microbiology* **78**(13): 4561-4570.

Harris, W.R., Carrano, C.J., and Raymond, K.N. (1979) 'Coordination chemistry of microbial iron transport compounds. 16. Isolation, characterization, and formation constants of ferric aerobactin', *Journal of the American Chemical Society* **101**(10): 2722-2727.

Hartz, A., Cuvelier, M., Nowosielski, K., Bonilla, T.D., Green, M., Esiobu, N., McCorquodale, D.S., and Rogerson, A. (2008) 'Survival potential of *Escherichia coli* and enterococci in subtropical beach sand: implications for water quality managers', *Journal of Environmental Quality* **37**(3): 898-905.

Hayashi, F., Smith, K.D., Ozinsky, A., Hawn, T.R., Yi, E.C., Goodlett, D.R., Eng, J.K., Akira, S., Underhill, D.M., and Aderem, A. (2001) 'The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5', *Nature* 410(6832): 1099-1103.

Hayashi, H., Sakamoto, M., and Benno, Y. (2002) 'Phylogenetic analysis of the human gut microbiota using 16s rDNA clone libraries and strictly anaerobic culture-based methods', *Microbiology and Immunology* **46**(8): 535-548.

He, G., Shankar, R.A., Chzhan, M., Samouilov, A., Kuppusamy, P., and Zweier, J.L. (1999) 'Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging', *Proceedings of the National Academy of Sciences of the United States of America* **96**(8): 4586-4591.

He, X., Qian, W., Wang, Z., Li, Y., and Zhang, J. (2010) 'Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks', *Nature Genetics* **42**(3): 272-276.

Heesemann, J., Hantke, K., Vocke, T., Saken, E., Rakin, A., Stojiljkovic, I., and Berner, R. (1993) 'Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65,000 Da and pesticin sensitivity', *Molecular Microbiology* **8**(2): 397-408.

Henderson, J.P., Crowley, J.R., Pinkner, J.S., Walker, J.N., Tsukayama, P., Stamm, W.E., Hooton, T.M., and Hultgren, S.J. (2009) 'Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic *Escherichia coli*', *PLoS Pathogens* **5**(2): e1000305.

Hengge-Aronis, R. (2002) 'Recent insights into the general stress response regulatory network in *Escherichia coli*', *Journal of Molecular Microbiology and Biotechnology* **4**(3): 341-346.

Herías, V.M., Midtvedt, T., Hanson, L.A., and Wold, A.E. (1995) 'Role of *Escherichia coli* P fimbriae in intestinal colonization in gnotobiotic rats', *Infection and Immunity* **63**(12): 4781-4789.

Herías, V.M., Midtvedt, T., Hanson, L.A., and Wold, A.E. (1997) '*Escherichia coli* K5 capsule expression enhances colonization of the large intestine in the gnotobiotic rat', *Infection and Immunity* **65**(2): 531-536.

Herías, V.M., Robertson, A.K., Midtvedt, T., and Wold, A.E. (2001) '*Escherichia coli* S fimbriae do not contribute to intestinal colonization or translocation in the gnotobiotic rat', *Microbial Pathogenesis* **31**(2): 103-107.

Herman, K.M., Hall, A.J., and Gould, L.H. (2015) 'Outbreaks attributed to fresh leafy vegetables, United States, 1973-2012', *Epidemiology and infection* **143**(14): 3011-3021.

Herzer, P.J., Inouye, S., Inouye, M. and Whittam, T.S. (1990) 'Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*', *Journal of Bacteriology* **172**(11): 6175-6181.

Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010) 'Bacterial competition: surviving and thriving in the microbial jungle', *Nature Reviews Microbiology* **8**(1): 15-25.

Hilali, F., Ruimy, R., Saulnier, P., Barnabe, C., Lebouguenec, C., Tibayrenc, M., and Andremont, A. (2000) 'Prevalence of virulence genes and clonality in *Escherichia coli* strains that cause bacteremia in cancer patients', *Infection and Immunity* **68**(7): 3983-3989.

Hirano, S.S., and Upper, C.D. (2000) 'Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* – a pathogen, ice nucleus and epiphyte', *Microbiology and Molecular Biology Reviews* **64**(3): 624-653.

Hohnadel, D., and Meyer, J.M. (1988) 'Specificity of pyoverdine-mediated iron uptake among fluorescent *Pseudomonas* strains', *Journal of Bacteriology* **170**(10): 4865-4873.

Holden, N., Pritchard, L., and Toth, I. (2009) 'Colonization outwith the colon: plants as an alternative environmental reservoir for human pathogenic enterobacteria', *FEMS Microbiology Reviews* **33**(4): 689-703.

Hood, M.I., and Skaar, E.P. (2012) 'Nutritional immunity: transition metals at the pathogen-host interface', *Nature Reviews Microbiology* **10**(8): 525-537.

Hora, R., Warriner, K., Shelp, B.J., and Griffiths, M.W. (2005) 'Internalization of *Escherichia coli* O157:H7 following biological and mechanical disruption of growing spinach plants', *Journal of Food Protection* **68**(12): 2506-2509.

Hurrell, R., and Egli, I. (2010) 'Iron bioavailability and dietary reference values', *The American Journal of Clinical Nutrition* **91**(5): 1461S-1467S.

Ibekwe, A.M., Grieve, C.M., and Yang, C.H. (2007) 'Survival of *Escherichia coli* O157:H7 in soil and on lettuce after soil fumigation', *Canadian Journal of Microbiology* **53**(5): 623-635.

Iguchi, A., Iyoda, S., Terajima, J., Watanabe, H., and Osawa, R. (2006) 'Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome', *Gene* **372**: 199-207.

Ihssen, J., Grasselli, E., Bassin, C., François, P., Piffaretti, J.C., Köster, W., Schrenzel, J., and Egli, T. (2007) 'Comparative genomic hybridization and physiological characterization of environmental isolates indicate that significant (eco-)physiological properties are highly conserved in the species *Escherichia coli*', *Microbiology* **153**(7): 2052-2066.

Imbert, M., and Blondeau, R. (1998) 'On the iron requirement of lactobacilli grown in chemically defined medium', *Current Microbiology* **37**(1): 64-66.

Imlay, J.A. (2003) 'Pathways of oxidative damage', *Annual Review of Microbiology* **57**: 395-418.

Ishii, S., Hansen, D.L., Hicks, R.E., and Sadowsky, M.J. (2007) 'Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior', *Environmental Science & Technology* **41**(7): 2203-2209.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006) 'Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds', *Applied and Environmental Microbiology* **72**(1): 612-621.

Ishii, S., Yan, T., Vu, H., Hansen, D.L., Hicks, R.E., and Sadowsky, M.J. (2010) 'Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils', *Microbes and Environments* **25**(1): 8-14.

Islam, M., Doyle, M.P., Phatak, S.C., Millner, P., and Jiang, X. (2004) 'Persistence of enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water', *Journal of Food Protection* **67**(7): 1365-1370.

Jablasone, J., Warriner, K., and Griffiths, M. (2005) 'Interactions of *Escherichia coli* O157:H7, *Salmonella typhimurium* and *Listeria monocytogenes* on plants cultivated in a gnotobiotic system', *International Journal of Food Microbiology* **99**(1): 7-18.

Jackson, S.A., Patel, I.R., Barnaba, T., LeClerc, J.E., and Cebula, T.A. (2011) 'Investigating the global genomic diversity of *Escherichia coli* using a

multi-genome DNA microarray platform with novel gene prediction strategies', *BMC Genomics* **12**: 349.

Janda, J.M., and Abbott, S.L. (2007) '16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls', *Journal of Clinical Microbiology* **45**(9): 2761-2764.

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X., and Brisse, S. (2008) 'Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains', *BMC Genomics* **9**: 560.

Jeffery, I.B., Claesson, M.J., O'Toole, P.W., and Shanahan, F. (2012) 'Categorization of the gut microbiota: enterotypes or gradients?', *Nature Reviews Microbiology* **10**(9): 591-592.

Johansson, M.E., Gustafsson, J.K., Sjöberg, K.E., Petersson, J., Holm, L., Sjövall, H., and Hansson, G.C. (2010) 'Bacteria penetrate the inner mucus layer before inflammation in the Dextran Sulfate colitis model', *PLoS One* **5**(8): e12238.

Johansson, M.E., Larsson, J.M., and Hansson, G.C. (2011) 'The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions', *Proceedings of the National Academy of Sciences of the United States of America* **108**(Suppl 1): 4659-4665.

Johansson, M.E., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and Hansson, G.C. (2008) 'The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria', *Proceedings of the National Academy of Sciences of the United States of America* **105**(39): 15064-15069.

Johnson, J.R., Delavari, P., Kuskowski, M., and Stell, A.L. (2001) 'Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*', *Journal of Infectious Diseases* **183**(1): 78-88.

Johnson, J.R., Delavari, P., Stell, A.L., Prats, A.G., Carlino, U., and Russo, T.A. (2001) 'Integrity of archival strain collections: the ECOR collection', *ASM News* **67**(6): 288-289.

Johnson, J.R., Johnston, B., Kuskowski, M.A., Nougayrede, J.P., and Oswald, E. (2008) 'Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island', *Journal of Clinical Microbiology* **46**(12): 3906-3911.

Johnson, J.R., Owens, K., Gajewski, A., and Kuskowski, M.A. (2005) 'Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women', *Journal of Clinical Microbiology* **43**(12): 6064-6072.

Johnson, J.R., Owens, K.L., Clabots, C.R., Weissman, S.J., and Cannon, S.B. (2006) 'Phylogenetic relationships among clonal groups of extraintestinal pathogenic *Escherichia coli* as assessed by multi-locus sequence analysis', *Microbes and Infection* **8**(7): 1702-1713.

Johnson, T.J., and Nolan, L.K. (2009) 'Pathogenomics of the virulence plasmids of *Escherichia coli*', *Microbiology and Molecular Biology Reviews* **73**(4): 750-774.

Johnstone, T.C., and Nolan, E.M. (2015) 'Beyond iron: non-classical biological functions of bacterial siderophores', *Dalton Transactions* **44**(14): 6320-6339.

Jones, S.A., Chowdhury, F.Z., Fabich, A.J., Anderson, A., Schreiner, D.M., House, A.L., Autieri, S.M., Leatham, M.P., Lins, J.J., Jorgensen, M., Cohen, P.S., and Conway, T. (2007) 'Respiration of *Escherichia coli* in the mouse intestine', *Infection and Immunity* **75**(10): 4891-4899.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011) 'Anaerobic respiration of *Escherichia coli* in the mouse intestine', *Infection and Immunity* **79**(10): 4218-4226.

Jones, S.A., Jorgensen, M., Chowdhury, F.Z., Rodgers, R., Hartline, J., Leatham, M.P., Struve, C., Krogfelt, K.A., Cohen, P.S., and Conway, T. (2008) 'Glycogen and maltose utilization by *Escherichia coli* O157:H7 in the mouse intestine', *Infection and Immunity* **76**(6): 2531-2540.

Joshi, F., Archana, G., and Desai, A. (2006) 'Siderophore cross-utilization amongst rhizospheric bacteria and the role of their differential affinities for Fe^{3+} on growth stimulation under iron-limited conditions', *Current Microbiology* **53**(2): 141-147.

Joyner, D.C., and Lindow, S.E. (2000) 'Heterogeneity of iron bioavailability on plants assessed with a whole-cell GFP-based bacterial biosensor', *Microbiology* **146**(10): 2435-2445.

Juge, N. (2012) 'Microbial adhesins to gastrointestinal mucus', *Trends in Microbiology* **20**(1): 30-39.

Jurkevitch, E., Hadar, Y., and Chen, Y. (1992) 'Differential siderophore utilization and iron uptake by soil and rhizosphere bacteria', *Applied and Environmental Microbiology* **58**(1): 119-124.

Kamada, N., Chen, G.Y., Inohara, N., and Nunez, G. (2013) 'Control of pathogens and pathobionts by the gut microbiota', *Nature Immunology* **14**(7): 685-690.

Kammler, M., Schön, C., and Hantke, K. (1993) 'Characterization of the ferrous iron uptake system of *Escherichia coli*', *Journal of Bacteriology* **175**(19): 6212-6219.

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

Karamanoli, K., Bouligaraki, P., Constantinidou, H.I., and Lindow, S.E. (2011) 'Polyphenolic compounds on leaves limit iron availability and affect growth of epiphytic bacteria', *Annals of Applied Biology* **159**(1): 99-108.

Karch, H., Tarr, P.I., and Bielaszewska, M. (2005) 'Enterohaemorrhagic *Escherichia coli* in human medicine', *International Journal of Medical Microbiology* **295**(6–7): 405-418.

Kehl-Fie, T.E., Chitayat, S., Hood, M.I., Damo, S., Restrepo, N., Garcia, C., Munro, K.A., Chazin, W.J., and Skaar, E.P. (2011) 'Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense enhancing neutrophil killing of *Staphylococcus aureus*', *Cell Host & Microbe* **10**(2): 158-164.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Schröder, I., Shearer, A.G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R.P., Paulsen, I., and Karp, P.D. (2013) 'EcoCyc: fusing model organism databases with systems biology', *Nucleic Acids Research* **41**(D1): D605-D612.

Keyer, K., and Imlay, J.A. (1996) 'Superoxide accelerates DNA damage by elevating free-iron levels', *Proceedings of the National Academy of Sciences of the United States of America* **93**(24): 13635-13640.

Khachatryan, Z.A., Ktsoyan, Z.A., Manukyan, G.P., Kelly, D., Ghazaryan, K.A., and Aminov, R.I. (2008) 'Predominant role of host genetics in controlling the composition of gut microbiota', *PLoS One* **3**(8): e3064.

Kim, J.C., Yoon, J.W., Kim, C.H., Park, M.S., and Cho, S.H. (2012) 'Repression of flagella motility in enterohemorrhagic *Escherichia coli* O157:H7 by mucin components', *Biochemical and Biophysical Research Communications* **423**(4): 789-792.

Kimmitt, P.T., Harwood, C.R., and Barer, M.R. (2000) 'Toxin gene expression by shiga toxin-producing *Escherichia coli*: the role of antibiotics and the bacterial SOS response', *Emerging Infectious Diseases* **6**(5): 458-465.

Kimura, M. (1990) 'Some models of neutral evolution, compensatory evolution, and the shifting balance process', *Theoretical Population Biology* **37**(1): 150-158.

King, T., Ishihama, A., Kori, A., and Ferenci, T. (2004) 'A regulatory trade-off as a source of strain variation in the species *Escherichia coli*', *Journal of Bacteriology* **186**(17): 5614-5620.

Kirjavainen, P.V., Ouwehand, A.C., Isolauri, E., and Salminen, S.J. (1998) 'The ability of probiotic bacteria to bind to human intestinal mucus', *FEMS Microbiology Letters* **167**(2): 185-189.

Klein, D.A. and Casida, L.E. Jr (1967) '*Escherichia coli* die-out from normal soil as related to nutrient availability and the indigenous microflora', *Canadian Journal of Microbiology* **13**(11): 1461-1470.

Koch, A. L. (1987). 'Why *Escherichia coli* should be renamed *Escherichia ilei*', in Torriani-Gorini, A.L., Rothman, F.G., Silver, S., Wright, A., and Yagil, E.

(eds.) *Phosphate metabolism and cellular regulation in microorganisms*. Washington, D.C., ASM Press: 300-305.

Köhler, C.D. and Dobrindt, U. (2011) 'What defines extraintestinal pathogenic *Escherichia coli*?', *International Journal of Medical Microbiology* **301**(8): 642-647.

Kondrashov, F.A. (2012) 'Gene duplication as a mechanism of genomic adaptation to a changing environment', *Proceedings of the Royal Society of London B: Biological Sciences* **279**(1749): 5048-5057.

Kosek, M., Bern, C., and Guerrant, R.L. (2003) 'The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000', *Bulletin of the World Health Organisation* **81**(3): 197-204.

Köster, W. (1991) 'Iron(III) hydroxamate transport across the cytoplasmic membrane of *Escherichia coli*', *Biology of Metals* **4**(1): 23-32.

Köster, W. and Braun, V. (1990) 'Iron (III) hydroxamate transport into *Escherichia coli*. Substrate binding to the periplasmic FhuD protein', *Journal of Biological Chemistry* **265**(35): 21407-21410.

Kotloff, K.L. (1999) 'Bacterial diarrheal pathogens', *Advances in Pediatric Infectious Diseases* **14**: 219-267.

Krogfelt, K.A., Hjulgaard, M., Sørensen, K., Cohen, P.S., and Givskov, M. (2000) '*rpoS* gene function is a disadvantage for *Escherichia coli* BJ4 during competitive colonization of the mouse large intestine', *Infection and Immunity* **68**(5): 2518-2524.

Kroupitski, Y., Pinto, R., Brandl, M.T., Belausov, E. and Sela, S. (2009) 'Interactions of *Salmonella enterica* with lettuce leaves', *Journal of Applied Microbiology* **106**(6): 1876-1885.

Kubo, K.M., and Craig, N.L. (1990) 'Bacterial transposon Tn*7* utilizes two different classes of target sites', *Journal of Bacteriology* **172**(5): 2774-2778.

Kudinha, T., Kong, F., Johnson, J.R., Andrew, S.C., Anderson, P., and Gilbert, G.L. (2012) 'Multiplex PCR-based reverse line blot assay for simultaneous detection of 22 virulence genes in uropathogenic *Escherichia coli*', *Applied and Environmental Microbiology* **78**(4): 1198-1202.

Kühn, I., Möllby, R., and Tullus, K. (1986) 'Colonization and persistence of *Escherichia coli* phenotypes in the intestines of children aged 0 to 18 months', *Infection* **14**(1): 7-12.

Kümmerli, R., Griffin, A.S., West, S.A., Buckling, A., and Harrison, F. (2009) 'Viscous medium promotes cooperation in the pathogenic bacterium *Pseudomonas aeruginosa*', *Proceedings of the Royal Society of London B: Biological Sciences* **276**(1672): 3531-3538.

Kümmerli, R., van den Berg, P., Griffin, A.S., West, S.A., and Gardner, A. (2010) 'Repression of competition favours cooperation: experimental evidence from bacteria', *Journal of Evolutionary Biology* **23**(4): 699-706.

Kuo, C.H., and Ochman, H. (2010) 'The extinction dynamics of bacterial pseudogenes', *PLoS Genetics* **6**(8): e1001050.

Kupz, A., Fischer, A., Nies, D., Grass, G., Göbel, U., Bereswill, S., and Heimesaat, M.M. (2013) 'Impact of metal ion homeostasis of genetically modified *Escherichia coli* Nissle 1917 and K12 (W3110) strains on colonization properties in the murine intestinal tract', *European Journal of Microbiology and Immunology* **3**(3): 229-235.

Lafont, J.P., Dho, M., D'Hauteville, H.M., Bree, A., and Sansonetti, P.J. (1987) 'Presence and expression of aerobactin genes in virulent avian strains of *Escherichia coli*', *Infection and Immunity* **55**(1): 193-197.

Lang, E., and Malik, K.A. (1996) 'Maintenance of biodegradation capacities of aerobic bacteria during long-term preservation', *Biodegredation* **7**(1): 65-71.

Langman, L., Young, I.G., Frost, G.E., Rosenberg, H., and Gibson, F. (1972) 'Enterochelin system of iron transport in *Escherichia coli*: mutations affecting ferric-enterochelin esterase', *Journal of Bacteriology* **112**(3): 1142-1149.

Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., and Turner, A.K. (2009) 'Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants', *Genome Research* **19**(12): 2308-2316.

Lautenbach, E., Bilker, W.B., Tolomeo, P., and Maslow, J.N. (2008) 'Impact of diversity of colonizing strains on strategies for sampling *Escherichia coli* from faecal specimens', *Journal of Clinical Microbiology* **46**(9): 3094-3096.

Lawley, T.D., Clare, S., Walker, A.W., Stares, M.D., Connor, T.R., Raisen, C., Goulding, D., Rad, R., Schreiber, F., Brandt, C., Deakin, L.J., Pickard, D.J., Duncan, S.H., Flint, H.J., Clark, T.G., Parkhill, J., and Dougan, G. (2012) 'Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice', *PLoS Pathogens* **8**(10): e1002995.

Lawrence, J.G., Ochman, H., and Hartl, D.L. (1991) 'Molecular and evolutionary relationships among enteric bacteria', *Journal of General Microbiology* **137**(8): 1911-1921.

Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E., and Tenaillon, O. (2007) 'Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains', *Molecular Biology and Evolution* **24**(11): 2373-2384.

Leatham, M.P., Banerjee, S., Autieri, S.M., Mercado-Lubo, R., Conway, T., and Cohen, P.S. (2009) 'Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157:H7 growth in the streptomycin-treated mouse intestine', *Infection and Immunity* **77**(7): 2876-2886.

Lecointre, G., Rachdi, L., Darlu, P., and Denamur, E. (1998) '*Escherichia coli* molecular phylogeny using the incongruence length difference test', *Molecular Biology and Evolution* **15**(12): 1685-1695.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012) 'Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing', *Proceedings of the National Academy of Sciences of the United States of America* **109**(41): E2774-2783.

Lee, J.W., and Helmann, J.D. (2007) 'Functional specialization within the Fur family of metalloregulators', *BioMetals* **20**(3-4): 485-499.

Lee, S., Yu, J.K., Park, K., Oh, E.J., Kim, S.Y., and Park, Y.J. (2010) 'Phylogenetic groups and virulence factors in pathogenic and commensal strains of *Escherichia coli* and their association with bla_{CTX-M}', *Annals of Clinical and Laboratory Science* **40**(4): 361-367.

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

Lehmann, L. (2007) 'The evolution of trans-generational altruism: kin selection meets niche construction', *Journal of Evolutionary Biology* **20**(1): 181-189.

Leopold, S.R., Sawyer, S.A., Whittam, T.S., and Tarr, P.I. (2011) 'Obscured phylogeny and possible recombinational dormancy in *Escherichia coli*', *BMC Evolutionary Biology* **11**: 183.

Lerat, E., and Ochman, H. (2004) 'Ψ-Φ: exploring the outer limits of bacterial pseudogenes', *Genome Research* **14**(11): 2273-2278.

Lescat, M., Hoede, C., Clermont, O., Garry, L., Darlu, P., Tuffery, P., Denamur, E., and Picard, B. (2009) '*aes*, the gene encoding the esterase B in *Escherichia coli*, is a powerful phylogenetic marker of the species', *BMC Microbiology* **9**: 273.

Leveau, J.H., and Lindow, S.E. (2001) 'Appetite of an epiphyte: quantitative monitoring of bacterial sugar consumption in the phyllosphere', *Proceedings of the National Academy of Sciences of the United States of America* **98**(6): 3446-3453.

Lewis, S.B., Cook, V., Tighe, R., and Schüller, S. (2012) 'Enterohemorrhagic *Escherichia coli* colonisation of human colonic epithelium *in vitro* and *ex vivo*', *Infection and Immunity* **83**(3): 942-949.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008) 'Worlds within worlds: evolution of the vertebrate gut microbiota', *Nature Reviews Microbiology* **6**(10): 776-788.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006) 'Ecological and evolutionary forces shaping microbial diversity in the human intestine', *Cell* **124**(4): 837-848.

Li, W.H., Gojobori, T., and Nei, M. (1981) 'Pseudogenes as a paradigm of neutral evolution', *Nature* **292**(5820): 237-239.

Liao, D. (1999) 'Concerted evolution: molecular mechanism and biological implications', *American Journal of Human Genetics* **64**(1): 24-30.

Liao, D. (2000) 'Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea', *Journal of Molecular Evolution* **51**(4): 305-317.

Licht, T.R., Tolker-Nielsen, T., Holmstrøm, K., Krogfelt, K.A., and Molin, S. (1999) 'Inhibition of *Escherichia coli* precursor-16S rRNA processing by mouse intestinal contents', *Environmental Microbiology* **1**(1): 23-32.

Lim, E.L., Amaral, L.A., Caron, D.A., and DeLong, E.F. (1993) 'Application of rRNA-based probes for observing marine nanoplanktonic protists', *Applied and Environmental Microbiology* **59**(5): 1647-1655.

Lin, H., Fiscchbach, M.A., Liu, D.R., and Walsh, C.T. (2005) '*In vitro* characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes', *Journal of the American Chemical Society* **127**(31): 11075-11084.

Lindow, S.E., and Brandl, M.T. (2003) 'Microbiology of the phyllosphere', *Applied and Environmental Microbiology* **69**(4): 1875-1883.

Linggood, M.A., Roberts, M., Ford, S., Parry, S.H., and Williams, P.H. (1987) 'Incidence of the aerobactin iron uptake system among *Escherichia coli* isolates from infections of farm animals', *Journal of General Microbiology* **133**(4): 835-842.

Linninge, C., Ahrné, S., and Molin, G. (2015) 'Pre-treatment with antibiotics and *Escherichia coli* to equalise the gut microbiota in conventional mice', *Antonie van Leeuwenhoek* **107**(1): 149-156.

Liu, J., Rutz, J.M., Feix, J.B., and Klebba, P.E. (1993) 'Permeability properties of a large gated channel within the ferric enterobactin receptor, FepA', *Proceedings of the National Academy of Sciences of the United States of America* **90**(22): 10653-10657.

Livak, K.J., and Schmittgen, T.D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method', *Methods* **25**(4): 402-408.

Lloyd, A.L., Rasko, D.A., and Mobley, H.L. (2007) 'Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*', *Journal of Bacteriology* **189**(9): 3532-3546.

Loomis, L.D., and Raymond, K.N. (1991) 'Solution equilibria of enterobactin and metal-enterobactin complexes', *Inorganic Chemistry* **30**(5): 906-911.

Lopez-Velasco, G., Davis, M., Boyer, R.R., Williams, R.C., and Ponder, M.A. (2010) 'Alterations of the phylloepiphytic bacterial community associated with interactions of *Escherichia coli* O157:H7 during storage

of packaged spinach at refrigeration temperatures', *Food Microbiology* **27**(4): 476-486.

Lukjancenko, O., Wassenaar, T., and Ussery, D. (2010) 'Comparison of 61 sequenced *Escherichia coli* genomes', *Microbial Ecology* **60**(4): 708-720.

Lundberg, J.O., Hellstrom, P.M., Lundberg, J.M., and Alving, K. (1994) 'Greatly increased luminal nitric oxide in ulcerative colitis', *Lancet* **344**(8938): 1673-1674.

Lv, H., and Henderson, J.P. (2011) '*Yersinia* High Pathogenicity Island genes modify the *Escherichia coli* primary metabolome independently of siderophore production', *Journal of Proteome Research* **10**(12): 5547-5554.

Lv, H., Hung, C.S., and Henderson, J.P. (2014) 'Metabolomic analysis of siderophore cheater mutants reveals metabolic costs of expression in uropathogenic *Escherichia coli*', *Journal of Proteome Research* **13**(3): 1397-1404.

Mahajan, A., Currie, C.G., Mackie, S., Tree, J., McAteer, S., McKendrick, I., McNeilly, T.N., Roe, A., La Ragione, R.M., Woodward, M.J., Gally, D.L., and Smith, D.G. (2009) 'An investigation of the expression and adhesion function of H7 flagella in the interaction of *Escherichia coli* O157:H7 with bovine intestinal epithelium', *Cellular Microbiology* **11**(1): 121-137.

Majeed, H., Gillor, O., Kerr, B., and Riley, M.A. (2011) 'Competitive interactions in *Escherichia coli* populations: the role of bacteriocins', *The ISME Journal* **5**(1): 71-81.

Makui, H., Roig, E., Cole, S.T., Helmann, J.D., Gros, P., and Cellier, M.F. (2000) 'Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter', *Molecular Microbiology* **35**(5): 1065-1078.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and 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**(1): e53957.

Manichanh, C., Reeder, J., Gibert, P., Varela, E., Llopis, M., Antolin, M., Guigo, R., Knight, R., and Guarner, F. (2010) 'Reshaping the gut microbiome with bacterial transplantation and antibiotic intake', *Genome Research* **20**: 1411-1419.

Mao, B.H., Chang, Y.F., Scaria, J., Chang, C.C., Chou, L.W., Tien, N., Wu, J.J., Tseng, C.C., Wang, M.C., Chang, C.C., Hsu, Y.M., and Teng, C.H. (2012) 'Identification of *Escherichia coli* genes associated with urinary tract infections', *Journal of Clinical Microbiology* **50**(2): 449-456.

Maringanti, S., and Imlay, J.A. (1999) 'An intracellular iron chelator pleiotropically suppresses enzymatic and growth defects of superoxide dismutase-deficient *Escherichia coli*', *Journal of Bacteriology* **181**(12): 3792-3802.

Marlovits, T.C., Haase, W., Herrmann, C., Aller, S.G., and Unger, V.M. (2002) 'The membrane protein FeoB contains an intramolecular G protein essential for Fe(II) uptake in bacteria', *Proceedings of the National Academy of Sciences of the United States of America* **99**(25): 16243-16248.

Marteyn, B., Scorza, F.B., Sansonetti, P.J., and Tang, C. (2011) 'Breathing life into pathogens: the influence of oxygen on bacterial virulence and host responses in the gastrointestinal tract', *Cellular Microbiology* **13**(2): 171-176.

Marteyn, B., West, N.P., Browning, D.F., Cole, J.A., Shaw, J.G., Palm, F., Mounier, J., Prevost, M.C., Sansonetti, P., and Tang, C.M. (2010) 'Modulation of *Shigella* virulence in response to available oxygen *in vivo*', *Nature* **465**(7296): 355-358.

Martin-Galiano, A.J., Ferrandiz, M.J., and de la Campa, A.G. (2001) 'The promoter of the operon encoding the F0F1 ATPase of *Streptococcus pneumoniae* is inducible by pH', *Molecular Microbiology* **41**(6): 1327-1338.

Martinez-Medina, M., and Garcia-Gil, L.J. (2014) '*Escherichia coli* in chronic inflammatory bowel diseases: An update on adherent invasive *Escherichia coli* pathogenicity', *World Journal of Gastrointestinal Pathophysiology* **5**(3): 213-227.

Massé, E., and Gottesman, S. (2002) 'A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **99**(7): 4620-4625.

Massé, E., Majdalani, N., and Gottesman, S. (2003) 'Regulatory roles for small RNAs in bacteria', *Current Opinion in Microbiology* **6**(2): 120-124.

Massé, E., Vanderpool, C.K., and Gottesman, S. (2005) 'Effect of RyhB small RNA on global iron use in *Escherichia coli*', *Journal of Bacteriology* **187**(20): 6962-6971.

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., and de La Cochetiere, M.F. (2013) 'Development of intestinal microbiota in infants and its impact on health', *Trends in Microbiology* **21**(4): 167-173.

Matin, A. (1999) 'pH homeostasis in acidophiles', *Novartis Foundation Symposium* **221**: 152-163.

Matin, A., Wilson, B., Zychlinsky, E., and Matin, M. (1982) 'Proton motive force and the physiological basis of delta pH maintenance in *Thiobacillus acidophilus*', *Journal of Bacteriology* **150**(2): 582-591.

Matte-Tailliez, O., Brochier, C., Forterre, P., and Philippe, H. (2002) 'Archaeal phylogeny based on ribosomal proteins', *Molecular Biology and Evolution* **19**(5): 631-639.

Matthysse, A.G., Deora, R., Mishra, M., and Torres, A.G. (2008) 'Polysaccharides cellulose, poly-β-1,6-*N*-acetyl-D-glucosamine, and colanic acid are

required for optimal binding of *Escherichia coli* O157:H7 strains to alfalfa sprouts and K-12 strains to plastic but not for binding to epithelial cells', *Applied and Environmental Microbiology* **74**(8): 2384-2390.

Maurelli, A.T. (2007) 'Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens', *FEMS Microbiology Letters* **267**(1): 1-8.

Maurelli, A.T., Fernández, R.E., Bloch, C.A., Rode, C.K., and Fasano, A. (1998) '"Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* sand enteroinvasive *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **95**(7): 3943-3948.

McCutcheon, J.P., and Moran, M.A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology* **10**(1): 13-26.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. (1995) 'A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens', *Proceedings of the National Academy of Sciences of the United States of America* **92**(5): 1664-1668.

McHugh, J.P., Rodriguez-Quinones, F., Abdul-Tehrani, H., Svistunenko, D.A., Poole, R.K., Cooper, C.E., and Andrews, S.C. (2003) 'Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis', *Journal of Biological Chemistry* **278**(32): 29478-29486.

McKenzie, G.J., and Craig, N.L. (2006) 'Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn*7* without need for selection of the insertion event', *BMC Microbiology* **6**: 39.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005) 'The microbial pan-genome', *Current Opinion in Genetics & Development* **15**(6): 589-594.

Melican, K., Sandoval, R.M., Kader, A., Josefsson, L., Tanner, G.A., Molitoris, B.A., and Richter-Dahlfors, A. (2011) 'Uropathogenic *Escherichia coli* P and type 1 fimbriae act in synergy in a living host to facilitate renal colonization leading to nephron obstruction', *PLoS Pathogens* **7**(2): e1001298.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M., and Karch, H. (2011) 'Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology', *PLoS One* **6**(7): e22751.

Melotto, M., Panchal, S., and Roy, D. (2014) 'Plant innate immunity against human bacterial pathogens', *Frontiers in Microbiology* **5**: 411.

Mercier, J. and Lindow, S.E. (2000) 'Role of leaf surface sugars in colonization of plants by bacterial epiphytes', *Applied and Environmental Microbiology* **66**(1): 369-374.

Méric, G. (2011) *Molecular and ecological characterisation of Escherichia coli from plants*, PhD. thesis, University of East Anglia.

Méric, G., Kemsley, E.K., Falush, D., Saggers, E.J., and Lucchini, S. (2013) 'Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*', *Environmental Microbiology* **15**(2): 487-501.

Merrell, D.S., Thompson, L.J., Kim, C.C., Mitchell, H., Tompkins, L.S., Lee, A., and Falkow, S. (2003) 'Growth phase-dependent response of *Helicobacter pylori* to iron starvation', *Infection and Immunity* **71**(11): 6510-6525.

Miethke, M., and Marahiel, M.A., (2007) 'Siderophore-based iron acquisition and pathogen control', *Microbiology and Molecular Biology Reviews* **71**(3): 413-451.

Mignard, S., and Flandrois, J.P. (2006) '16S rRNA sequencing in routine bacterial identification: a 30-month experiment', *Journal of Microbiological Methods* **67**(3): 574-581.

Miller, R.D., and Hartl, D.L. (1986) 'Biotyping confirms a nearly clonal population structure in *Escherichia coli*', *Evolution* **40**(1): 1-12.

Mira, A., Ochman, H., and Moran, N.A. (2001) 'Deletional bias and the evolution of bacterial genomes', *Trends in Genetics* **17**(10): 589-596.

Miranda, R.L., Conway, T., Leatham, M.P., Chang, D.E., Norris, W.E., Allen, J.H., Stevenson, S.J., Laux, D.C., and Cohen, P.S. (2004) 'Glycolytic and gluconeogenic growth of *Escherichia coli* O157:H7 (EDL933) and *E. coli* K-12 (MG1655) in the mouse intestine', *Infection and Immunity* **72**(3): 1666-1676.

Moeck, G.S., and Coulton, J.W. (1998) 'TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport', *Molecular Microbiology* **28**(4): 675-681.

Moghadam, M.S., Foo, H.L., Leow, T.C., Rahim, R.A., and Loh, T.C. (2010) 'Novel bacteriocinogenic *Lactobacillus plantarum* strains and their differentiation by sequence analysis of 16S rDNA, 16S-23S and 23S-5S intergenic spacer regions and randomly amplified polymorphic DNA analysis', *Food Technology and Biotechnology* **48**(4): 476-483.

Monsen, E.R., Hallberg, L., Layrisse, M., Hegsted, D.M., Cook, J.D., Mertz, W., and Finch, C.A. (1978) 'Estimation of available dietary iron', *The American Journal of Clinical Nutrition* **31**(1): 134-141.

Moors, M.A., Li, L., and Mizel, S.B. (2001) 'Activation of interleukin-1 receptor-associated kinase by Gram-negative flagellin', *Infection and Immunity* **69**(7): 4424-4429.

Mueller, S.I., Valdebenito, M., and Hantke, K. (2009) 'Salmochelin, the long-overlooked catecholate siderophore of *Salmonella*', *BioMetals* **22**(4): 691-695.

Mulvey, M.R., Switala, J., Borys, A., and Loewen, P.C. (1990) 'Regulation of transcription of *katE* and *katF* in *Escherichia coli*', *Journal of Bacteriology* **172**(12): 6713-6720.

Murphy, K.C. (1998) 'Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*', *Journal of Bacteriology* **180**(8): 2063-2071.

Nagy, T.A., Moreland, S.M., Andrews-Polymenis, H., and Detweiler, C.S. (2013) 'The ferric enterobactin transporter Fep is required for persistent *Salmonella enterica* serovar Typhimurium infection', *Infection and Immunity* **81**(11): 4063-4070.

Nandal, A., Huggins, C.C., Woodhall, M.R., McHugh, J., Rodriguez-Quinones, F., Quail, M.A., Guest, J.R., and Andrews, S.C. (2010) 'Induction of the ferritin gene (*ftnA*) of *Escherichia coli* by Fe^{2+}-Fur is mediated by reversal of H-NS silencing and is RyhB independent', *Molecular Microbiology* **75**(3): 637-657.

Negre, V.L., Bonacorsi, S., Schubert, S., Bidet, P., Nassif, X., and Bingen, E. (2004) 'The siderophore receptor IroN, but not the high-pathogenicity island or the hemin receptor ChuA, contributes to the bacteremic step of *Escherichia coli* neonatal meningitis', *Infection and Immunity* **72**(2): 1216-1220.

Neidhardt, F.C. (1999) 'Bacterial growth: constant obsession with *dN*/*dt*', *Journal of Bacteriology* **181**(24): 7405-7408.

Nguyen, A.T., O'Neill, M.J., Watts, A.M., Robson, C.L., Lamont, I.L., Wilks, A., and Oglesby-Sherrouse, A.G. (2014) 'Adaptation of iron homeostasis pathways by a *Pseudomonas aeruginosa* pyoverdine mutant in the cystic fibrosis lung', *Journal of Bacteriology* **196**(12): 2265-2276.

Niederhoffer, E.C., Naranjo, C.M., Bradley, K.L., and Fee, J.A. (1990) 'Control of *Escherichia coli* superoxide dismutase (*sodA* and *sodB*) genes by the ferric uptake regulation (*fur*) locus', *Journal of Bacteriology* **172**(4): 1930-1938.

Nielsen, K.L., Dynesen, P., Larsen, P., and Frimodt-Moller, N. (2014) 'Faecal *Escherichia coli* from patients with *E. coli* urinary tract infection and healthy controls who have never had a urinary tract infection', *Journal of Medical Microbiology* **63**(4): 582-589.

Niemira, B.A. and Cooke, P.H. (2010) '*Escherichia coli* O157:H7 biofilm formation on Romaine lettuce and spinach leaf surfaces reduces efficacy of irradiation and sodium hypochlorite washes', *Journal of Food Science* **75**(5): M270-M277.

Niki, H., Yamaichi, Y., and Hiraga, S. (2000) 'Dynamic organization of chromosomal DNA in *Escherichia coli*', *Genes & Development* **14**(2): 212-223.

Notley-McRobb, L., King, T., and Ferenci, T. (2002) '*rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses', *Journal of Bacteriology* **184**(3): 806-811.

Nowrouzian, F., Adlerberth, I., and Wold, A.E. (2001a) 'P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli*', *Epidemiology and Infection* **126**(1): 11-18.

Nowrouzian, F., Adlerberth, I., and Wold, A.E. (2006) 'Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells', *Microbes and Infection* **8**(3): 834-840.

Nowrouzian, F., Hesselmar, B., Saalman, R., Strannegard, I.L., Aberg, N., Wold, A.E., and Adlerberth, I. (2003) '*Escherichia coli* in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage', *Pediatric Research* **54**(1): 8-14.

Nowrouzian, F., Ostblom, A.E., Wold, A.E., and Adlerberth, I. (2009) 'Phylogenetic group B2 *Escherichia coli* strains from the bowel microbiota of Pakistani infants carry few virulence genes and lack the capacity for long-term persistence', *Clinical Microbiology and Infection* **15**(5): 466-472.

Nowrouzian, F., Wold, A.E., and Adlerberth, I. (2001b) 'P fimbriae and aerobactin as intestinal colonization factors for *Escherichia coli* in Pakistani infants', *Epidemiology and Infection* **126**(1): 19-23.

Nowrouzian, F., Wold, A.E., and Adlerberth, I. (2005) '*Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants', *Journal of Infectious Diseases* **191**(7): 1078-1083.

Nzakizwanayo, J., Kumar, S., Ogilvie, L.A., Patel, B.A., Dedi, C., Macfarlane, W.M., and Jones, B.V. (2015) 'Disruption of *Escherichia coli* Nissle 1917 K5 capsule biosynthesis, through loss of distinct *kfi* genes, modulates interaction with intestinal epithelial cells and impact on cell health', *PLoS One* **10**(3): e120430.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., and Palsson, B.O. (2013) 'Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction', *Molecular Systems Biology* **9**: 693.

Obata-Yasuoka, M., Ba-Thein, T. Tsukamoto, W., Yoshikawa, H., and Hayashi, H. (2002) 'Vaginal *Escherichia coli* share common virulence factor profiles, serotypes and phylogeny with other extraintestinal *E. coli*', *Microbiology* **148**(9): 2745-2752.

Ochman, H., and Selander, R.K. (1984) 'Standard reference strains of *Escherichia coli* from natural populations', *Journal of Bacteriology* **157**(2): 690-693.

Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M., and Hayashi, T. (2009) 'Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*', *Proceedings of the National Academy of Science of the United States of America* **106**(42): 17939-17944.

Okhuysen, P.C., and DuPont, H.L. (2010) 'Enteroaggregative *Escherichia coli* (EAEC): a cause of acute and persistent diarrhoea of worldwide importance', *Journal of Infectious Diseases* **202**(4): 503-505.

Onderdonk, A., Marshall, B., Cisneros, R., and Levy, S.B. (1981) 'Competition between congenic *Escherichia coli* K-12 strains *in vivo*', *Infection and Immunity* **32**(1): 74-79.

Ongeng, D., Vasquez, G.A., Muyanja, C., Ryckeboer, J., Geeraerd, A.H., and Springael, D. (2011) 'Transfer and internalisation of *Escherichia coli* O157:H7 and *Salmonella enterica* serovar Typhimurium in cabbage cultivated on contaminated manure-amended soil under tropical field conditions in Sub-Saharan Africa', *International Journal of Food Microbiology* **145**(1): 301-310.

Outten, C.E., and O'Halloran, T.V. (2001) 'Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis', *Science* **292**(5526): 2488-2492.

Outten, F.W., Djaman, O., and Storz, G. (2004) 'A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*', *Molecular Microbiology* **52**(3): 861-872.

Ozenberger, B.A., Brickman, T.J., and McIntosh, M.C. (1989) 'Nucleotide sequence of *Escherichia coli* isochorismate synthetase gene *entC* and evolutionary relationship of isochorismate synthetase and other chorismate-utilizing enzymes', *Journal of Bacteriology* **171**(2): 775-783.

Palchevskiy, V. and Finkel, S.E. (2006) '*Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient', *Journal of Bacteriology* **188**(11): 3902-3910.

Pallecchi, L., Lucchetti, C., Bartoloni, A., Bartalesi, F., Mantella, A., Gamboa, H., Carattoli, A., Paradisi, F., and Rossolini, G.M. (2007) 'Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure', *Antimicrobial Agents and Chemotherapy* **51**(4): 1179-1184.

Panda, S., El Khader, I., Casellas, F., Vivancos, J.L., Cors, M.G., Santiago, A., Cuenca, S., Guarner, F., and Manichanh, C. (2014) 'Short-term effect of antibiotics on human gut microbiota', *PLoS One* **9**(4): e95476.

Pande, S., Shitut, S., Freund, L., Westermann, M., Bertels, F., Colesie, C., Bischofs, I.B., and Kost, C. (2015) 'Metabolic cross-feeding via

intercellular nanotubes among bacteria', *Nature Communications* **6**: 6238.

Park, S.J. and Gunsalus, R.P. (1995) 'Oxygen, iron, carbon, and superoxide control of the fumarase *fumA* and *fumC* genes of *Escherichia coli*: role of the *arcA*, *fnr*, and *soxR* gene products', *Journal of Bacteriology* **177**(21): 6255-6262.

Payne, S.M. (1994) 'Detection, isolation, and characterization of siderophores', *Methods in Enzymology* **235**: 329-344.

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., van den Brandt, P.A., and Stobberingh, E.E. (2006) 'Factors influencing the composition of the intestinal microbiota in early infancy', *Pediatrics* **118**(2): 511-521.

Perna, N.T., Mayhew, G.F., Pósfai, G., Elliott, S., Donnenberg, M.S., Kaper, J.B., and Blattner, F.B. (1998) 'Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7', *Infection and Immunity* **66**(8): 3810-3817.

Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R. (2001) 'Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7', *Nature* **409**(6819): 529-533.

Perry, R.D., Balbo, P.B., Jones, H.A., Fetherston, J.D., and DeMoll, E. (1999) 'Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation', *Microbiology* **145**(5): 1181-1190.

Perry, R.D., Shah, J., Bearden, S.W., Thompson, J.M., and Fetherston, J.D. (2003) '*Yersinia pestis* TonB: role in iron, heme, and hemoprotein utilization', *Infection and Immunity* **71**(7): 4159-4162.

Peters, J.E., and Craig, N.L. (2001) 'Tn*7*: smarter than we thought', *Nature Reviews Molecular Cell Biology* **2**(11): 806-814.

Peterson, C.N., Mandel, M.J., and Silhavy, T.J. (2005) '*Escherichia coli* starvation diets: essential nutrients weigh in distinctly', *Journal of Bacteriology* **187**(22): 7549-7553.

Pfeifer, B.A., Wang, C.C., Walsh, C.T., and Khosla, C. (2003) 'Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host', *Applied and Environmental Microbiology* **69**(11): 6698-6702.

Pfeiffer, T. and Bonhoeffer, S. (2004) 'Evolution of cross-feeding in microbial populations', *The American Naturalist* **163**(6): E126-E135.

Phillips, A.D., Navabpour, S., Hicks, S., Dougan, G., Wallis, T., and Frankel, G. (2000) 'Enterohaemorrhagic *Escherichia coli* O157:H7 target Peyer's patches in humans and cause attaching/effacing lesions in both human and bovine intestine', *Gut* **47**(3): 377-381.

Phillips, P.C. (2008) 'Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems', *Nature Reviews Genetics* **9**(11): 855-867.

Pi, H., Jones, S.A., Mercer, L.E., Meador, J.P., Caughron, J.E., Jordan, L., Newton, S.M., Conway, T., and Klebba, P.E. (2012) 'Role of catecholate siderophores in Gram-negative bacterial colonization of the mouse gut', *PLoS One* **7**(11): e50020.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999) 'The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection', *Infection and Immunity* 67(2): 546-553.

Plésiat, P., and Nikaido, H. (1992) 'Outer membranes of Gram-negative bacteria are permeable to steroid probes', *Molecular Microbiology* **6**(10): 1323-1333.

Portal-Celhay, C., and Blaser, M.J. (2012) 'Competition and resilience between founder and introduced bacteria in the *Caenorhabditis elegans* gut', *Infection and Immunity* **80**(3): 1288-1299.

Posey, J.E., and Gherardini, F.C. (2000) 'Lack of a role for iron in the Lyme disease pathogen', *Science* **288**(5471): 1651-1653.

Posey, J.E., Hardham, J.M., Norris, S.J., and Gherardini, F.C. (1999) 'Characterization of a manganese-dependent regulatory protein, TroR, from *Treponema pallidum*', *Proceedings of the National Academy of Sciences of the United States of America* **96**(19): 10887-10892.

Poulsen, L.K., Lan, F., Kristensen, C.S., Hobolth, P., Molin, S., and Krogfelt, K.A. (1994) 'Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA in situ hybridization', *Infection and Immunity* **62**(11): 5191-5194.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995) 'Physiological state of *Escherichia coli* BJ4 growing in the large intestines of streptomycin-treated mice', *Journal of Bacteriology* **177**(20): 5840-5845.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) 'Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes', *Environmental Microbiology* **7**(5): 631-640.

Prakash, O., Nimonkar, Y., and Shouche, Y.S. (2012) 'Practice and prospects of microbial preservation', *FEMS Microbiology Letters* **339**(1): 1-9.

Pu, S., Beaulieu, J.C., Prinyawiwatkul, W., and Ge, B. (2009) 'Effects of plant maturity and growth media bacterial inoculum level on the surface contamination and internalization of *Escherichia coli* O157:H7 in growing spinach leaves', *Journal of Food Protection* **72**(11): 2313-2320.

Py, B., and Barras, F. (2010) 'Building Fe-S proteins: bacterial strategies', *Nature Reviews Microbiology* **8**(6): 436-446.

Rabsch, W., Methner, U., Voigt, W., Tschäpe, H., Reissbrodt, R., and Williams, P.H. (2003) 'Role of receptor proteins for enterobactin and 2,3-dihydroxybenzoylserine in virulence of *Salmonella enterica*', *Infection and Immunity* **71**(12): 6953-6961.

Rabsch, W., and Winkelmann, G. (1991) 'The specificity of bacterial siderophore receptors probed by bioassays', *Biology of Metals* **4**(4): 244-250.

Raffatellu, M., George, M.D., Akiyama, Y., Hornsby, M.J., Nuccio, S.P., Paixao, T.A., Butler, B.P., Chu, H., Santos, R.L., Berger, T., Mak, T.W., Tsolis, R.M., Bevins, C.L., Solnick, J.V., Dandekar, S., and Bäumler, A.J. (2009) 'Lipocalin-2 resistance confers an advantage to *Salmonella enterica* serotype typhimurium for growth and survival in the inflamed intestine', *Cell Host & Microbe* **5**(5): 476-486.

Rajilić-Stojanović, M., Smidt, H., and De Vos, W.M. (2007) 'Diversity of the human gastrointestinal tract microbiota revisited', *Environmental Microbiology* **9**(9): 2125-2136.

Rang, C.U., Licht, T.R., Midtvedt, T., Conway, P.L., Chao, L., Krogfelt, K.A., Cohen, P.S., and Molin, S. (1999) 'Estimation of growth rates of *Escherichia coli* BJ4 in streptomycin-treated and previously germfree mice by *in situ* rRNA hybridization', *Clinical and Diagnostic Laboratory Immunology* **6**(3): 434-436.

Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., Henderson, I.R., Sperandio, V., and Ravel, J. (2008) 'The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates', *Journal of Bacteriology* **190**(20): 6881-6893.

Ratledge, C., and Dover, L.G. (2000) 'Iron metabolism in pathogenic bacteria', *Annual Review of Microbiology* **54**(1): 881-941.

Raymond, K.N., Dertz, E.A., and Kim, S.S. (2003) 'Enterobactin: an archetype for microbial iron transport', *Proceedings of the National Academy of Sciences of the United States of America* **100**(7): 3584-3588.

Redfield, R.J. (1993) 'Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation', *Journal of Heredity* **84**(5): 400-404.

Redfield, R.J. (2001) 'Do bacteria have sex?' *Nature Reviews Genetics* **2**(8): 634-639.

Reid, R.K., Reid, C.P., Powell, P.E., and Szaniszlo, P.J. (1984) 'Comparison of siderophore concentrations in aqueous extracts of rhizosphere and adjacent bulk soils', *Pedobiologia* **26**(4): 263-266.

Reid, S.D., Herbelin, C.J., Bumbaugh, A.C., Selander, R.K., and Whittam, T.S. (2000) 'Parallel evolution of virulence in pathogenic *Escherichia coli*', *Nature* **406**(6791): 64-67.

Reigstad, C.S., Hultgren, S.J., and Gordon, J.I. (2007) 'Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled', *Journal of Biological Chemistry* **282**(29): 21259-21267.

Rendón, M.A., Saldaña, Z., Erdem, A.L., Monteiro-Neto, V., Vázquez, A., Kaper, J.B., Puente, J.L., and Girón, J.A. (2007) 'Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization', *Proceedings of the National Academy of Sciences of the United States of America* **104**(25): 10637-10642.

Rezzonico, E., Mestdagh, R., Delley, M., Combremont, S., Dumas, M.E., Holmes, E., Nicholson, J., and Bibiloni, R. (2011) 'Bacterial adaptation to the gut environment favours successful colonization: microbial and metabonomic characterization of a simplified microbiota mouse model', *Gut Microbes* **2**(6): 307-318.

Richard, H., and Foster, J.W. (2004) '*Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential', *Journal of Bacteriology* **186**(18): 6032-6041.

Ridlon, J.M., Kang, D.J., and Hylemon, P.B. (2006) 'Bile salt biotransformations by human intestinal bacteria', *Journal of Lipid Research* **47**(2): 241-259.

Rigottier-Gois, L. (2013) 'Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis', *The ISME Journal* **7**(7): 1256-1261.

Rosenberg, E.Y., Bertenthal, D., Nilles, M.L., Bertrand, K.P., and Nikaido, H. (2003) 'Bile salts and fatty acids induce the expression of *Escherichia coli* AcrAB multidrug efflux pump through their interaction with Rob regulatory protein', *Molecular Microbiology* **48**(6): 1609-1619.

Ross-Gillespie, A., Dumas, Z., and Kümmerli, R. (2015) 'Evolutionary dynamics of interlinked public goods traits: an experimental study of siderophore production in *Pseudomonas aeruginosa* **28**(1): 29-39.

Rossez, Y., Holmes, A., Wolfson, E.B., Gally, D.L., Mahajan, A., Pedersen, H.L., Willats, W.G., Toth, I.K., and Holden, N.J. (2014) 'Flagella interact with ionic plant lipids to mediate adherence of pathogenic *Escherichia coli* to fresh produce plants', *Environmental Microbiology* **16**(7): 1462-2920.

Rouf, M.A. (1964) 'Spectrochemical analysis of inorganic elements in bacteria', *Journal of Bacteriology* **88**(6): 1545-1549.

Rozen, D.E., de Visser, J.A., and Gerrish, P.J. (2002) 'Fitness effects of fixed beneficial mutations in microbial populations', *Current Biology* **12**(12): 1040-1045.

Russo, T.A., and Johnson, J.R. (2003) 'Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem', *Microbes and Infection* **5**(5): 449-456.

Russo, T.A., McFadden, C.D., Carlino-MacDonald, U.B., Beanan, J.M., Barnard, T.J., and Johnson, J.R. (2002) 'IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*', *Infection and Immunity* **70**(12): 7156-7160.

Sabine, D.B., and Vaseleko, J. (1967) 'Trace element requirements of *Lactobacillus acidophilus*', *Nature* **214**(5087): 520.

Sabri, M., Caza, M., Proulx, J., Lymberopoulos, M.H., Brée, A., Moulin-Schouleur, M., Curtiss, R., and Dozois, C.M. (2008) 'Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic *Escherichia coli* O78 strain χ7122', *Infection and Immunity* **76**(2): 601-611.

Saldeña, T.A., Saraví, F.D., Hwang, H.J., Cincunegui, L.M., and Carra, G.E. (2000) 'Oxygen diffusive barriers of rat distal colon: role of subepithelial tissue, mucosa, and mucus gel layer', *Digestive Diseases and Sciences* **45**(11): 2108-2114.

Sankar, T.S., Neelakanta, G., Sangal, V., Plum, G., Achtman, M., and Schnetz, K. (2009) 'Fate of the H-NS-repressed *bgl* operon in evolution of *Escherichia coli*', *PLoS Genetics* **5**(3): e1000405.

Savageau, M.A. (1974) 'Genetic regulatory mechanisms and the ecological niche of *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **71**(6): 2453-2455.

Savageau, M.A. (1983) '*Escherichia coli* habitats, cell types, and molecular mechanisms of gene control', *The American Naturalist* **122**(6): 732-744.

Schierack, P., Römer, A., Jores, J., Kaspar, H., Guenther, S., Filter, M., Eichberg, J., and Wieler, L.H. (2009) 'Isolation and characterisation of intestinal *Escherichia coli* clones from wild boars in Germany', *Applied and Environmental Microbiology* **75**(3): 695-702.

Schroeder, G.N., and Hilbi, H. (2008) 'Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion', *Clinical Microbiology Reviews* **21**(1): 134-156.

Schubert, S., Darlu, P., Clermont, O., Wieser, A., Magistro, G., Hoffmann, C., Weinert, K., Tenaillon, O., Matic, I., and Denamur, E. (2009) 'Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species', *PLoS Pathogens* **5**(1): e1000257.

Schubert, S., Picard, B., Gouriou, S., Heesemann, J., and Denamur, E. (2002) '*Yersinia* high-pathogenicity island contributes to virulence in *Escherichia*

coli causing extraintestinal infections', *Infection and Immunity* **70**(9): 5335-5337.

Schwartz, C.J., Giel, J.L., Patschkowski, T., Luther, C., Ruzicka, F.J., Beinert, H., and Kiley, P.J. (2001) 'IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins', *Proceedings of the National Academy of Sciences of the United States of America* **98**(26): 14895-14900.

Searle, L.J., Méric, G., Porcelli, I., Sheppard, S.K., and Lucchini, S. (2015) 'Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *Escherichia coli*', *PLoS One* **10**(3): e117906.

Sears, H.J., Brownlee, I., and Uchiyama, J.K. (1950) 'Persistence of individual strains of *Escherichia coli* in the intestinal tract of man', *Journal of Bacteriology* **59**(2): 293-301.

Selander, R., and Levin, B. (1980) 'Genetic diversity and structure in *Escherichia coli* populations', *Science* **210**(4469): 545-547.

Selander, R., Korhonen, T.K., Väisänen-Rhen, V., Williams, P.H., Pattison, P.E., and Caugant, D.A. (1986) 'Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis', *Infection and Immunity* **52**(1): 213-222.

Seo, S.W., Kim, D., Latif, H., O'Brien, E.J., Szubin, R., and Palsson, B.O. (2014) 'Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*', *Nature Communications* **5**: 4910.

Serres, M.H., Kerr, A.R., McCormack, T.J., and Riley, M. (2009) 'Evolution by leaps: gene duplication in bacteria', *Biology Direct* **4**: 46.

Servin, A.L. (2005) 'Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*', *Clinical Microbiology Reviews* **18**(2): 264-292.

Shaw, R.K., Berger, C.N., Feys, B., Knutton, S., Pallen, M.J., and Frankel, G. (2008) 'Enterohemorrhagic *Escherichia coli* Exploits EspA filaments for attachment to salad leaves', *Applied and Environmental Microbiology* **74**(9): 2908-2914.

Shaw, R.K., Berger, C.N., Pallen, M.J., Sjoling, A., and Frankel, G. (2011) 'Flagella mediate attachment of enterotoxigenic *Escherichia coli* to fresh salad leaves', *Environmental Microbiology Reports* **3**(1): 112-117.

Shen, P., and Huang, H.V. (1986) 'Homologous recombination in *Escherichia coli*: dependence on substrate length and homology', *Genetics* **112**(3): 441-457.

Shin, R., Suzuki, M., and Morishita, Y. (2002) 'Influence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic *Escherichia coli* O157:H7', *Journal of Medical Microbiology* **51**(3): 201-206.

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

Singer, I.I., Kawka, D.W., Scott, S., Weidner, J.R., Mumford, R.A., Riehl, T.E., and Stenson, W.F. (1996) 'Expression of inducible nitric oxide synthase and nitrotyrosine in colonic epithelium in inflammatory bowel disease', *Gastroenterology* **111**(4): 871-885.

Singh, V., Yeoh, B.S., Xiao, X., Kumar, M., Bachman, M., Borregaard, N., Joe, B., and Vijay-Kumar, M. (2015) 'Interplay between enterobactin, myeloperoxidase and lipocalin 2 regulates *E. coli* survival in the inflamed gut', *Nature Communications* **6**: 7113.

Sinha, S., and Redfield, R.J. (2012) 'Natural DNA uptake by *Escherichia coli*', *PLoS One* **7**(4): e35620.

Sjogren, R. (1995) 'Thirteen-year survival study of an environmental *Escherichia coli* in field mini-plots', *Water, Air, and Soil Pollution* **81**(3): 315-335.

Skippington, E., and Ragan, M.A. (2011) 'Lateral genetic transfer and the construction of genetic exchange communities', *FEMS Microbiology Reviews* **35**(5): 707-735.

Small, P., Blankenhorn, D., Welty, D., Zinser, E., and Slonczewski, J.L. (1994) 'Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of *rpoS* and growth pH', *Journal of Bacteriology* **176**(6): 1729-1737.

Smati, M., Clermont, O., Le Gal, F., Schichmanoff, O., Jauréguy, F., Eddi, A., Denamur, E., and Picard, B. (2013) 'Real-time PCR for quantitative analysis of human commensal *Escherichia coli* populations reveals a high frequency of sub-dominant phylogroups', *Applied and Environmental Microbiology* **79**(16): 5005-5012.

Smith, E.E., Buckley, D.G., Wu, Z., Saenphimmachak, C., Hoffman, L.R., D'Argenio, D.A., Miller, S.I., Ramsey, B.W., Speert, D.P., Moskowitz, S.M., Burns, J.L., Kaul, R., and Olson, M.V. (2006) 'Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients', *Proceedings of the National Academy of Sciences of the United States of America* **103**(22): 8487-8492.

Smith, J.L., Fratamico, P.M., and Gunther, N.W. (2007) 'Extraintestinal pathogenic *Escherichia coli*', *Foodborne Pathogens and Disease* **4**(2): 134-163.

Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993) 'How clonal are bacteria?' *Proceedings of the National Academy of Sciences of the United States of America* **90**(10): 4384-4388.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000) 'Sources of *Escherichia coli* in a coastal subtropical environment', *Applied and Environmental Microbiology* **66**(1): 230-237.

Solomon, E.B., Pang, H.J., and Matthews, K.R. (2003) 'Persistence of *Escherichia coli* O157:H7 on lettuce plants following spray irrigation with contaminated water', *Journal of Food Protection* **66**(12): 2198-2202.

Solomon, E.B., Yaron, S., and Matthews, K.R. (2002) 'Transmission of *Escherichia coli* O157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization', *Applied and Environmental Microbiology* **68**(1): 397-400.

Spees, A.M., Wangdi, T., Lopez, C.A., Kingsbury, D.D., Xavier, M.N., Winter, S.E., Tsolis, R.M., and Bäumler, A.J. (2013) 'Streptomycin-induced inflammation enhances *Escherichia coli* gut colonization through nitrate respiration', *mBio* **4**(4): e430-e413.

Spira, B., de Almeida Toledo, R., Maharjan, R.P., and Ferenci, T. (2011) 'The uncertain consequences of transferring bacterial strains between laboratories - *rpoS* instability as an example', *BMC Microbiology* **11**: 248.

Staib, L. and Fuchs, T.M. (2014) 'From food to cell: nutrient exploitation strategies of enteropathogens', *Microbiology* **160**(6): 1020-1039.

Staley, T.E., Jones, E.W., and Corley, L.D. (1969) 'Attachment and penetration of *Escherichia coli* into intestinal epithelium of the ileum in newborn pigs', *American Journal of Pathology* **56**(3): 371-392.

Stecher, B., Hapfelmeier, S., Muller, C., Kremer, M., Stallmach, T., and Hardt, W.D. (2004) 'Flagella and chemotaxis are required for efficient induction of *Salmonella enterica* serovar Typhimurium colitis in streptomycin-pretreated mice', *Infection and Immunity* **72**(7): 4138-4150.

Stecher, B., and Hardt, W.D. (2011) 'Mechanisms controlling pathogen colonization of the gut', *Current Opinion in Microbiology* **14**(1): 82-91.

Stenutz, R., Weintraub, A., and Widmalm, G. (2006) 'The structures of *Escherichia coli* O-polysaccharide antigens', *FEMS Microbiology Reviews* **30**(3): 382-403.

Stins, M.F., Prasadarao, N.V., Ibric, L., Wass, C.A., Luckett, P., and Kim, K.S. (1994) 'Binding characteristics of S fimbriated *Escherichia coli* to isolated brain microvascular endothelial cells', *American Journal of Pathology* **145**(5): 1228-1236.

Stojiljkovic, I., Baumler, A.J., and Hantke, K. (1994) 'Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a Fur titration assay', *Journal of Molecular Biology* **236**(2): 531-545.

Storz, G., and Imlay, J.A. (1999) 'Oxidative stress', *Current Opinion in Microbiology* **2**(2): 188-194.

Stępniewski, W., and Stępniewska, Z. (2009) 'Selected oxygen-dependent process – response to soil management and tillage', *Soil and Tillage Research* **102**(2): 193-200.

Strober, W. (2011) 'Adherent-invasive *E. coli* in Crohn disease: bacterial "agent provocateur"', *Journal of Clinical Investigation* **121**(3): 841-844.

Stülke, J., and Hillen, W. (1999) 'Carbon catabolite repression in bacteria', *Current Opinion in Microbiology* **2**(2): 195-201.

Sun, S., Ke, R., Hughes, D., Nilsson, M., and Andersson, D.I. (2012) 'Genome-wide detection of spontaneous chromosomal rearrangements in bacteria', *PLoS One* **7**(8): e42639.

Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2010) 'The siderophore pyoverdine of *Pseudomonas syringae* pv. tabaci 6605 is an intrinsic virulence factor in host tobacco infection', *Journal of Bacteriology* **192**(1): 117-126.

Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A., and Andersson, S.G. (2002) '50 million years of genomic stasis in endosymbiotic bacteria', *Science* **296**(5577): 2376-2379.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010) 'The population genetics of commensal *Escherichia coli*', *Nature Reviews Microbiology* **8**(3): 207-217.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008) 'Comparative genomics: the bacterial pan-genome', *Current Opinion in Microbiology* **11**(5): 472-477.

Texier, S., Prigent-Combaret, C., Gourdon, M.H., Poirier, M.A., Faivre, P., Dorioz, J.M., Poulenard, J., Jocteur-Monrozier, L., Moenne-Loccoz, Y., and Trevisan, D. (2008) 'Persistence of culturable *Escherichia coli* faecal contaminants in dairy alpine grassland soils', *Journal of Environmental Quality* **37**(6): 2299-2310.

Thanassi, D.G., Cheng, L.W., and Nikaido, H. (1997) 'Active efflux of bile salts by *Escherichia coli*', *Journal of Bacteriology* **179**(8): 2512-2518.

Thomas, C.M., and Nielsen, K.M. (2005) 'Mechanisms of, and barriers to, horizontal gene transfer between bacteria', *Nature Reviews Microbiology* **3**(9): 711-721.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research* **22**(22): 4673-4680.

Tiedje, J.M., Asuming-Brempong, S., Nüsslein, K., Marsh, T.L., and Flynn, S.J. (1999) 'Opening the black box of soil microbial diversity', *Applied Soil Ecology* **13**(2): 109-122.

Tiso, M., and Schechter, A.N. (2015) 'Nitrate reduction to nitrite, nitric oxide and ammonia by gut bacteria under physiological conditions', *PLoS One* **10**(3): e119712.

Tobe, T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., and Pallen, M.J. (2006) 'An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination', *Proceedings of the National Academy of Sciences of the United States of America* **103**(40): 14941-14946.

Torres, A.G., Jeter, C., Langley, W., and Matthysse, A.G. (2005) 'Differential binding of *Escherichia coli* O157:H7 to alfalfa, human epithelial cells, and plastic is mediated by a variety of surface structures', *Applied and Environmental Microbiology* **71**(12): 8008-8015.

Torres, A.G., Redford, P., Welch, R.A., and Payne, S.M. (2001) 'TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and TonB are required for virulence in the mouse', *Infection and Immunity* **69**(10): 6179-6185.

Toshima, H., Yoshimura, A., Arikawa, K., Hidaka, A., Ogasawara, J., Hase, A., Masaki, H., and Nishikawa, Y. (2007) 'Enhancement of Shiga toxin production in enterohemorrhagic *Escherichia coli* serotype O157:H7 by DNase colicins', *Applied and Environmental Microbiology* **73**(23): 7582-7588.

Toth, I.K., Pritchard, L., and Birch, P.R. (2006) 'Comparative genomics reveals what makes an enterobacterial plant pathogen', *Annual Review of Phytopathology* **44**: 305-336.

Touati, D., Jacques, M., Tardat, B., Bouchard, L., and Despied, S. (1995) 'Lethal oxidative damage and mutagenesis are generated by iron in *Δfur* mutants of *Escherichia coli*: protective role of superoxide dismutase', *Journal of Bacteriology* **177**(9): 2305-2314.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le Bouguénec, C., Lescat, M., Mangenot, S., Martinez-Jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E.P., and Denamur, E. (2009) 'Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths', *PLoS Genetics* **5**(1): e1000344.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) 'The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes', *Genome Biology* **15**(11): 524.

Tullus, K., Kuhn, I., Orskov, I., Orskov, F., and Mollby, R. (1992) 'The importance of P and type 1 fimbriae for the persistence of *Escherichia coli* in the human gut', *Epidemiology and Infection* **108**(3): 415-421.

Tuomola, E.M., Ouwehand, A.C., and Salminen, S.J. (1999) 'The effect of probiotic bacteria on the adhesion of pathogens to human intestinal mucus', *FEMS Immunology and Medical Microbiology* **26**(2): 137-142.

Turner, S.M., Scott-Tucker, A., Cooper, L.M., and Henderson, I.R. (2006) 'Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*', *FEMS Microbiology Letters* **263**(1): 10-20.

Unno, T., Han, D., Jang, J., Widmer, K., Ko, G., Sadowsky, M.J., and Hur, H.G. (2011) 'Genotypic and phenotypic trends in antibiotic resistant pathogenic *Escherichia coli* isolated from humans and farm animals in South Korea', *Microbes and Environments* **26**(3): 198-204.

Ussery, D.W., Binnewies, T.T., Gouveia-Oliveira, R., Jarmer, H., and Hallin, P.F. (2004) 'Genome update: DNA repeats in bacterial genomes', *Microbiology* **150**(11): 3519-3521.

Valdebenito, M., Bister, B., Reissbrodt, R., Hantke, K., and Winkelmann, G. (2005) 'The detection of salmochelin and yersiniabactin in uropathogenic *Escherichia coli* strains by a novel hydrolysis-fluorescence-detection (HFD) method', *International Journal of Medical Microbiology* **295**(2): 99-107.

Valdebenito, M., Crumbliss, A.L., Winkelmann, G., and Hantke, K. (2006) 'Environmental factors influence the production of enterobactin, salmochelin, aerobactin, and yersiniabactin in *Escherichia coli* strain Nissle 1917', *International Journal of Medical Microbiology* **296**(8): 513-520.

van der Waaij, D., Berghuis, J.M., and Lekkerkerk, J.E. (1972) 'Colonization resistance of the digestive tract of mice during systemic antibiotic treatment', *Epidemiology and Infection* **70**(4): 605-610.

van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011) 'Survival of *Escherichia coli* in the environment: fundamental and public health aspects', *The ISME Journal* **5**(2): 173-183.

van Hoek, A.H., Aarts, H.J., Bouw, E., van Overbeek, W.M., and Franz, E. (2013) 'The role of *rpoS* in *Escherichia coli* O157 manure-amended soil survival and distribution of allelic variations among bovine, food and chemical isolates', *FEMS Microbiology Letters* **338**(1): 18-23.

van Opijnen, T., and Camilli, A. (2010) 'Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms', *Current Protocols in Microbiology* **36**: 1E.3.1-1E.3.24.

van Passel, M.W., Marri, P.R., and Ochman, H. (2008) 'The emergence and fate of horizontally acquired genes in *Escherichia coli*', *PLoS Computational Biology* **4**(4): e1000059.

Vassinova, N., and Kozyrev, D. (2000) 'A method for direct cloning of Fur-regulated genes: identification of seven new Fur-regulated loci in *Escherichia coli*', *Microbiology* **146**(12): 3171-3182.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E., and Gewirtz, A.T. (2010) 'Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5', *Science* **328**(5975): 228-231.

Vital, M., Hammes, F., and Egli, T. (2008) '*Escherichia coli* O157 can grow in natural freshwater at low carbon concentrations', *Environmental Microbiology* **10**(9): 2387-2396.

Vogel-Scheel, J., Alpert, C., Engst, W., Loh, G., and Blaut, M. (2010) 'Requirement of purine and pyrimidine synthesis for colonization of the mouse intestine by *Escherichia coli*', *Applied and Environmental Microbiology* **76**(15): 5181-5187.

Vollmerhausen, T.L., Ramos, N.L., Gundogdu, A., Robinson, W., Brauner, A., and Katouli, M. (2011) 'Population structure and uropathogenic virulence-associated genes of faecal *Escherichia coli* from healthy young and elderly adults', *Journal of Medical Microbiology* **60**(5): 574-581.

Vos, M., and Didelot, X. (2009) 'A comparison of homologous recombination rates in bacteria and archaea', *The ISME Journal* **3**(2): 199-208.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., and Whittam, T.S. (2007) 'Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches', *Environmental Microbiology* **9**(9): 2274-2288.

Wallick, H., and Stuart, C.A. (1943) 'Antigenic relationships of *Escherichia coli* isolated from one individual', *Journal of Bacteriology* **45**(2): 121-126.

Wandersman, C., and Delepelaire, P. (2004) 'Bacterial iron sources: from siderophores to hemophores', *Annual Review of Microbiology* **58**: 611-647.

Warriner, K., Spaniolas, S., Dickinson, M., Wright, C., and Waites, W.M. (2003) 'Internalization of bioluminescent *Escherichia coli* and *Salmonella* Montevideo in growing bean sprouts', *Journal of Applied Microbiology* **95**(4): 719-727.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) 'Identification of novel small RNAs using comparative genomics and microarrays', *Genes & Development* **15**(13): 1637-1651.

Wassenaar, T.M., Beimfohr, C., Geske, T., and Zimmermann, K. (2014) 'Voluntarily exposure to a single, high dose of probiotic *Escherichia coli* results in prolonged colonisation', *Beneficial Microbes* **5**(4): 367-375.

Wassenaar, T.M., and Gunzer, F. (2015) 'The prediction of virulence based on presence of virulence genes in *E. coli* may not always be accurate', *Gut Pathogens* **7**: 15.

Watts, R.E., Totsika, M., Challinor, V.L., Mabbett, A.N., Ulett, G.C., De Voss, J.J., and Schembri, M.A. (2012) 'Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*', *Infection and Immunity* **80**(1): 333-344.

Weaver, V.B., and Kolter, R. (2004) '*Burkholderia* spp. alter *Pseudomonas aeruginosa* physiology through iron sequestration', *Journal of Bacteriology* **186**(8): 2376-2384.

Weinberg, E.D. (1997) 'The *Lactobacillus* anomaly: total iron abstinence', *Perspectives in Biology and Medicine* **40**(4): 578-583.

Weinreich, D.M., Watson, R.A., and Chao, L. (2005) 'Perspective: Sign epistasis and genetic constraint on evolutionary trajectories', *Evolution* **59**(6): 1165-1174.

Welch, R.A., Burland, V., Plunkett, G. 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., and Blattner, F.R. (2002) 'Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **99**(26): 17020-17024.

White, A.P., Sibley, K.A., Sibley, C.D., Wasmuth, J.D., Schaefer, R., Surette, M.G., Edge, T.A., and Neumann, N.F. (2011) 'Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity', *Applied and Environmental Microbiology* **77**(21): 7620-7632.

White, C., Lee, J., Kambe, T., Fritsche, K., and Petris, M.J. (2009) 'A role for the ATP7A copper-transporting ATPase in macrophage bactericidal activity', *Journal of Biological Chemistry* **284**(49): 33949-33956.

Whitman, R.L., and Nevers, M.B. (2003) 'Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach', *Applied and Environmental Microbiology* **69**(9): 5555-5562.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) 'Prokaryotes: The unseen majority', *Proceedings of the National Academy of Sciences of the United States of America* **95**(12): 6578-6583.

Whittam, T.S. (1989) 'Clonal dynamics of *Escherichia coli* in its natural habitat', *Antonie van Leeuwenhoek* **55**(1): 23-32.

Whittam, T.S., Ochman, H., and Selander, R.K. (1983a) 'Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*', *Molecular Biology Evolution* **1**(1): 67-83.

Whittam, T.S., Ochman, H., and Selander, R.K. (1983b) 'Multilocus genetic structure in natural populations of *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **80**(6): 1751-1755.

Wielgoss, S., Barrick, J.E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R.E., and Schneider, D. (2011) 'Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*', *G3: Genes, Genomes, Genetics* **1**(3): 183-186.

Willenbrock, H., Hallin, P.F., Wassenaar T.M., and Ussery, D.W. (2007) 'Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray', *Genome Biology* **8**(12): R267.

Williams, A.P., Avery, L.M., Killham, K., and Jones, D.L. (2007) 'Survival of *Escherichia coli* O157:H7 in the rhizosphere of maize grown in waste-amended soil', *Journal of Applied Microbiology* **102**(2): 319-326.

Williams, P.H., and Carbonetti, N.H. (1986) 'Iron, siderophores, and the pursuit of virulence: independence of the aerobactin and enterochelin iron uptake systems in *Escherichia coli*.' *Infection and Immunity* **51**(3): 942-947.

Wilson, K.H. (1993) 'The microecology of *Clostridium difficile*', *Clinical Infectious Diseases* **16**(Suppl 4): S214-S218.

Winfield, M.D., and Groisman, E.A. (2003) 'Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*', *Applied and Environmental Microbiology* **69**(7): 3687-3694.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., Laughlin, R.C., Gomez, G., Wu, J., Lawhon, S.D., Popova, I.E., Parikh, S.J., Adams, L.G., Tsolis, R.M., Stewart, V.J., and Bäumler, A.J. (2013) 'Host-derived nitrate boosts growth of *E. coli* in the inflamed gut', *Science* **339**(6120): 708-711.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C., Ochman, H., and Achtman, M. (2006) 'Sex and virulence in *Escherichia coli*: an evolutionary perspective', *Molecular Microbiology* **60**(5): 1136-1151.

Wold, A.E., Caugant, D.A., Lidin-Janson, G., de Man, P., and Svanborg, C. (1992) 'Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics', *Journal of Infectious Diseases* **165**(1): 46-52.

Woodmansee, A.N., and Imlay, J.A. (2002) 'Reduced flavins promote oxidative DNA damage in non-respiring *Escherichia coli* by delivering electrons to intracellular free iron', *Journal of Biological Chemistry* **277**(37): 34055-34066.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D., and Lewis, J.D. (2011) 'Linking long-term dietary patterns with gut microbial enterotypes', *Science* **334**(6052): 105-108.

Xicohtencatl-Cortes, J., Sánchez Chacón, E., Saldaña, Z., Freer, E., and Girón, J.A. (2009) 'Interaction of *Escherichia coli* O157:H7 with leafy green produce', *Journal of Food Protection* **72**(7): 1531-1537.

Xu, Q., Schlabach, M.R., Hannon, G.J., and Elledge, S.J. (2009) 'Design of 240,000 orthogonal 25mer DNA barcode probes', *Proceedings of the National Academy of Sciences of the United States of America* **106**(7): 2289-2294.

Yamamoto, S., Tsukamoto, T., Terai, A., Kurazono, H., Takeda, Y., and Yoshida, O. (1997) 'Genetic evidence supporting the faecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*', *Journal of Urology* **157**(3): 1127-1129.

Yancey, R.J., Breeding, S.A., and Lankford, C.E. (1979) 'Enterochelin (enterobactin): virulence factor for *Salmonella typhimurium*', *Infection and Immunity* **24**(1): 174-180.

Yaron, S., and Romling, U. (2014) 'Biofilm formation by enteric pathogens and its role in plant colonization and persistence', *Microbial Biotechnology* **7**(6): 496-516.

Youssif, S. (2001) 'Recent trends in the chemistry of pyridine N-oxide', *Archive for Organic Chemistry* **1**: 242-268

Zambrano, M., Siegele, D., Almiron, M., Tormo, A., and Kolter, R. (1993) 'Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures', *Science* **259**(5102): 1757-1760.

Zdziarski, J., Svanborg, C., Wullt, B., Hacker, J., and Dobrindt, U. (2008) 'Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation?', *Infection and Immunity* **76**(2): 695-703.

Zhang, L., Foxman, B., and Marrs, C. (2002) 'Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2', *Journal of Clinical Microbiology* **40**(11): 3951-3955.

Zhang, Z., Gosset, G., Barabote, R., Gonzalez, C.S., Cuevas, W.A., and Saier M.H. (2005) 'Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in *Escherichia coli*', *Journal of Bacteriology* **187**(3): 980-990.

Zhao, K., Liu, M., and Burgess, R.R. (2007) 'Adaptation in bacterial flagellar and motility systems: from regulon members to 'foraging'-like behavior in *E. coli*', *Nucleic Acids Research* **35**(13): 4441-4452.

Zheng, M., Doan, B., Schneider, T.D., and Storz, G. (1999) 'OxyR and SoxRS regulation of *fur*', *Journal of Bacteriology* **181**(15): 4639-4643.

Zhu, H., Hart, C.A., Sales, D., and Roberts, N.B. (2006) 'Bacterial killing in gastric juice – effect of pH and pepsin on *Escherichia coli* and *Helicobacter pylori*', *Journal of Medical Microbiology* **55**(9): 1265-1270.

Zhu, M., Valdebenito, M., Winkelmann, G., and Hantke, K. (2005) 'Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization', *Microbiology* **151**(7): 2363-2372.

Zhu, X.Y., Zhong, T., Pandya, Y., and Joerger, R.D. (2002) '16S rRNA-based analysis of microbiota from the cecum of broiler chickens', *Applied and Environmental Microbiology* **68**(1): 124-137.

Zinser, E.R., and Kolter, R. (2000) 'Prolonged stationary-phase incubation selects for *lrp* mutations in *Escherichia coli* K-12', *Journal of Bacteriology* **182**(15): 4361-4365.

Zoetendal, E.G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A.D., and De Vos, W.M. (2002) 'Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from faeces', *Applied and Environmental Microbiology* **68**(7): 3401-3407.

Appendices

Appendix A: ECOR and GMB isolates

Table A.1: List of all ECOR and GMB strains. 72 ECOR and 96 GMB isolates were used throughout this study. ECOR isolates (Ochman and Selander, 1984) were collected from healthy host faecal samples except 11 strains isolated from human urine. GMB isolates (Méric et al., 2013) were predominantly taken from the aerial parts of plants except four strains collected from soil in a rocket field. *Animal host was located in a zoo rather than wild. [†]Mizuna is the common name for the plant species *Brassica rapa nipposinica* and *Brassica juncea* var. *japonica*. [‡]Mixed plant sources were from mixed salad leaves post-harvest. [#]Plants listed as other include tatsoi (*Brassica narinosa*), amaranth leaves (*Amaranthus* spp.), red chard (*Beta vulgaris*) and watercress (*Nasturtium officinale*). Strains are labelled to indicate which experiments they were used in as follows: [a]*in vitro* barcode competition study (section 3.4), [b]mouse competition study 1 (section 3.5.1), [c]mouse competition study 2 (section 3.5.2), [d]soil competition study (section 3.6), [e]siderophore gene expression (sections 5.2 and 5.3), [f]siderophore mutant experiments (sections 5.4-5.6). All strains listed were used in multiplex PCRs to determine siderophore gene presence.

Strain	Phylogeny	Source	
		Plant/Host	Location
ECOR01	A	Human	Iowa (USA)
ECOR02	A	Human	New York (USA)
ECOR03	A	Dog (*Canis lupus*)	Massachusetts (USA)
ECOR04	A	Human	Iowa (USA)
ECOR05	A	Human	Iowa (USA)
ECOR06	A	Human	Iowa (USA)
ECOR07	A	Orangutan (*Pongo* sp.)	Washington (USA)*
ECOR08	A	Human	Iowa (USA)
ECOR09	A	Human	Sweden
ECOR10	A	Human	New York (USA)
ECOR11	A	Human (urine)	Sweden
ECOR12	A	Human	Sweden
ECOR13	A	Human	Sweden
ECOR14	A	Human (urine)	Sweden
ECOR15	A	Human	Sweden
ECOR16[ab]	A	Leopard (*Panthera pardus*)	Washington (USA)*
ECOR17	A	Pig (*Sus* sp.)	Indonesia
ECOR18	A	Celebese ape (*Macaca nigra*)	Washington (USA)*
ECOR19	A	Celebese ape (*Macaca nigra*)	Washington (USA)*
ECOR20	A	Steer (*Bos taurus*)	Bali
ECOR21	A	Steer (*Bos taurus*)	Bali
ECOR22	A	Steer (*Bos taurus*)	Bali
ECOR23	A	Elephant (*Elephas maximus*)	Washington (USA)*

ECOR24	A	Human	Sweden
ECOR25	A	Dog (*Canis lupus*)	New York (USA)
ECOR26	B1	Human	Massachusetts (USA)
ECOR27	B1	Giraffe (*Giraffa camelopardalis*)	Washington (USA)*
ECOR28	B1	Human	Iowa (USA)
ECOR29	B1	Kangaroo rat (*Dipodomys* sp.)	Nevada (USA)
ECOR30	B1	Bison (*Bison* sp.)	Alberta (Canada)
ECOR31	E	Leopard (*Panthera pardus*)	Washington (USA)*
ECOR32	B1	Giraffe (*Giraffa camelopardalis*)	Washington (USA)*
ECOR33	B1	Sheep (*Ovis* sp.)	California (USA)
ECOR34	B1	Dog (*Canis lupus*)	Massachusetts (USA)
ECOR35	F	Human	Iowa (USA)
ECOR36	F	Human	Sweden
ECOR37	E	Marmoset (*Cebuella pygmaea*)	Washington (USA)*
ECOR38	F	Human	Iowa (USA)
ECOR39	F	Human	Sweden
ECOR40	F	Human (urine)	Sweden
ECOR41	F	Human	Tonga
ECOR42	E	Human	Massachusetts (USA)
ECOR43	E	Human	Sweden
ECOR44	D	Cougar (*Puma concolor*)	Washington (USA)*
ECOR45	B1	Pig (*Sus* sp.)	Indonesia
ECOR46	D	Celebese ape (*Macaca nigra*)	Washington (USA)*
ECOR47	D	Sheep (*Ovis* sp.)	Papua New Guinea
ECOR48	D	Human (urine)	Sweden
ECOR49[a]	D	Human	Sweden
ECOR50	D	Human (urine)	Sweden
ECOR51	B2	Human	Massachusetts (USA)
ECOR52	B2	Orangutan (*Pongo* sp.)	Washington (USA)*
ECOR53	B2	Human	Iowa (USA)
ECOR54	B2	Human	Iowa (USA)
ECOR55[ab]	B2	Human	Sweden
ECOR56	B2	Human (urine)	Sweden
ECOR57	B2	Gorilla (*Gorilla gorilla*)	Washington (USA)*
ECOR58	B1	Lion (*Panthera leo*)	Washington (USA)*
ECOR59	B2	Human	Massachusetts (USA)
ECOR60	B2	Human (urine)	Sweden
ECOR61	B2	Human	Sweden
ECOR62	B2	Human (urine)	Sweden
ECOR63	B2	Human	Sweden
ECOR64	B2	Human (urine)	Sweden
ECOR65	B2	Celebese ape (*Macaca nigra*)	Washington (USA)*
ECOR66	B2	Celebese ape (*Macaca nigra*)	Washington (USA)*

ECOR67	B1	Goat (*Capra aegagrus*)	Indonesia
ECOR68	B1	Giraffe (*Giraffa camelopardalis*)	Washington (USA)*
ECOR69	B1	Celebese ape (*Macaca nigra*)	Washington (USA)*
ECOR70	A	Gorilla (*Gorilla gorilla*)	Washington (USA)*
ECOR71	A	Human (urine)	Sweden
ECOR72	A	Human (urine)	Sweden
GMB01	A	Rocket (*Eruca sativa*)	Italy
GMB02[c]	A	Rocket (*Eruca sativa*)	Italy
GMB03	B1	Rocket *(Eruca sativa)*	King's Lynn (UK)
GMB04	D	Rocket (*Eruca sativa*)	King's Lynn (UK)
GMB05	B1	Mizuna (*Brassica* spp.)[†]	King's Lynn (UK)
GMB06	A	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB07[abcd]	B1	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB10	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB100[e]	B2	Spinach (*Spinacia oleracea*)	Mixed
GMB101	B2	Mixed[‡]	Mixed
GMB102	B2	Rocket (*Eruca sativa*)	Dorset (UK)
GMB103	B1	Other[#]	Other
GMB104[cdef]	A	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB105	B1	Spinach (*Spinacia oleracea*)	Italy
GMB106	A	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB107	A	Mixed[‡]	Mixed
GMB108	A	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB13	B1	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB14	B1	Mizuna (*Brassica* spp.)[†]	King's Lynn (UK)
GMB15	D	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB16	B1	Mizuna (*Brassica* spp.)[†]	Martham (UK)
GMB17	B1	Mizuna (*Brassica* spp.)[†]	Martham (UK)
GMB18[cd]	B1	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB19	D	Other[#]	Dorset (UK)
GMB20	A	Rocket (*Eruca sativa*)	Martham (UK)
GMB21	D	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB22	B1	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB23[cde]	A	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB24	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB25	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB26	B1	Other [#]	Dorset (UK)
GMB27	A	Mizuna (*Brassica* spp.)[†]	Martham (UK)
GMB28	B1	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB29	B1	Mixed[‡]	Martham (UK)
GMB30[e]	D	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB31	A	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB32[c]	A	Spinach (*Spinacia oleracea*)	Dorset (UK)

GMB33	A	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB34[abcd]	A	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB35	A	Other [#]	Martham (UK)
GMB36	B1	Mizuna (*Brassica* spp.)[†]	Dorset (UK)
GMB37	A	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB38	B1	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB39	E	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB40[de]	B2	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB41	B1	Other [#]	Dorset (UK)
GMB43	B1	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB44	D	Mixed[‡]	Mixed
GMB45[abcd]	B2	Spinach (*Spinacia oleracea*)	Mixed
GMB47	B2	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB48	A	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB49	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB50	D	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB51	D	Spinach (*Spinacia oleracea*)	Berkshire (UK)
GMB52	D	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB53[e]	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB54[cd]	D	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB58	B1	Other [#]	Mixed
GMB59	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB60	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB61	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB63	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB64	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB65	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB66	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB67	A	Rocket (*Eruca sativa*)	Martham (UK)
GMB68	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB69	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB70	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB71[abcd]	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB72[cd]	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB73	B2	Rocket (*Eruca sativa*)	Martham (UK)
GMB74	D	Rocket (*Eruca sativa*)	Martham (UK)
GMB76	B2	Rocket (*Eruca sativa*)	Martham (UK)
GMB77	E	Rocket (*Eruca sativa*)	Martham (UK)
GMB78	E	Rocket (*Eruca sativa*)	Martham (UK)
GMB79	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB80	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB81	B1	Rocket (*Eruca sativa*)	Martham (UK)
GMB83	B1	Soil	Martham (UK)

GMB84	A	Soil	Martham (UK)
GMB85	B1	Soil	Martham (UK)
GMB86	B1	Soil	Martham (UK)
GMB87	D	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB88[e]	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB89	B1	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB90	A	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB91[ef]	A	Spinach (*Spinacia oleracea*)	Mixed
GMB92	B1	Rocket (*Eruca sativa*)	Dorset (UK)
GMB93	B2	Spinach (*Spinacia oleracea*)	Martham (UK)
GMB94	B2	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB95	A	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB96	D	Spinach (*Spinacia oleracea*)	Italy
GMB97	D	Spinach (*Spinacia oleracea*)	Dover (UK)
GMB98[c]	B2	Spinach (*Spinacia oleracea*)	Dorset (UK)
GMB99	B1	Rocket (*Eruca sativa*)	Italy

Appendix B: Barcode mutant characterisation

B.1 Barcoded and parental strain E. coli growth curves

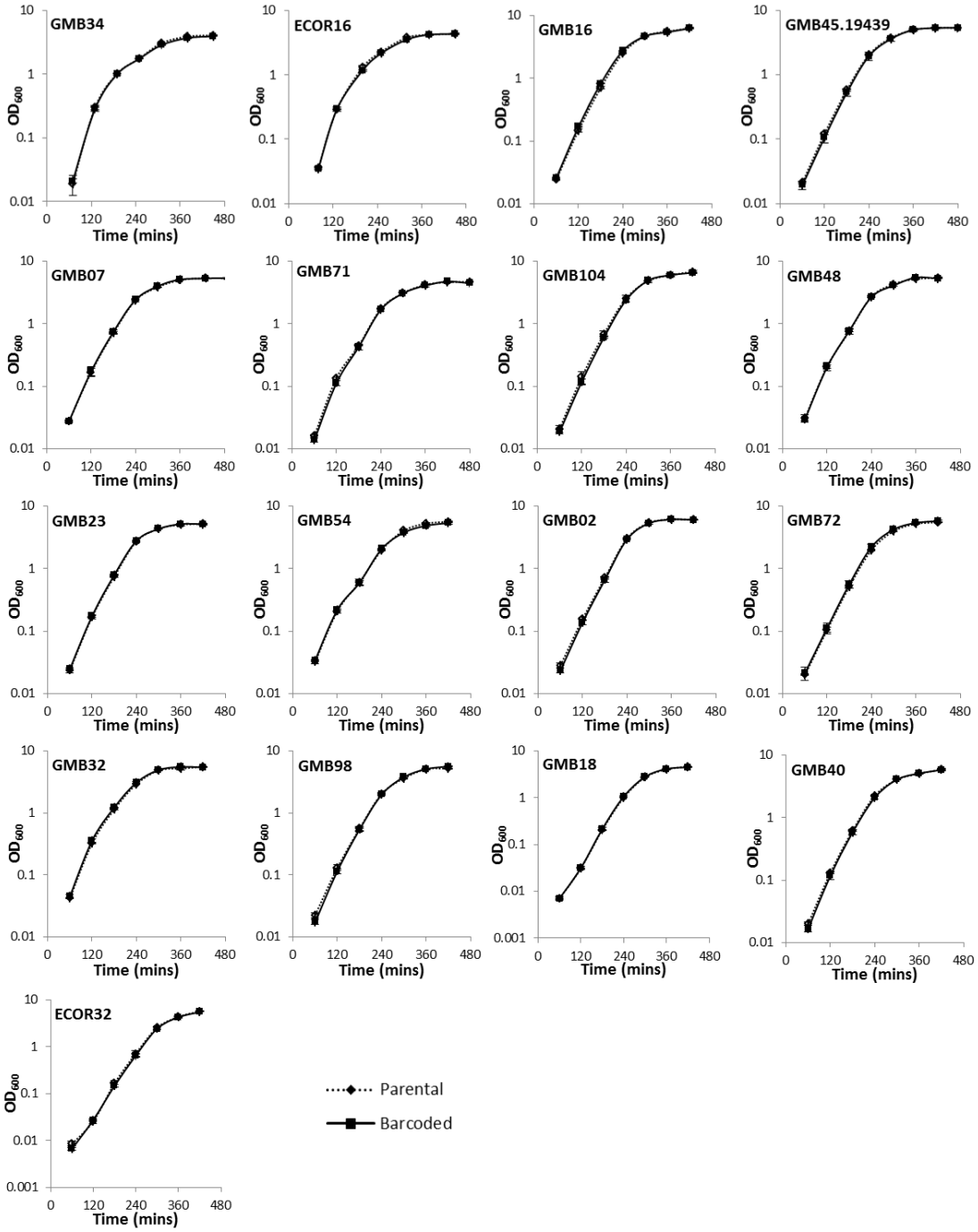


Figure B.1: Growth curves of barcoded and parental strains grown in LB medium at 37°C. All barcoded strains show equivalent growth to their parental strains. Results show the mean ± standard error (*n* = 3).

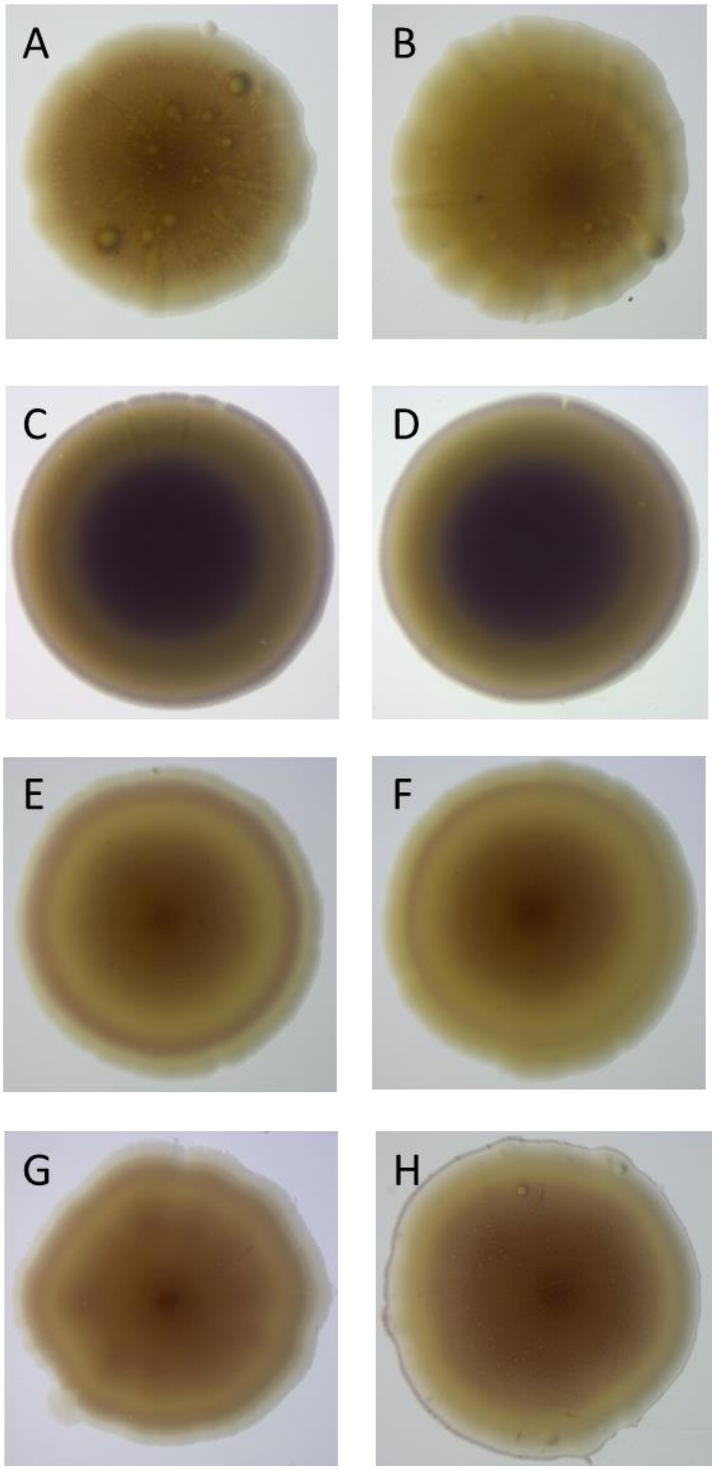
B.2 GN2 BIOLOG plate list of carbon sources

Table B.2: Carbon sources on a GN2 BIOLOG plate.

Well	Carbon Source	Well	Carbon Source
A01	Water (blank)	**E01**	p-Hydroxy Phenylacetic Acid
A02	α-Cyclodextrin	**E02**	Itaconic Acid
A03	Dextrin	**E03**	α-Keto Butyric Acid
A04	Glycogen	**E04**	α-Keto Glutaric Acid
A05	Tween 40	**E05**	α-Keto Valeric Acid
A06	Tween 80	**E06**	D,L-Lactic Acid
A07	N-Acetyl-D-Galactosamine	**E07**	Malonic Acid
A08	N-Acetyl-D-Glucosamine	**E08**	Propionic Acid
A09	Adonitol	**E09**	Quinic Acid
A10	L-Arabinose	**E10**	D-Saccharic Acid
A11	D-Arabitol	**E11**	Sebacic Acid
A12	D-Cellobiose	**E12**	Succinic Acid
B01	i-Erythritol	**F01**	Bromosuccinic Acid
B02	D-Fructose	**F02**	Succinamic Acid
B03	L-Fucose	**F03**	Glucuronamide
B04	D-Galactose	**F04**	L-Alaninamide
B05	Gentiobiose	**F05**	D-Alanine
B06	α-D-Glucose	**F06**	L-Alanine
B07	m-Inositol	**F07**	L-Alanylglycine
B08	α-D-Lactose	**F08**	L-Asparagine
B09	Lactulose	**F09**	L-Aspartic Acid
B10	Maltose	**F10**	L-Glutamic Acid
B11	D-Mannitol	**F11**	Glycyl-L-aspartic Acid
B12	D-Mannose	**F12**	Glycyl-L-Glutamic Acid
C01	D-Melibiose	**G01**	L-Histidine
C02	β-Methyl-D-Glucoside	**G02**	Hydroxy-LProline
C03	D-Psicose	**G03**	L-Leucine
C04	D-Raffinose	**G04**	L-Ornithine
C05	L-Rhamnose	**G05**	LPhenylalanine
C06	D-Sorbitol	**G06**	L-Proline
C07	Sucrose	**G07**	L-Pyroglutamic Acid
C08	D-Trehalose	**G08**	D-Serine
C09	Turanose	**G09**	L-Serine
C10	Xylitol	**G10**	L-Threonine
C11	Pyruvic Acid Methyl Ester	**G11**	D,L-Carnitine
C12	Succinic Acid Mono-Methyl-Ester	**G12**	γ-Amino Butyric Acid
D01	Acetic Acid	**H01**	Urocanic Acid
D02	Cis-Aconitic Acid	**H02**	Inosine
D03	Citric Acid	**H03**	Uridine

D04	Formic Acid	**H04**	Thymidine
D05	D-Galactonic Acid Lactone	**H05**	Phenyethylamine
D06	D-Galacturonic Acid	**H06**	Putrescine
D07	D- Gluconic Acid	**H07**	2-Aminoethanol
D08	D-Glucosaminic Acid	**H08**	2,3-Butanediol
D09	D-Glucuronic Acid	**H09**	Glycerol
D10	α-Hydroxybutyric Acid	**H10**	D,L-α-Glycerol Phosphate
D11	β-Hydroxybutyric Acid	**H11**	α-D-Glucose-1-Phosphate
D12	γ-Hydroxybutyric Acid	**H12**	D-Glucose-6-Phosphate

B.3 Colony morphology on YESCA plates of barcoded and parental strains



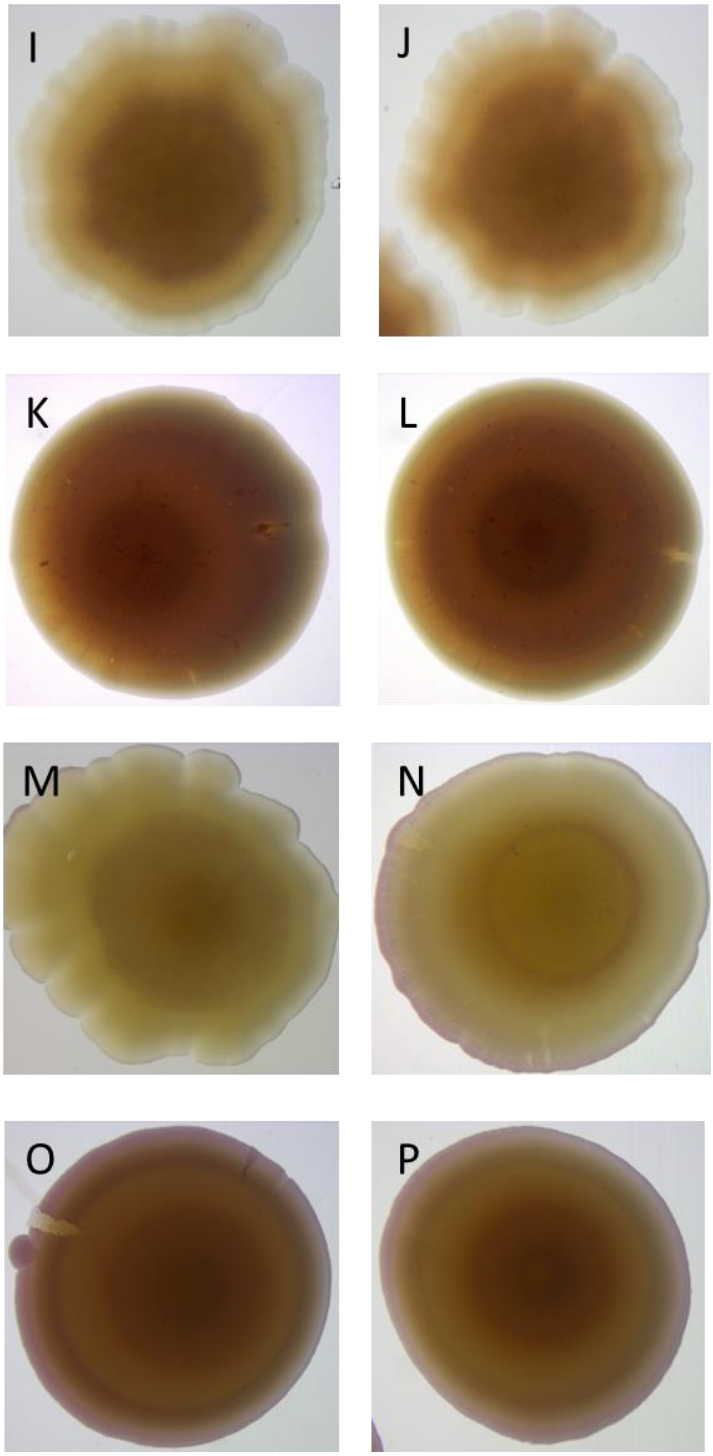


Figure B.3: Colony morphology on YESCA plates of two strains of E. coli and their barcoded versions. (A) GMB18, (B) GMB18-B5, (C) GMB23, (D) GMB23-23277, (E) GMB71, (F) GMB71-B10, (G) GMB34, (H) GMB34-33233, (I) ECOR32, (J) ECOR32-23457, (K) GMB02, (L) GMB02-23796, (M) GMB98, (N) GMB98-30438, (O) GMB104 and (P) GMB104-29884 were all imaged using a Leica M165C Stereo microscope after YESCA plates were incubated at 37°C for 48h followed by 7 days at RT.

Appendix C: Siderophore gene presence tables

Table C.1: Detection of siderophore genes in ECOR isolates by multiplex PCR.
+ = gene detected, - = gene not detected. [a]Result different from array-based study (Jackson et al., 2011). [b]Result different from array-based and PCR-based studies (Johnson, Delavari, Kuskowski and Stell, 2001; Jackson et al., 2011), [c]Result different from array-based study, but supported by PCR-based study (Johnson, Delavari, Kuskowski and Stell, 2001; Jackson et al., 2011), [d]Result different from array-based and PCR-based studies (Schubert et al., 2009; Jackson et al., 2011), [e]Strain isolated from urine sampled from women with urinary tract infections, [#]Independent literature data are not available for *irp3-5*, **irp4* primers amplified a region spanning both *irp4* and *irp5* genes. Enterobactin results are not shown as all five genes tested for were detected in all strains. ECOR35 and ECOR36 showed different results from the array-based study (Jackson et al., 2011) for one enterobactin gene (*entE)*.

		Aerobactin					Salmochelin					Yersiniabactin				
Strain	**Phylogroup**	***iucA***	***iucB***	***iucC***	***iucD***	***iutA***	***iroB***	***iroC***	***iroD***	***iroE***	***iroN***	***irp2***	***irp1***	***irp3#***	***irp4#****	***fyuA***
ECOR01	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR02	A	-[a]	-[a]	-[a]	-[a]	-[b]	-	-	-	-	-	+	+	+	+	+
ECOR03	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR04	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR05	A	+	+	+	+	+	-	-	+	+	-	-	-	-	-	-
ECOR06	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR07	A	+[a]	+	+	+	+[c]	+	+	+	+	+	+	+	+	+	+
ECOR08	A	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
ECOR09	A	-	-	-[a]	-[a]	-[c]	-	-	-	-	-	+	+	+	+	+
ECOR10	A	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR11[e]	A	+	+	+	+	+[c]	-	-	-	-	-	+[d]	+[d]	+	+	+[d]
ECOR12	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR13	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR14[e]	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR15	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR16	A	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR17	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR18	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR19	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR20	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR21	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR22	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR23	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR24	A	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
ECOR25	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR26	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR27	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR28	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR29	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR30	B1	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
ECOR31	E	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR32	B1	-	-	-	-	-	-[a]	-[a]	-[a]	-[a]	-[c]	-	-	-	-	-
ECOR33	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR34	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR35	F	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
ECOR36	F	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
ECOR37	E	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
ECOR38	F	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+

ECOR39	F	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+
ECOR40[e]	F	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+
ECOR41	F	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+
ECOR42	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR43	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR44	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR45	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR46	D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR47	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR48[e]	D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR49	D	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+
ECOR50[e]	D	+	+	+	+	+[c]	-	+	+	+	-	+	+	+	+	+
ECOR51	B2	+	+	+	+	+[c]	+	+	+	+	+	+	+	+	+	+
ECOR52	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR53	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR54	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR55	B2	+	+	+	+	+[c]	-	-	-	-	-	+	+	+	+	+
ECOR56[e]	B2	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
ECOR57	B2	+	+	+	+	+[c]	+	+	+	+	+	+	+	+	+	+
ECOR58	B1	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
ECOR59	B2	-	-	-	-	-	-	-[a]	-[a]	-[a]	-	+	+	+	+	+
ECOR60[e]	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR61	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR62[e]	B2	+[a]	+	+	+	+[c]	+	+	+	+	+	+	+	+	+	+
ECOR63	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR64[e]	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR65	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR66	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
ECOR67	B1	-	-	-	-	-	-[a]	-[a]	-[a]	-[a]	-[b]	-	-	-	-	-
ECOR68	B1	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR69	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ECOR70	A	+[a]	+	+	+	+[c]	+	+	+	+	+	+	+	+	+	+
ECOR71[e]	A	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
ECOR72[e]	A	-	-	-	-	-	-	-	-	-	-	-[d]	-[d]	-	-	-[d]

Table C.2: Detection of siderophore genes in GMB isolates by multiplex PCR. + = gene detected, - = gene not detected. **irp4* primers amplified a region spanning both *irp4* and *irp5* genes. Enterobactin results are not displayed as all genes were detected in all strains.

		Aerobactin					Salmochelin					Yersiniabactin				
Strain	**Phylogroup**	***iucA***	***iucB***	***iucC***	***iucD***	***iutA***	***iroB***	***iroC***	***iroD***	***iroE***	***iroN***	***irp2***	***irp1***	***irp3***	***irp4****	***fyuA***
GMB01	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB02	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB03	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB04	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB05	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB06	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB07	B1	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB10	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB100	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB101	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB102	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB103	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB104	A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GMB105	B1	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
GMB106	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB107	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB108	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB13	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB14	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB15	D	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
GMB16	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB17	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB18	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB19	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB20	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB21	D	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
GMB22	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB23	A	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+
GMB24	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB25	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB26	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB27	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB28	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB29	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB30	D	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
GMB31	A	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB32	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB33	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB34	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB35	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB36	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB37	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB38	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB39	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB40	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB41	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB43	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB44	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB45	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB47	B2	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB48	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

GMB49	B1	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-
GMB50	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB51	D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB52	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB53	B1	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB54	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB58	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB59	B1	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
GMB60	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB61	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB63	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB64	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB65	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB66	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB67	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB68	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB69	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB70	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB71	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB72	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB73	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB74	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB76	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB77	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB78	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB79	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB80	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB81	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB83	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB84	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB85	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB86	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB87	D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB88	B1	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
GMB89	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB90	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB91	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB92	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB93	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB94	B2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB95	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB96	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GMB97	D	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
GMB98	B2	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
GMB99	B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Appendix D: Publication

PLOS | ONE

RESEARCH ARTICLE

Variation in Siderophore Biosynthetic Gene Distribution and Production across Environmental and Faecal Populations of *Escherichia coli*

Laura J. Searle[1☯], Guillaume Méric[1,2☯], Ida Porcelli[1], Samuel K. Sheppard[2,3], Sacha Lucchini[1]*

1 Gut Health and Food Safety, Institute of Food Research, Norwich, United Kingdom, 2 Institute of Life Science, College of Medicine, Swansea University, Swansea, United Kingdom, 3 Department of Zoology, University of Oxford, Oxford, United Kingdom

☯ These authors contributed equally to this work.
* sacha.lucchini@ifr.ac.uk

OPEN ACCESS

Citation: Searle LJ, Méric G, Porcelli I, Sheppard SK, Lucchini S (2015) Variation in Siderophore Biosynthetic Gene Distribution and Production across Environmental and Faecal Populations of *Escherichia coli*. PLoS ONE 10(3): e0117906. doi:10.1371/journal.pone.0117906

Received: August 8, 2014

Accepted: January 5, 2015

Published: March 10, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: SL, IP, GM received funding from Biotechnology and Biological Sciences Research Council (http://www.bbsrc.ac.uk), grant number: BB/J004529/1. LJS received funding from Biotechnology and Biological Sciences Research Council (http://www.bbsrc.ac.uk), DTG Studentship. SKS received funding from Biotechnology and Biological Sciences Research Council (http://www.bbsrc.ac.uk), grant number: BB/I02464X/1; Wellcome Trust (http://www.wellcome.ac.uk), grant number 088786/C/09/Z. The funders had no role in study design, data collection

Abstract

Iron is essential for *Escherichia coli* growth and survival in the host and the external environment, but its availability is generally low due to the poor solubility of its ferric form in aqueous environments and the presence of iron-withholding proteins in the host. Most *E. coli* can increase access to iron by excreting siderophores such as enterobactin, which have a very strong affinity for Fe^{3+}. A smaller proportion of isolates can generate up to 3 additional siderophores linked with pathogenesis; aerobactin, salmochelin, and yersiniabactin. However, non-pathogenic *E. coli* are also able to synthesise these virulence-associated siderophores. This raises questions about their role in the ecology of *E. coli*, beyond virulence, and whether specific siderophores might be linked with persistence in the external environment. Under the assumption that selection favours phenotypes that confer a fitness advantage, we compared siderophore production and gene distribution in *E. coli* isolated either from agricultural plants or the faeces of healthy mammals. This population-level comparison has revealed that under iron limiting growth conditions plant-associated isolates produced lower amounts of siderophores than faecal isolates. Additionally, multiplex PCR showed that environmental isolates were less likely to contain loci associated with aerobactin and yersiniabactin synthesis. Although aerobactin was linked with strong siderophore excretion, a significant difference in production was still observed between plant and faecal isolates when the analysis was restricted to strains only able to synthesise enterobactin. This finding suggests that the regulatory response to iron limitation may be an important trait associated with adaptation to the non-host environment. Our findings are consistent with the hypothesis that the ability to produce multiple siderophores facilitates *E. coli* gut colonisation and plays an important role in *E. coli* commensalism.

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Iron is an essential element implicated in many cellular processes such as DNA replication, energy generation and protection from oxidative stress. When oxygen is present, free ferrous iron (Fe^{2+}) is rapidly oxidised to the insoluble ferric iron (Fe^{3+}). To cope with the reduced bioavailability of ferric iron, bacteria such as *E. coli* have evolved mechanisms to scavenge iron molecules in order to maintain their intracellular iron concentration between 10^{-7} and 10^{-5} M [1]. It has been estimated that in order to survive and multiply in most environments, Gram-negative bacteria require 10^5 to 10^6 Fe^{3+} ions per generation [2]. The main mechanism of ferric iron uptake in *E. coli* involves the synthesis of up to 4 distinct siderophore molecules with very high affinity for iron. After synthesis, siderophores are secreted into the external environment. Siderophore-iron complexes are then imported through specific transporters and degraded to release the iron in the bacterial cytosol [1]. In *E. coli*, siderophores have been described as virulence factors in pathogenic strains and are therefore suggested as targets for antibacterial compounds to limit pathogenic growth [3]. However, their presence in commensal strains highlights the need to re-assess the role of siderophores in the general ecology of *E. coli*, which also includes the external non-host environment, where *E. coli* can persist for several weeks [4]. In particular, vegetables, an increasingly recognised secondary reservoir for *E. coli* [5], might represent an iron limited environment where siderophore-mediated iron uptake is an important determinant of bacterial fitness [6–8]. In agreement with this hypothesis, gene expression of the enterobactin and salmochelin siderophores was induced in *Salmonella* Typhimurium in alfalfa root exudates [9].

The diversity in siderophore production displayed by the *E. coli* species suggests that the distribution of siderophores among a population is shaped by the environmental requirements, as with other traits [10,11]. Taking advantage of our recently described GMB collection, which groups 96 environmental isolates from agricultural plants, mainly grown in the UK [11], we compared siderophore production and distribution in these plant-associated strains with *E. coli* isolated from healthy mammalian hosts. We used this approach to investigate the influence of the environment on siderophore production by *E. coli*.

Materials and Methods

Bacterial strains

The plant-associated *E. coli* strains (GMB collection) used in this study have been described previously [11], and mostly comprise strains isolated between 2008 and 2009 in England from the aerial parts of salad crops such as spinach and rocket (76/96). A minority of strains have been isolated from other crops, salad bags and field soil. The ECOR collection includes 61 isolates obtained from the faeces of healthy mammals and 11 isolates from the urine of women with urinary tract infections [12]. The subset of 61 faecal isolates will be referred to as ECOR-F and comprises 29 human and 32 animal isolates mainly isolated in the USA and Europe in the 1980s. The ECOR collection was kindly provided by Beth Whittam (Michigan State University, USA).

Siderophore production

To assess siderophore production *in vitro*, bacterial colonies were grown on chrome azurol S (CAS) [13] plates for 48h at 37°C with glucose as the carbon source. Colonies of siderophore-producing bacteria grown on this medium are surrounded by a yellow or orange halo. The ability to produce siderophores was quantified by measuring the halo diameter using the ImageJ software (http://rsbweb.nih.gov/ij/index.html). To assess any siderophore-specific differences in diffusion that may influence halo size on CAS agar plates, a liquid CAS assay was performed. Single colonies were grown in modified M9 media with shaking for 16h at 37°C [14]. OD_{600}

measurements were performed before cultures were pelleted and supernatants collected [13]. Supernatant from strains positive for siderophore production cause a colour change from blue to orange when mixed with CAS assay solution. This colour change was measured using OD_{630} and percent siderophore units calculated and standardised to culture OD_{600} [13].

Multiplex PCR

To maximise the sensitivity of the multiplex PCR in detecting siderophore production/receptor genes, ClustalW multiple sequence alignments were performed using publically available *E. coli* sequences from EcoCyc and the National Centre for Biotechnology Information (NCBI). Primers were designed to target conserved regions of 4 biosynthetic and 1 receptor genes for all 4 siderophore systems (Table 1). Primers for the salmochelin PCR targeted export and degradation genes within the salmochelin operon as it only has 1 biosynthesis gene. Product sizes were designed to be of different lengths within one system to ensure that bands could be distinguished on a gel. Template DNA was extracted from overnight cultures using the QIAGEN DNeasy Blood and Tissue extraction kit as per manufacturer's instructions. 10 ng of DNA was used for each multiplex PCR in a 25 µl reaction volume containing 12.5µl Go-Taq Green Master Mix (Promega) and 0.1 mM of each primer. Amplification for each PCR was as follows: 35 cycles at

Table 1. Primers used in this study for multiplex and real-time PCR.

	Gene	Sequence (Forward)	Sequence (Reverse)	Product Size
Primers for multiplex PCR				
Enterobactin	*entA*	GTGCGCTGTAATGTGGTTTC	CAGAGGCGAGGAACAAAATC	184
	entB	GCGACTACTGCAAACAGCAC	TTCAGCGACATCAAATGCTC	382
	entC	GACTCAGGCGATGAAAGAGG	TGCAATCCAAAAACGTTCAA	438
	entE	CGTAGCGTCGAGATTTGTCA	CCCATCAGCTCATCTTCCAT	776
	fepA	TTTGTCGAGGTTGCCATACA	CACGCTGATTTTGATTGACG	349
Salmochelin	*iroB*	CAACCATCGGTTTGACAGTG	GACGTAACACCGCCGAGTAT	166
	iroC	TGCCACACAGGATTTTACCA	CTCACTCTGGGTGCAGCATA	388
	iroD	GGTAAGCAGTTGTCCGGTGT	GTTACTGCGGCTCCTATTCG	227
	iroE	ATCATAACCTCTGCCCAACG	ACCAACCTCCCTTTCGATCT	300
	iroN	CTTCCTCTACCAGCCTGACG	GCTCCGAAGTGATCATCCAT	648
Yersiniabactin	*irp1*	AGAGCGGAAATAACCGAACA	GTAAACAGGCCGTGACGATT	221
	irp2	CTGGTGATGGTGATGGAAAA	CCATCGCGATAAATTGTCCT	247
	irp3	GTATACCTCGCCGGAACAGA	GCCAGCGTTTGTAAGGAACT	177
	irp4&5	GCGCCACAAGGACTGATTAT	GTCTCTCCAGCGACCAGAAC	905
	fyuA	GGGAATGTGAAACTGCGTCT	CGGGTGCCAAGTTCATAGTT	791
Aerobactin	*iucA*	ATAAGGGAAATAGCGCAGCA	TTACGGCTGAAGCGGATTAC	212
	iucB	CCACGAATAGTGACGACCAA	GTTTTTGATGCAGAGCGTGA	339
	iucC	ATTTCGGGAAACGCTTCTTT	GTGGTTCCGCTGTATCACCT	158
	iucD	TCTTCCTTCAGTCCGGAGAA	TCCTCATTTTTCCTGGCATC	630
	iutA	CCAGCCTCAAACTCCATCAT	ACAGCCGACAACTGGACTCT	157
Primers for real-time PCR				
Enterobactin	*entC*	CGAGCGTTTTAGCTCCATTC	CCTCTTTCATCGCCTGAGTC	143
Salmochelin	*iroB*	TATACCGGTCGTGATGCAAA	ATACTCGGCGGTGTTACGTC	150
Yersiniabactin	*irp2*	TAAAACTGAAGCCGGGTCAC	CCGTTGTGTCACCAGAAATG	122
Aerobactin	*iucA*	CTGCCGGTCGGATTTATTTA	ATAAGGGAAATAGCGCAGCA	138
RpoB	*rpoB*	GTGGTGAAACCGCATCTTTT	CGATGTACTCAACCGGGACT	138

dci:10.1371/journal.pone.0117906.t001

95°C for 30s, 55°C for 30s, 72°C for 1 min, and 1 cycle at 72°C for 5 min. Yersiniabactin and aerobactin multiplex PCRs had slight alterations, with the annealing temperature raised to 60°C for the yersiniabactin PCR and elongation step shortened to 40 s for the aerobactin PCR.

RNA extraction

LB overnight cultures were washed with water and diluted to a final OD_{600} of 0.05 in 250 ml flasks containing 25 ml of modified M9 medium without iron addition [14]. The cultures were incubated at 37°C in a shaking incubator until the exponential phase of growth (OD_{600} = 0.2). Synthesis and degradation of RNA were blocked by adding 1/5 volume of stop-solution (90% ethanol/10% phenol) [15]. The RNA was purified using the Promega SV total RNA purification kit according to the manufacturer's instructions. Quantification of the RNA was performed by measuring the 260 nm absorbance on a Nanodrop 1000 spectrophotometer.

Real-time PCR

To determine the level of expression of each siderophore system, 5 μg of total RNA from each sample was reverse-transcribed in 20 μl of buffer containing 0.5 mM dNTPs, 0.2 μg random hexamers and 200 U reverse transcriptase (Fermentas). Target mRNAs were then detected by real-time PCR using SYBR Green JumpStart Taq ReadyMix following the manufacturer's instructions (Sigma). Quantification was performed using the comparative Ct method relative to *rpoB* as an internal standard. The real-time PCR was performed using gene-specific primer pairs designed *in silico* (http://frodo.wi.mit.edu/) to target conserved regions (Table 1). Amplicons were designed to be 100–150 bp in size and their amplification efficiency determined to avoid potential bias. The amplification cycle was as follows: 40 cycles at 94°C for 30s, 60°C for 30s, 72°C for 40s, and 1 cycle at 72°C for 5 min.

Results

Comparison of siderophore production in *E. coli* populations of plant-associated and faecal isolates from healthy mammals

To assess whether environmental adaptation influences the ability of the associated *E. coli* populations to generate siderophores, we compared siderophore production of previously described *E. coli* strains either isolated from plants (GMB strains) or faecal samples of healthy mammals (ECOR-F subset; S1 Table). The isolates were grown on chrome azurol S (CAS) solid medium and the size of the halo resulting from the colour change following the accumulation of the free indicator dye was measured. Differences in growth rate were normalised by dividing the diameter of the halo by the size of the colony (Fig. 1A). No growth and no halo could be detected for 2 environmental strains (GMB37 and GMB69) and 2 faecal isolates (ECOR29 and ECOR52). Growth was not restored by adding ferrous sulphate to a 100μM final concentration to MM9, suggesting the presence of auxotrophic mutations not related to iron acquisition. These 4 strains were therefore excluded from the siderophore production comparison. Plant-associated strains produced significantly less siderophores (unpaired t-test; t = 4.39, $p<0.0001$) than faecal isolates on the CAS indicator agar medium (Fig. 1B). The frequency distributions of siderophore production for GMB and ECOR-F strains highlighted how a larger (37% vs 11%; Fischer's exact test, $p<0.001$) proportion of ECOR-F included high siderophore producers (80^{th} percentile) compared to GMB (Fig. 1C).

Previous epidemiological studies have shown that the *iutA*, *iroN* and *fyuA* siderophore receptor genes are more frequently associated with strains belonging to particular phylogenetic groups, with *E. coli* isolates belonging to phylogroup B2 generally possessing a greater number

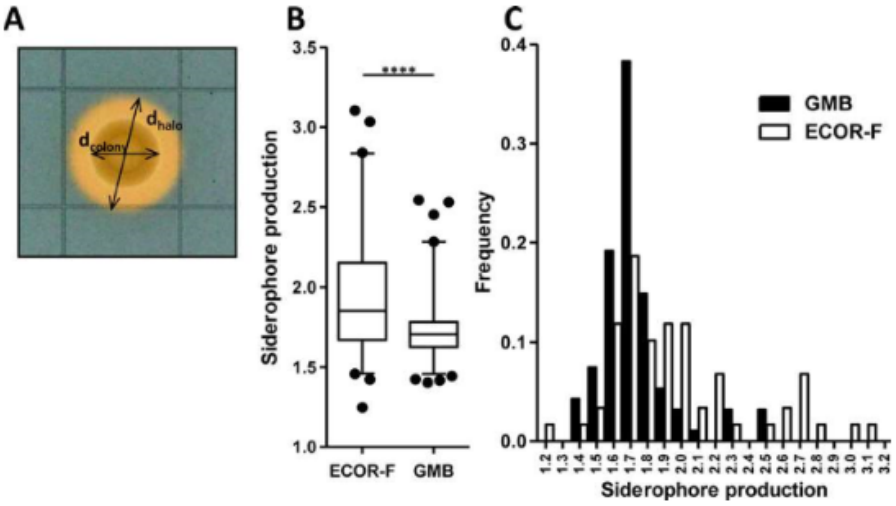


Fig 1. Plant associated *E. coli* display lower siderophore production compared to faecal isolates. A) Example of siderophore production levels obtained by dividing the halo diameter (d_{halo}) by the colony diameter (d_{colony}) measured on CAS agar plates. **B**) Box plot showing siderophore production for ECOR-F and GMB. The central rectangle of the plot spans the interquartile range (*IQR*). The segment inside the rectangle shows the median, while the whiskers span the 5–95 percentile. Black circles represent outliers. Statistical significance was determined using the Student *t*-test. ****$P<0.0001$. **C**) Frequency plot showing siderophore production for both strain collections.

dci:10.1371/journal.pone.0117906.g001

of siderophore systems [16–18]. Differences in siderophore production between ECOR-F and GMB could therefore be the simple reflection of the unequal phylogenetic composition of these collections [11]. However, the major *E. coli* phylogenetic groups were found to generate similar siderophore levels (One way ANOVA, $p>0.05$; S1 Fig.), indicating that the observed differences are likely to be linked to sample origin rather than phylogenetic composition.

Design and validation of multiplex PCR for the detection of siderophore utilisation and biosynthetic genes in *E. coli*

To determine whether the greater capacity to produce siderophores under iron limitation displayed by the faecal isolates was also reflected at the genome level by a different complement of siderophore biosynthesis genes. We designed a multiplex PCR aimed at detecting both receptor and biosynthesis genes of the four known *E. coli* siderophores. We took advantage of the large number of available *E. coli* genome sequences to generate multiple alignments and identify suitable conserved regions to design highly sensitive and specific PCR primers (Table 1). The specificity of the multiplex PCR was assessed on the ECOR collection [12], which has been previously monitored by PCR for the presence of siderophore receptor genes [16]. More recently, the gene content of the strains from the ECOR collection has been analysed using multi-genome arrays, which cover the entire enterobactin, aerobactin and salmochelin loci [19].

We found a good correlation with the array data with 96.7% (523/541) of positive and 96.1% (518/539) of negative PCR signals matching both datasets (S2 Table). Results on the presence/absence of the yersiniabactin locus matched previously PCR-based published data in 97.2% (70/72) of cases [20].

Published PCR-based data on the distribution of siderophore receptor genes supported our results when in conflict with the array data (full details are provided in S2 Table). However, our observations that ECOR02 and ECOR67 respectively lack the aerobactin and salmochelin loci were not supported. To exclude the possibility of a sample mix up, we took advantage of existing CRISPR sequence data that can differentiate between the different ECOR strains [21]. The DNA sequences of the CRISPR2.1 and CRISPR2.3 regions of ECOR02 and ECOR67 matched those found in the databases, thus confirming the identity of the strains. We also found differences for 2 strains (ECOR11 and ECOR72) with previously published PCR-based data in the distribution of the yersiniabactin locus (S2 Table) [20] despite the fact that both CRISPR sequences confirmed the identity of these strains. The discrepancies we have identified reflect previous observations that the phenotype and genotype of ECOR strains can differ between laboratories [22], and highlight the importance of characterising the strains used in independent studies.

Distribution of the aerobactin, enterobactin, salmochelin and yersiniabactin loci in plant and faecal *E. coli* isolates

The distribution of the siderophore biosynthetic loci in plant-associated *E. coli* was assessed by performing the validated multiplex PCRs on 96 strains from the GMB collection (S3 Table) [11]. Importantly, the siderophore receptor genes were always detected alongside the corresponding biosynthetic genes (133/133, S3 Table). This observation highlights the sensitivity of our approach and indicates that each siderophore locus is generally evolutionarily maintained as a complete unit.

The prevalence of siderophore production genes observed for the strains isolated from plants was compared with that of the healthy host isolates from the ECOR collection (ECOR-F). The results showed that, in accordance with siderophore production data, plant-associated *E. coli* generally possessed fewer siderophore production and uptake systems (Fig. 2). Yersiniabactin and aerobactin were found in significantly lower proportions in plant-associated *E. coli* (Table 2). This is also reflected by a significant difference in the proportion of strains only possessing the enterobactin locus; 76% (73/96) and 42.6% (26/61) for GMB and ECOR-F respectively (Fischer's exact test, $p<0.0001$, S2 and S3 Tables).

To expand our comparison beyond ECOR-F, assuming that the presence of the receptor genes generally correlates with the ability to synthesise the corresponding siderophores, we investigated the available recent literature for data describing the distribution of siderophore receptor genes in faecal and rectal *E. coli* isolates obtained from healthy humans [17,18,23–30]. Published data was consistent with data obtained for ECOR-F, confirming the narrower distribution of the aerobactin and yersiniabactin loci among the plant-associated isolates (Table 2).

Our results suggest that high siderophore production might be linked to the presence of the aerobactin and/or yersiniabactin loci. Accordingly, the presence of the aerobactin synthesis genes was nearly always associated with strong siderophore production (Table 3). However, no correlation between siderophore production and the presence of the yersiniabactin locus could be observed (Table 3). To exclude any potential effect linked to siderophore-specific differences in diffusion through the agar, the link between aerobactin and strong *in vitro* siderophore production was confirmed by assaying a representative subset of strains (n = 33) in liquid CAS medium (S2A Fig.). The liquid assays also reflected the CAS agar assays in showing lower siderophore production for the plant-associated strains (S2B Fig.).

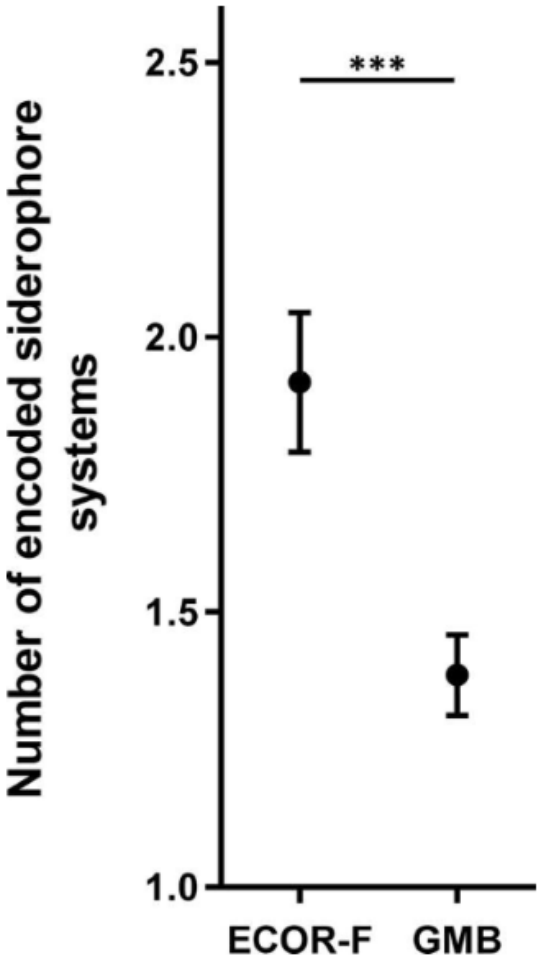


Fig 2. Plant associated *E. coli* isolates encode fewer siderophore production systems than faecal isolates at the population level. The graph displays the mean number of detected siderophore systems for the GMB and ECOR-F strain collections. Error bars display the standard error of the mean. Statistical significance was determined using the Student *t*-test. ***P<0.001.

doi:10.1371/journal.pone.0117906.g002

Strain-specific diversity in siderophore gene expression under low iron conditions

The differential distribution of siderophore loci was not sufficient to fully explain the difference in siderophore production between GMB and ECOR-F isolates. Strains only possessing the enterobactin locus displayed a wide range of siderophore production levels, with significant differences in enterobactin production, observable between plant-associated and faecal isolates at

Table 2. Distribution of siderophore biosynthetic systems in the genome of plant and host-associated *E. coli* populations.

Siderophore	Proportion of detected systems (%)			Statistical significance[a]		
	GMB (n = 96)	ECOR-F (n = 61)	Faecal[b] (n = 618–1042)[c]	GMB vs ECOR-F	GMB vs Faecal	ECOR-F vs Faecal
Aerobactin	5	25	29	<0.001	<0.001	NS[d]
Yersiniabactin	19	48	35	<0.001	<0.01	NS
Salmochelin	16	20	21	NS	NS	NS

[a]Significance tests were performed using the Fisher's exact test. The Benjamini and Hochberg False discovery rate method was used to correct for multiple comparisons.

[b]The data relative to healthy humans were obtained from the literature (see text for references).

[c]The range of isolates tested reflects the fact that not every study included analysed the distribution of all the siderophore systems (aerobactin, n = 1042; yersiniabactin, n = 618; salmochelin, n = 808).

[d]NS: no statistical significance detected.

dci:10.1371/journal.pone.0117906.t002

Table 3. Distribution of siderophore biosynthetic systems relative to siderophore production.

	top 25%	*25–50%*	*50–75%*	*bottom 25%*
Aerobactin (%)	42***	5	3	3
Yersiniabactin (%)	42	24	32	31
Salmochelin (%)	26*	21	13	5

Significance tests were performed comparing the top and bottom quartiles using the Fischer's exact test. The Benjamini and Hochberg False discovery rate method was used to correct for multiple comparisons.

*$P<0.05$,

***$P<0.001$.

dci:10.1371/journal.pone.0117906.t003

the population level (unpaired t-test; t = 3.13, p = 0.002, Fig. 3). This suggested widespread differences in the regulation of siderophore production between individual *E. coli* strains, and thus prompted the investigation of the possible role of differences in gene regulation of enterobactin and the 3 other siderophore systems. We examined the expression levels of one biosynthesis gene for each siderophore in 8 environmental strains displaying a range of siderophore production abilities under low iron conditions. Analysis of the mRNA levels by real-time PCR showed that enterobactin and aerobactin were the most highly expressed systems, both displaying 4 and 600-fold higher expression levels than salmochelin and yersiniabactin, respectively (Fig. 4A). Importantly, enterobactin as well as salmochelin displayed a wide range of different expression levels, which strongly correlated with total siderophore production (Fig. 4B).

In contrast, aerobactin and yersiniabactin displayed a much narrower range of expression, high and very low respectively in all strains tested. This provides a possible explanation for the strong link between presence of the aerobactin locus and high siderophore production in the CAS assays (Table 3, S2A Fig.). Conversely, the generally low yersiniabactin gene expression is in good agreement with the absence of correlation between the yersiniabactin locus and siderophore production levels.

Discussion

Iron acquisition is vital to microbial physiology and host colonisation. Production of the enterobactin siderophore has been shown to be important for the colonisation of the healthy GI-

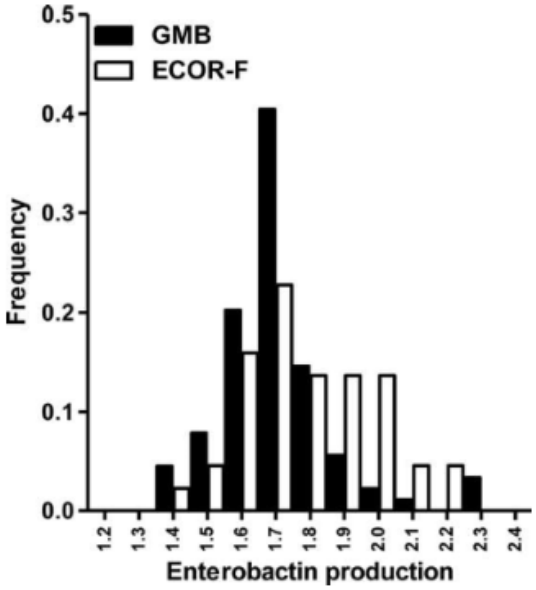


Fig 3. Plant associated *E. coli* display lower enterobactin production compared to faecal isolates. Frequency plot comparing siderophore production of GMB and ECOR-F strains only encoding the enterobactin siderophore locus. Sample sizes in the data sets were $n_{GMB} = 69$; $n_{ECOR\text{-}F} = 28$.

doi:10.1371/journal.pone.0117906.g003

tract by *E. coli* [31]. However, the ability to synthesise additional siderophores is generally associated with virulence, particularly for extraintestinal pathogenic *E. coli* that have reached the urinary tract or the blood stream [1,32,33]. The observation that many commensal *E. coli* isolates can produce these virulence-associated siderophores suggests another possible role in the ecology of *E. coli*, perhaps including persistence outside animal hosts [4]. Reasoning that a particular bacterial function will be more prevalent in the environment where it provides greater fitness advantage, we compared the siderophore production and distribution of siderophore biosynthetic and receptor genes in non-pathogenic *E. coli* isolated from plants or mammalian hosts. This population-wide approach has previously been used to identify environment-specific differences in *E. coli* carbon source utilisation and extra-cellular matrix production [11].

We found that plant-associated isolates excrete significantly lower amounts of siderophores compared to faecal isolates when grown under iron limiting conditions. Analysis of the distribution of the 4 siderophore biosynthetic systems in both *E. coli* populations showed that the aerobactin and yersiniabactin loci were less prevalent in the *E. coli* strains isolated from plants. The correlation between genotype and total siderophore production suggested that the greater ability of faecal isolates to generate siderophores may be partly linked to their greater association with the aerobactin locus. Interestingly, the aerobactin biosynthesis genes are often plasmid-encoded [1,34]. This could indicate that the lower frequency of the aerobactin locus among the GMB isolates might be linked to differences in plasmid carriage. When the comparison was restricted to strains only able to synthesise enterobactin, the difference in siderophore

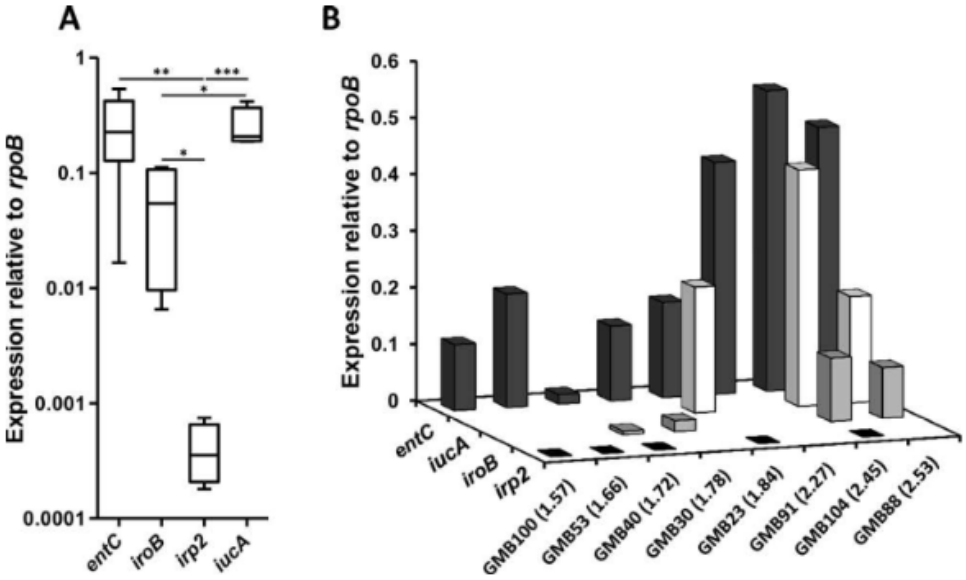


Fig 4. Gene expression level of siderophore biosynthetic genes in plant-associated *E. coli*. The expression of one biosynthetic gene was determined for each siderophore locus, when present, in 8 GMB strains; *entC* (enterobactin), *iroB* (salmochelin), *irp2* (yersiniabactin), *iucA* (aerobactin). The strains analysed were GMB23 (*entC*, *irp2*, *iucA*), GMB30 (*entC*, *iroB*), GMB40 (*entC*, *iroB*, *irp2*), GMB53 (*entC*, *irp2*), GMB88 (*entC*, *iroB*, *iucA*), GMB91 (*entC*), GMB100 (*entC*, *irp2*) and GMB104 (*entC*, *iroB*, *irp2*, *iucA*). **A**) Box plot showing the gene expression levels of *entC*, *iroB*, *irp2*, and *iucA* relative to the internal reference *rpoB*. The central rectangle of the plot spans the interquartile range (*IQR*). The segment inside the rectangle shows the median, while the bars above and below show the location of the maximum and minimum, respectively. Statistical significance was determined using the Student *t*-test. In case of multiple tests, the significance of individual *t*-tests was determined using the Benjamini and Hochberg False discovery rate method; $*P<0.05$, $**P<0.01$, $***P<0.001$. **B**) To visualise the link between gene expression and siderophore production as measured on CAS agar plates, expression levels are shown for each individual strain ranked from low to high producer. Siderophore production is indicated in brackets after the corresponding strain name.

dci:10.1371/journal.pone.0117906.g004

production was still significant between plant and faecal isolates, suggesting that diversification in the regulatory response to iron limitation may also be an important environmental adaptive trait. The possible role of differential expression of enterobactin in *E. coli* adaptation mirrors previously observed differences in yersiniabactin and aerobactin synthesis between single-patient urinary and rectal *E. coli* isolates [35].

Our results suggest that the environment influences the distribution of siderophore production genes among the associated *E. coli* populations, and that the gut selects for *E. coli* commensal strains with multiple siderophore systems and with the ability to generate higher siderophore levels under iron limitation. In contrast, the plant environment seems to exert a weaker selective pressure on *E. coli* to maintain a diverse siderophore production, probably as a consequence of differences in iron source availability compared to the gut. In association with plants, *E. coli* could rely on alternative iron acquisition systems such as the direct uptake of ferrichrome or ferric citrate, molecules which can be found both in rhizosphere and phyllosphere [36–38]. It would therefore be interesting to expand this type of analysis to *E. coli* populations associated with other plant-hosts and climates to determine whether our observations can be generalised. It is also important to note that the differences in geography and time of isolation between GMB and ECOR-F could be potential confounding factors in our study. However, geography explains only 2% of the genetic diversity between the ECOR strains isolated in Europe

and the USA [39]. Similarly, the time of isolation has probably had little impact on the distribution of siderophore receptor genes as more recent studies of human commensal *E. coli* have reported comparable gene distribution to ECOR-F (Table 2).

Without excluding a role outside the host, production of aerobactin and yersiniabactin might therefore be important for gut colonisation by commensal *E. coli*. This hypothesis is supported by the observation that in contrast to transient *E. coli* clones, resident isolates which could be detected in the GI-tract of the same individual over 3 months or more were enriched for aerobactin biosynthesis genes [40]. Similarly, it is possible that yersiniabactin provides a fitness advantage to commensal *E. coli* since the rapid spread of the yersiniabactin locus in both virulent and commensal *E. coli* implies a role beyond pathogenesis [4,16,20].

Despite the availability of information about siderophore-mediated iron acquisition and the importance of enterobactin for *E. coli* gut colonisation [31], it is not clear how having higher and more diverse siderophore production benefits gut commensal *E. coli*. It has been shown that the 4 siderophores encoded by *E. coli* Nissle 1917 are differently affected by environmental factors such as pH and carbon source [41]. An expanded siderophore repertoire would therefore make a cell more competitive in acquiring iron in the spatially and temporally heterogeneous environment found in the gut. Strains possessing several siderophore systems could also have an advantage in the densely populated gut environment when in competition with bacteria that utilise a narrower range of siderophores.

The metabolic cost of generating siderophores is significant and has a clear impact on the *E. coli* metabolome [42]. Indeed, it has been shown that under specific circumstances bacteria can lose the ability to make siderophores while retaining the machinery for acquiring those produced by other strains, thus opening the possibility of cheating [43,44]. However, we observed that the siderophore receptor genes were always detected alongside the corresponding biosynthetic genes. This suggests that the natural environment selects against *E. coli* cheaters. One possible explanation might be the poor diffusion of siderophores in viscous media such as mucus [45], where *E. coli* is thought to reside, making it less likely for a cheater to be in contact with significant amounts of the required siderophore. This hypothesis is supported by the observation that enterobactin hyperexcretion makes catecholate transport-deficient *E. coli* iron-starved *in vivo* despite retaining the ability to import ferric iron through other systems [31]. This phenomenon could be explained by the entrapment of enterobactin around mucus-embedded cells, thus making the iron inaccessible to other acquisition systems.

In summary, our data suggest that the ability to produce yersiniabactin and aerobactin is associated with the colonisation of the gut by *E. coli*. The conditions under which each siderophore system is expressed and influences *E. coli* fitness in the GI-tract requires further investigation, as does the extent to which our findings can be extended to the persistence of pathogenic *E. coli*, such as those strains causing infections of the urinary tract, for which the physiologically normal gut is a likely reservoir [46,47].

Supporting Information

S1 Fig. Phylogenetic distribution of siderophore production. Box plots showing siderophore production by GMB and ECOR-F strains split according to the major *E. coli* phylogenetic groups. The central rectangle of the plot spans the interquartile range (*IQR*). The segment inside the rectangle shows the median, while the whiskers span the 5–95 percentile. Black circles represent outliers. Statistical significance was determined using the Student *t*-test. No significant differences were detected.
(PDF)

S2 Fig. Siderophore production of GMB and ECOR-F strains as assessed by liquid CAS assay. Box plots showing siderophore production in liquid CAS for a subset of strains (n = 33) to visualise differences in siderophore production of **A**) strains with or without the aerobactin biosynthesis locus, and **B**) GMB and ECOR-F. The central rectangle of the plot spans the inter-quartile range (*IQR*). The segment inside the rectangle shows the median, while the whiskers span the 5–95 percentile. Black circles represent outliers. Statistical significance was determined using the Student *t*-test; $^{*}P<0.05$, $^{****}P<0.0001$.
(PDF)

S1 Table. Siderophore production of GMB and ECOR-F strains as assessed by CAS agar assay. The table displays the ratios between colony and halo diameter as measured on chrome azurol S agar plates for each GMB and ECOR-F strain.
(PDF)

S2 Table. Distribution of enterobactin, aerobactin, yersiniabactin and salmochelin loci in the ECOR collection as assessed by multiplex PCR. The table shows presence/absence of the genes encoded in siderophore production loci in *E. coli* isolates from the ECOR strain collection.
(PDF)

S3 Table. Distribution of enterobactin, aerobactin, yersiniabactin and salmochelin loci in the GMB collection as assessed by multiplex PCR. The table shows presence/absence of the genes encoded in siderophore production loci in *E. coli* isolates from the GMB strain collection.
(PDF)

Acknowledgments

We would like to thank Jack Dainty for assistance with the statistical analysis of the data.

Author Contributions

Conceived and designed the experiments: LJS GM SL. Performed the experiments: LJS GM. Analyzed the data: LJS GM IP SL. Contributed reagents/materials/analysis tools: SKS. Wrote the paper: LJS GM SL.

References

1. Garenaux A, Caza M, Dozois CM (2011) The Ins and Outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*. Vet Microbiol 153: 89–98. doi: 10.1016/j.vetmic.2011.05.023 PMID: 21680117
2. Braun V, Killmann H (1999) Bacterial solutions to the iron supply problem. Trends Biochem Sci 24: 104–109. PMID: 10203757
3. Miethke M, Marahiel MA (2007) Siderophore-based iron acquisition and pathogen control. Microbiol Mol Biol Rev 71: 413–451. PMID: 17804665
4. van Elsas JD, Semenov AV, Costa R, Trevors JT (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. ISME J 5: 173–183. doi: 10.1038/ismej.2010.80 PMID: 20574458
5. Beuchat LR (2002) Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. Microb Infect 4: 413–423.
6. Cornelis P (2010) Iron uptake and metabolism in *pseudomonads*. Appl Microbiol Biotechnol 86: 1637–1645. doi: 10.1007/s00253-010-2550-2 PMID: 20352420
7. Diallo S, Crepin A, Barbey C, Orange N, Burini J- F, et al. (2011) Mechanisms and recent advances in biological control mediated through the potato rhizosphere. FEMS Microbiol Ecol 75: 351–364. doi: 10.1111/j.1574-6941.2010.01023.x PMID: 21204870

8. Jurkevitch E, Hadar Y, Chen Y (1992) Differential siderophore utilization and iron uptake by soil and rhizospheric bacteria. Appl Environ Microbiol 58: 119–124. PMID: 16348618
9. Hao L-y, Willis DK, Andrews-Polymenis H, McClelland M, Barak JD (2012) Requirement of siderophore biosynthesis for plant colonization by *Salmonella enterica*. Appl Environ Microbiol 78: 4561–4570. doi: 10.1128/AEM.07867-11 PMID: 22522683
10. Bergholz PW, Noar JD, Buckley DH (2011) Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. Appl Environ Microbiol 77: 211–219. doi: 10.1128/AEM.01880-10 PMID: 21075897
11. Meric G, Kemsley EK, Falush D, Saggers EJ, Lucchini S (2013) Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. Environ Microbiol 15: 487–501. doi: 10.1111/j.1462-2920.2012.02852.x PMID: 22934605
12. Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157: 690–693. PMID: 6363394
13. Payne SM (1984) Detection, isolation, and characterization of siderophores. Methods Enzymol 235: 329–344.
14. Watts RE, Totsika M, Challinor VL, Mabbett AN, Ulett GC, et al. (2012) Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*. Infect Immun 80: 333–344. doi: 10.1128/IAI.05594-11 PMID: 21930757
15. Tedin K, Blasi U (1996) The RNA chain elongation rate of the lambda late mRNA is unaffected by high levels of ppGpp in the absence of amino acid starvation. J Biol Chem 271: 17675–17686. PMID: 8663373
16. Johnson JR, Delavari P, Kuskowski M, Stell AL (2001) Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. J Infect Dis 183: 78–88. PMID: 11106538
17. Johnson JR, Owens K, Gajewski A, Kuskowski MA (2005) Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. J Clin Microbiol 43: 6064–6072. PMID: 16333100
18. White AP, Sibley KA, Sibley CD, Wasmuth JD, Schaefer R, et al. (2011) Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. Appl Environ Microbiol 77: 7620–7632. doi: 10.1128/AEM.05909-11 PMID: 21908635
19. Jackson SA, Patel IR, Barnaba T, LeClerc JE, Cebula TA (2011) Investigating the global genomic diversity of *Escherichia coli* using a multi-genome DNA microarray platform with novel gene prediction strategies. BMC Genomics 12: 349. doi: 10.1186/1471-2164-12-349 PMID: 21733163
20. Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, et al. (2009) Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. PLoS Path 5: e1000257.
21. Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJM (2010) Diversity of CRISPR loci in *Escherichia coli*. Microbiology-SGM 156: 1351–1361. doi: 10.1099/mic.0.036046-0 PMID: 20133361
22. Johnson JR, Delavari P, Stell AL, Prats G, Carlino U, et al. (2001) Integrity of archival strain collections: the ECOR collection. ASM News 67: 288–289.
23. Hilali F, Ruimy R, Saulnier P, Barnabe C, Lebouguenec C, et al. (2000) Prevalence of virulence genes and clonality in *Escherichia coli* strains that cause bacteremia in cancer patients. Infect Immun 68: 3983–3989. PMID: 10858212
24. Johnson JR, Johnston B, Kuskowski MA, Nougayrede J- P, Oswald E (2008) Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. J Clin Microbiol 46: 3906–3911. doi: 10.1128/JCM.00949-08 PMID: 18945841
25. Kudinha T, Kong F, Johnson JR, Andrew SD, Anderson P, et al. (2012) Multiplex PCR-based reverse line blot assay for simultaneous detection of 22 virulence genes in uropathogenic *Escherichia coli*. Appl Environ Microbiol 78: 1198–1202. doi: 10.1128/AEM.06921-11 PMID: 22156422
26. Lee S, Yu JK, Park K, Oh E- J, Kim S- Y, et al. (2010) Phylogenetic groups and virulence factors in pathogenic and commensal strains of *Escherichia coli* and their association with *bla*(CTX-M). Ann Clin Lab Sci 40: 361–367. PMID: 20947811
27. Mao B- H, Chang Y- F, Scaria J, Chang C- C, Chou L- W, et al. (2012) Identification of *Escherichia coli* genes associated with urinary tract infections. J Clin Microbiol 50: 449–456. doi: 10.1128/JCM.00640-11 PMID: 22075599
28. Nowrouzian F, Hesselmar B, Saalman R, Strannegard IL, Aberg N, et al. (2003) *Escherichia coli* in infants' intestinal microflora: Colonization rate, strain turnover, and virulence gene carriage. Pediatr Res 54: 8–14. PMID: 12700366
29. Unno T, Han D, Jang J, Widmer K, Ko G, et al. (2011) Genotypic and phenotypic trends in antibiotic resistant pathogenic *Escherichia coli* isolated from humans and farm animals in South Korea. Microbes Environ 26: 198–204. PMID: 21558676

30. Vollmerhausen TL, Ramos NL, Guendogdu A, Robinson W, Brauner A, et al. (2011) Population structure and uropathogenic virulence-associated genes of faecal *Escherichia coli* from healthy young and elderly adults. J Med Microbiol 60: 574–581. doi: 10.1099/jmm.0.027037-0 PMID: 21292854

31. Pi H, Jones SA, Mercer LE, Meador JP, Caughron JE, et al. (2012) Role of catecholate siderophores in Gram-negative bacterial colonization of the mouse gut. PLoS One 7: e50020. doi: 10.1371/journal.pone.0050020 PMID: 23209633

32. Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, et al. (2006) Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. J Infect Dis 194: 1141–1150. PMID: 16991090

33. Caza M, Lepine F, Milot S, Dozois CM (2008) Specific roles of the *iro*BCDEN genes in virulence of an avian pathogenic *Escherichia coli* O78 strain and in production of salmochelins. Infect Immun 76: 3539–3549. doi: 10.1128/IAI.00455-08 PMID: 18541653

34. Vokes SA, Reeves SA, Torres AG, Payne SM (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. Mol Microbiol 33: 63–73. PMID: 10411724

35. Henderson JP, Crowley JR, Pinkner JS, Walker JN, Tsukayama P, et al. (2009) Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic *Escherichia coli*. PLoS Path 5: e1000305.

36. Crowley DE (2007) Microbial siderophores in the plant rhizosphere; Barton LL, Abadia J, editors. 169–198 p.

37. Reid RK, Reid CPP, Powell PE, Szaniszlo PJ (1984) Comparison of siderophore concentrations in aqueous extracts of rhizosphere and adjacent bulk soils. Pedobiologia 26: 263–266.

38. Devos CR, Lubberding HJ, Bienfait HF (1986) Rhizosphere acidification as a response to iron-deficiency in bean-plants. Plant Physiol 81: 842–846. PMID: 16664912

39. Miller RD, Hartl DL (1986) Biotyping confirms a nearly clonal population-structure in *Escherichia coli*. Evolution 40: 1–12.

40. Nowrouzian F, Adlerberth I, Wold PE (2001) P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli*. Epidemiol Infect 126: 11–18. PMID: 11293669

41. Valdebenito M, Crumbliss AL, Winkelmann G, Hantke K (2006) Environmental factors influence the production of enterobactin, salmochelin, aerobactin, and yersiniabactin in *Escherichia coli* strain Nissle 1917. Int J Med Microbiol 296: 513–520. PMID: 17008127

42. Lv HT, Hung CS, Henderson JP (2014) Metabolomic analysis of siderophore cheater mutants reveals metabolic costs of expression in uropathogenic *Escherichia coli*. J Prot Res 13: 1397–1404.

43. Buckling A, Harrison F, Vos M, Brockhurst MA, Gardner A, et al. (2007) Siderophore-mediated cooperation and virulence in *Pseudomonas aeruginosa*. FEMS Microbiol Ecol 62: 135–141. PMID: 17919300

44. De Vos D, De Chial M, Cochez C, Jansen S, Tummler B, et al. (2001) Study of pyoverdine type and production by *Pseudomonas aeruginosa* isolated from cystic fibrosis patients: prevalence of type II pyoverdine isolates and accumulation of pyoverdine-negative mutations. Arch Microbiol 175: 384–388. PMID: 11409549

45. Kuemmerli R, Griffin AS, West SA, Buckling A, Harrison F (2009) Viscous medium promotes cooperation in the pathogenic bacterium *Pseudomonas aeruginosa*. Proc R Soc B 276: 3531–3538. doi: 10.1098/rspb.2009.0861 PMID: 19605393

46. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, et al. (2013) Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med 5: 184ra160.

47. Russo TA, Stapleton A, Wenderoth S, Hooton TM, Stamm WE (1995) Chromosomal restrction-fragment-length-polymorphism analysis of *Escherichia coli* strains causing recurrent urinary-tract infections in young women. J Infect Dis 172: 440–445. PMID: 7622887