**The intra-host population structure of UPEC during UTI**

**Abstract**

Urinary tract infections (UTIs) are one of the most common bacterial infections with 60% of women likely to suffer a UTI at some point in their lives. Uropathogenic *Escherichia coli* (UPEC) are responsible for 80% of community acquired UTIs and result in billions of dollars in direct costs. During a UTI, UPEC invade and colonize the bladder, where they persist until they are cleared by therapeutic intervention or by the immune system. UPEC is considered a facultative pathogen because the bacterium resides in the distal mammalian gut as a commensal, but causes damage and disease when it colonizes the bladder. Accordingly, the recal-perineal-urethral hypothesis states that the gut acts a reservoir for UPEC, which can then escape the gut, colonize the periurethral area, and eventually invade and colonize the bladder. This hypothesis has been supported by a number of studies of UTI patients that have found that the urine isolates were identical to one of the strains isolated from the patient’s gut at the same time of the UTI. While these studies have focused on the relative abundance of strains of UPEC within the *E. coli* population, these investigations have ignored the community structure of the gut microbiota as a whole. The healthy human microbiota serves as a barrier to the invasion of harmful pathogens through occupation of the available niches. Disruption of the gut microbiota, termed dysbiosis, has been implicated in a number of diseases, such as infection with *Clostridium difficile*. In these situations, harmful bacteria that are absent or exist at low abundances in a healthy gut undergo a bloom of growth and increase their relative abundance and prevalence. A similar bloom of UPEC in the periurethral microbiome before the onset of a UTI has been documented; however, the abundance and prevalence of UPEC in the gut microbiota of patients with UTI has not been measured, and the composition of the gut microbiome during UTI has yet to be described. My research will investigate the composition of the gut microbiota, with particular focus on the community structure of UPEC, in patients during the onset and after the clearance of a UTI caused by UPEC. This information will aid our understanding of the gut reservoir of UPEC, which may impact the disease progression of UTIs.

**Sub-aim1: Compare the community structure of the gut microbiota during acute UTI and after recovery.**

Metagenomic sequencing of the bacterial 16S rRNA gene can been used to describe the bacterial community structure of the gut microbiome and measure the relative abundance of particular phylotypes therein. Comparisons of the community structure at different time points within the same patient can be used to identify changes in the phylotypes measured in the metagenomic sequence. In order to determine if the gut microbiota is different within a patient during UTI and after recovery, I will isolate genomic DNA from three fecal samples collected from patients during the onset of a UTI, after antibiotic treatment, and after recovery from the antibiotics. Results obtained from the metagenomic sequencing of this DNA will then be compared to publically available information on healthy gut microbiomes to determine if the gut microbiotas of patients during UTI are dysbiotic. Further, the abundance of sequencing reads that map to *Escherichia* in the fecal samples will be used to estimate the relative abundance of *Escherichia* species in the gut microbiome.

**Sub-aim2: Compare the relative abundance of urovirulence genes in the gut microbiome during acute UTI and after recovery.**

Not all *Escherichia* species are pathogenic and analysis of the 16S rRNA gene of *Escherichia* species is not capable of accurately determining the uropathogenicity of a particular strain. However, quantitative polymerase chain reaction (qPCR) can be used to measure the abundance of known UPEC urovirulence genes relative to the abundance of total bacterial 16S rRNA genes in a fecal sample, which can be used to estimate the relative abundance of UPEC in the gut microbiome. As above, longitudinal comparisons of three samples taken from patients at the onset of a UTI (before antibiotics are given), after antibiotic treatment, and after recovery from the antibiotics can be used to measure differences in UPEC relative abundance or changes in the urovirulence gene carriage.

Completion of these aims will result in a better understanding of the relationship between the gut microbiome, UPEC, and UTIs.

**Literature Review**

**Uropathogenic *Escherichia* *coli* – a facultative pathogen and model system**

Approximately 50% of women will suffer a urinary tract infection (UTI) by the age of 321 and up to 15 million cases of UTI occur each year2. Approximately 20-30% of these women will suffer a recurrent UTI within three to four months following an initial UT2-4. The rates of UTI increase in the immunosuppressed, including the elderly and children, and may result in significant complications, including renal scarring, septicemia, and pyelonephritis2,5. UTIs are commonly acquired in the community, but are also the most common nosocomial infection1,6,7. UTIs are responsible for over 8.4 million clinic visits in 2007 and more than $2.5 billion in direct costs per annum, and have been increasing in costs in the last decade2,8-12. Clinical manifestations of UTIs include dysuria, foul-smelling or cloudy urine, fever, and flank pain8,11,13. UTIs can be classified as lower UTIs, which are confined to the bladder, or upper UTIs, which involve the kidneys. Uncomplicated UTIs are defined as infections that occur in patients without structural (e.g. large ureters) or functional (e.g. inability to fully clear the bladder) abnormalities of the urinary tract, that are not pregnant, and who do not have catheters or other instruments installed; all other cases of UTIs are considered complicated7,13. This review will focus on uncomplicated, lower UTIs that are acquired in the community.

Nearly 80% of community acquired UTIs are caused by uropathogenic *E. coli* (UPEC)8, while the remaining 20% are mainly caused by other Enterobacteriaceae, such as *Klebsiella* and *Enterobacter*, as well as Gram-positive organisms, such as *Staphylococcus saprophyticus*6. Despite its prevalence and pathogenicity, UPEC is considered a facultative pathogen, as are many other types of *E. coli*1,14,15. Facultative pathogens, such as the virulent O157:H7 strain of *E. coli*, live commensally in one habitat, such as cattle gastrointestinal tracts, but are capable of causing disease in alternative habitats, such as the human gastrointestinal tract. This pattern differs from obligate pathogens, such as the *Shigella* species, which are unable to colonize a host without causing disease1,2,15. UPEC is known to exist in the human gut as a commensal, but is also capable of causing disease if it is able to invade and colonize the bladder or the kidneys. UPEC is also known to inhabit the periurethral area and vagina without causing disease, and these habitats have been suggested to be potential reservoirs that are capable of invading the urinary tract2-4,16,17.

In addition to being clinically important, UPEC is also an excellent model system to study virulence in facultative pathogens. UPEC offer a number of unique advantages as a model system, including the range of laboratory tools available specific to *E. coli*, the tractability of genetic modification, and the wealth of genomic data available for the pathogen. UPEC have been used to study biofilm formation, pili structure and function, epithelial cell invasion, toxin production, and population bottlenecks, in addition to its obvious use as a model for uropathogenicity2-5,18-20. The evolution of virulence in this facultative pathogen has also been studied, which has resulted in a number of competing theories, which will be discussed below. Finally, although there has been attention paid to the global phylogenetic structure of UPEC, relatively few investigations have sought to describe the within-host distribution of UPEC populations or elucidate the changes in population structure that occur within patients with recurrent UTIs. However, new technologies, such as second-generation sequencing, now enable high-resolution descriptions of bacterial population structures using genomic analyses, enabling research into these unexplored areas. These analyses into population structure will facilitate a better understanding of how virulence has evolved in *E. coli* by describing the selection pressures faced by UPEC in their host habitats. Furthermore, because UPEC UTI can be used as a model for mucosal infections, the information gained from these studies will aid our understanding of other mucosal infections1,2,5-7,19.

**The within-host population structure of UPEC**

*Classification of Escherichia coli isolates*

*Escherichia coli* are associated with a number clinical conditions, each caused by *E. coli* strains harboring different repertoires of gene sets and virulence factors, and, therefore, can be categorized according to their pathology and genomic content. *E. coli* that cause disease in the gastrointestinal tract are grouped together into a super-group labeled intestinal pathogenic *E. coli* (IPEC)21-24. A separate group consists of extra-intestinal pathogenic *E. coli* (ExPEC) and includes strains of uropathogenic *E. coli* (UPEC) that are capable of causing urinary tract infections 1,6-8,11,13,25. In addition to pathotype, *E. coli* can also be categorized according to their phylogenetic history. Interestingly, pathogenic potential, genomic content, and phylogenetic history are not always perfectly concordant (Figure 1). Currently, four main clades of *E. coli* have been described, A, B1, B2, and D, along with two smaller clades, C and E2,7-13,26-28. ExPEC fall predominately into clade B2, and to a lesser extent D and are generally absent from other clades8,11,13,26 and the majority of urine isolates of E. coli are from clade B26,7,13,29-31. Clade B2 can be further subdivided into 9 sub-clades, of which several are correlated with increased urpathogenicity8,14. Although there appears to be a connection between phylogeny and virulence, UPEC strains have been isolated from clades A, B1, B2, and D 6,26,32. Thus, the four main clades of *E. coli* differ in their phylogenetic history, in addition to niche preference and life history, but these differences are not absolute predictors of pathogenic potential1,14,15,26,27,32.

*UPEC population structure in the bladder*

In general, UPEC populations in the bladder are short lived due to a combination of innate immune response and therapeutic intervention. Before the expansion of antibiotic use, bacteria were known to reside in the bladder for long periods of time, despite palliative care to remove symptoms33, thus indicating that a large portion of people is unable to clear the bacterial infection without curative treatment. This is supported by a recent placebo trial in which only 37% of women were able to clear a UTI by 5-7 weeks34. As a result, antibiotic therapy is widely used as a curative, and in cases of frequent recurrent UTI, a preventative therapy. This widespread antibiotic therapy, such as treatments with trimethoprim**/**sulfamethoxazole or fluoroquinolones, have resulted in the spread of antibiotic resistance35 and given rise to multidrug resistant isolates that now represent a major public health concern36. The effects of antibiotic use on the population structure of UPEC in a community have been studied, but more information is necessary to understand the long-term effects of these antibiotic treatments on within-host distribution of this facultative pathogen.

The population dynamics of UPEC during the course of a UTI are complex and consist of a number of bottleneck events that occur both outside and within the host epithelium18 and result in a drift to clonality within the bladder37,38. A stringent bottleneck occurs during the formation of intracellular bacterial colonies (IBCs), which is a critical step of UPEC pathogenesis that occurs during the acute phase of UTIs39-41. Although IBCs allow for significant clonal expansion of UPEC40, formation of the IBCs occurs at a very low rate, with only 50-700 IBCs persisting at 6h after inoculation of 107 UPEC bacteria37. The precise mechanisms underlying this severe bottleneck have not been fully described, but are known to involve interactions between the host and pathogen and rely on both genetic and environmental factors18. Formation of these IBCs requires known virulence factors, including the adhesin *fimH*42. While the IBC bottleneck is important during the acute phase of UTI, the disappearance of IBCs at the end of the acute phase does not halt the continued loss in genetic diversity, suggesting a secondary bottleneck that occurs during the extracellular, chronic phase of UTI18. As with the IBC bottleneck, passage through the extracellular bottleneck may also be mediated by virulence factors. This hypothesis has been supported by inability of a mutant UTI89 lacking a pathogenicity associated island (PAI) containing known virulence factors, such as a-hemolysin and P pili, to persist during chronic UTI18. These findings show that virulence factors have a significant effect on population structure of UPEC in the bladder, which, in turn, affects disease progression through acute and chronic phases of UTI.

*Escherichia coli population structure in the gut*

*E. coli* are some of the first bacteria to colonize the gut43,44, although they become less abundant as the gut microbiome matures45. In adults, the gut population of *E. coli* is comprised of a dominant strain that accounts for the majority of *E. coli* in the gut, and a handful of minor strains that contribute the remainder31,44,46-50. Several of these longitudinal studies have indicated that the dominant strain in the gut, termed a “resident strain” is relatively stable for months or years while the minor strains, labeled “transient strains”, persist only for a few weeks to a month44,46-48; however, more recent evidence suggests that the rate of turnover for the dominant strain is much more likely to occur on the order of days or weeks49,50.

The changes that occur in gut populations of *E. coli* during UTI have yet to be fully explored. Currently, it is known that the average number of *E. coli* strains in the gut of women experiencing a UTI (~3) does not differ drastically from the average number of *E. coli* strains in the guts of healthy women (2.5) as determined by PCR typing30,31. Whether this maintenance of strain richness is mirrored by a maintenance bacterial abundance relative to the rest of the microbiota has not been investigated; however, an expansion of *E. coli* abundance, measured relatively to the rest of the microbiota, in the periurethral area is known to occur during the days preceding a UTI17. Given the connection between the gut and the periurethral area, it is reasonable to hypothesize that such a bloom of *E. coli* growth in the gut accompanies onset of UTI symptoms. Further research is warranted to fully understand how the gut microbiota shapes and is shaped by colonization of the bladder by UPEC.

*Transmission of UPEC between the gut and bladder*

The fecal-perineal-urethral hypothesis postulates that the vagina, perineum, and gut are reservoirs for UPEC that cause UTIs16,51-53. In this model, UPEC originate in the gut habitat and colonize the periurethral area or vagina and then ascend into the bladder. Evidence for this model comes from the clonality found between strains colonizing the bladder and either the gut or periurethral area, or both. During acute UTI, several studies have shown that the strains isolated from the urine are found to be the dominant strain in the rectal and fecal populations of *E. coli* and the dominant clone of the periurethral microbiota was found to be the same as the strain causing subsequent UTIs in the majority of cases17,30,51,52,54. These findings stand in contrast to two studies indicating that the UPEC strain causing the UTI may not be found regularly in the feces or periurethral area49,55, however the pattern of gut and bladder dominance is supported by recent research conducted in the Hultgren lab that found similar results (paper in review).

There are two competing hypotheses that purport to explain the pattern of colonization of the bladder by enteric bacteria. The “prevalence” hypothesis states that UTIs are most likely to be caused by the dominant enteric strain in the gut56, while the “special pathogenicity” hypothesis stipulates that the presence of virulence factors mediate the success of bladder colonization57. Recent research has indicated that the two models may not be mutually exclusive, which supports an integrated model in which the presence of urovirulence factors and gut prevalence are highly associated30,31,54. In this integrated model, increased abundance of UPEC in the gut was associated with increased rates of bladder colonization during acute UTI, but this abundance did not fully abrogate the need for urovirulence factors, thus indicating that both high gut titers and the presence of urovirulence factors are needed for bladder colonization30,54. Other studies of intestinal and bladder colonization have revealed trends that differ from these findings. In these studies, both dominant and minor *E. coli* strains from the gut and periurethral area were found to be the same as the strain isolated from the bladder, and, in many cases, the strains colonizing the bladder could not be isolated from the gut or periurethral area in the same week49,50. This indicates that bacterial prevalence may not be as important as the presence of urovirulence factors in mediating successful colonization of the bladder. Taken together, these studies imply that urovirulence factors may have a significant effect on the gut population structure, an implication that will be further discussed below.

**UPEC virulence factors**

*The UPEC armament*

The wide range of UPEC virulence factors can be broadly categorized into four groups, namely adherence factors, toxins, protectins, and iron acquisition systems(Table 1). While the definition of a virulence factor can be complicated, given the multiple uses of many “virulence factors” outside of pathogenicity58, the term here is defined as a product encoded in the bacterial genome that increases its ability to cause disease. Of course, the host immune system, as well as environmental factors, plays critical roles in the outcome of a UTI, but the following virulence factors have been shown increase pathogenicity.

Adherence to host cells is critical for UPEC virulence, particularly during the acute phases of UTI39,59. UPEC employ a number of adherence factors, including the ubiquitous type 1 pili, encoded by the *fim* gene cluster, have been shown to be critical for UPEC colonization of the bladder39. Type 1 pili, tipped with the FimH adhesin, bind to uroplakin-1 found in the host bladder, and assist in UPEC invasion of the urothelium60. The presence of P pili is strongly associated with strains that cause pyelonephritis61, while both S pili and Dr adhesins are capable of binding to the bladder epithelium62,63. F1C pili have been found to enhance bladder colonization, but the molecular mechanism of this activity has yet to be defined64. The variety of adherence mechanisms present in the UPEC repertoire are an indication of the complex interactions between host and pathogen and the necessity of adherence to bacterial pathogenesis.

Four toxins have been found to influence the rate of bladder colonization by UPEC, Hemolysin, Cytotoxic Necrotizing Factor 1 (CNF1), secreted autotransporter toxin (sat), and vacuolating autotransporter toxin (vat), which are encoded by the *hly*, *CNF1*, *sat,* and *vat* gene clusters, respectively65. Each of these virulence factors has been associated with increased ability to colonize the bladder in murine models of UTI, although their mechanisms of action are different64. Hemolysin has been associated with increased risk of septicemia through cytotoxic activity while CNF1 enhances host cell adhesion and invasion through activation of the host’s Rho GTP-binding proteins, which results in a remodeling of the host cell’s cytoskeleton. Both sat and vat have been shown to cause damage to urinary tract cells *in vitro*, and have been associated with cytopathic effects and tissue damage *in vivo*64,65. Together, these toxins are capable of remodeling the bladder epithelium, causing tissue damage, and enhancing UPEC persistence and pathogenicity.

Protectins are another class of virulence factors found in UPEC genomes. These factors enhance persistence of the bacteria in the host environment through a number of mechanisms, including immune cell evasion, disruption of antimicrobial activity, and elimination of bacterial competition and are encoded by the *traT*, *ompT*, and *cva* genes, respectively64,65. An additional gene cluster, labeled *iss* for increased serum survivability, has been associated with complement resistance and increased pathogenicity; however, whether these phenotypes are mediated by the *iss* gene itself or an associated gene has yet to be elucidated66.

Iron is a necessary factor for both prokaryotic and eukaryotic enzymatic processes, and the extracellular levels of iron in host environments, 10-25 M, are much lower than the level required for bacterial survival, which is estimate to be 10-6 M (reference 67). As a result, iron acquisition systems play an important role in bacterial pathogenesis and persistence in host environments (reviewed in Andrews et al. 2003). UPEC, like other bacterial pathogens, are capable of deploying a number of iron acquisition systems that sequester and internalize the iron from the host environment. These systems include siderophores, such as the enterobactin, yersiniabactin, salmochelin, and aerobactin siderophores, that have binding affinities of approaching 10-49, which are capable of outcompeting host iron acquisition systems, such as transferrin, which have weaker binding affinities of around 10-20 (reference 78). UPEC may also contain the *chu, feo,* or *Sit* acquisition systems, which encode the Hemin uptake system, ferrous iron autotransporter, and an iron/manganese transport system, respectively. Often, a single strain of UPEC will contain many redundant iron acquisition systems, and no single iron acquisition system is necessary for virulence64.

*UPEC genotypes are varied, but structured*

As with many other *E. coli* pathotypes, UPEC have evolved via horizontal gene transfer and recombination, which has resulted in a complex and shuffled pan-genome. The pangenome of a species, defined as the collection of all genes found in at least one strain of the species, consists of the core genome, which are genes found in >95% of strains from that species, and an accessory genome, comprised of genes that are found in at least one but less than 95% of strains for a species68,69. The composition of a bacterial pangenome has been shown to affect, and be affected by, the evolution of virulence within a bacterial species, as well as the abundance of virulence factors present in the bacteria20. The *E. coli* pangenome is heavily biased towards accessory genes, as estimates of the total number of non-prophage, non-transposase genes in the *E. coli* reservoir is estimated to be over 10,000, almost five times as many as are expected to constitute the core genome shared by all *E. coli* strains28,70. The UPEC genomes that have been sequenced thus far, such as the model strains 53671, CFT07372, and UTI8973, show similar patterns in pangenome composition. Additionally, like other *E. coli,* UPEC genomes contain a large number of accessory genes unique to specific strains, in part due to the prevalence of pathogenicity associated islands (PAIs) common to UPEC20,28,70. Despite the number of unique genes, members of the UPEC group have greater genomic similarity and are more genetically distinct, as a group, than other pathovars70, which indicates greater inter-group heterogeneity between pathovars and less intra-group diversity within UPEC. Taken together, these data indicate that, although the genomic content of the UPEC group is varied due to the presence and absence of accessory genes, they are generally similar in their total gene content. This is an indication that investigation of accessory genes is important in understanding phenotypic differences that exist between UPEC strains.

Although the UPEC group has a high degree of genetic similarity, a definitive set of virulence factors has yet to be defined. Many UPEC genotypes are capable of causing disease in the bladder and there is no single set of urovirulence factors28,32,74. Despite their variety, evidence suggests that the accumulation of virulence factors is non-random, as at least five virulence profiles can be delineated by analyzing the presence of known virulence factors and clade membership of UPEC strains75. This is an indication that there is a pattern of co-occurrence of virulence factors, despite the variety of UPEC genotypes. Analysis of these virulence factors has shown that many factors co-occur and display low levels of intra-group diversity, indicating that structured, though frequent, horizontal gene transfer of virulence genes76. This pattern mirrors the homologous recombination in core genes, which has been shown to be high in *E. coli* 28, which suggests that virulence factors move through *E. coli* populations through horizontal gene transfer and recombination. Additionally, although a definitive set of urovirulence genes has not been identified, evidence does show increased number of virulence factors is correlated with increased levels of extra-intestinal pathogenesis32, indicating that many genes may be necessary to cause disease in the bladder. These data are strong indicators that virulence gene networks, rather than single genes, define sets of virulent genotypes. As a result, single gene investigations may not capture a complete picture of UPEC pathogenicity32.

*Virulence factors, phylogeny, and phenotype.*

While great variety exists in UPEC genotypes, single genes, or even small sets of virulence genes encoding complete virulence factors, are not sufficient to cause disease by themselves32,77, which suggests that additional genetic factors are necessary for pathogenesis. Supporting this finding, several studies have shown that the genomic context of a urovirulence gene can modify the functional effect of the urovirulence gene27,78. In many cases, non-pathogenic and pathogenic bacteria contain similar sets of virulence factors23, as can be seen in commensal probiotic strain *E. coli* Nissle 1917 and the uropathogenic bacterium *E. coli* CFT07379, however, despite their similar gene content, these strains have widely different pathogenic potential. Additional evidence for the necessity of a proper genomic context for virulence gene pathogenicity comes phylogenetic analysis of the virulence genes. Virulence factors specific to pathogenic isolates are common in isolates from clades B2 and D and rare in other clades, indicating that they are ancestral to those clades B2 and D80. Furthermore, genomic hybridization shows a correlation between the presence and absence of specific gene content and the phylogenetic history of the core-genome of B2 isolates, indicating the co-evolution of the accessory and core genomes14 possibly through a process of “fine-tuning” 26. Taken together, these data indicate that the phenotypic effects of virulence genes is mediated by an interaction with the genomic milieu that has been fine-tuned by the evolutionary history of the strain, and that clades B2 and D may have the milieu most conducive to maximum virulence potential. Identification of the genetic factors, other than the accessory genes, that differ between clades B2 and D and other clades may reveal the context that enhances a virulent phenotype14

*Urovirulence factors and support for the coincidental pathogenesis hypothesis*

A number of models have been proposed to explain the evolution of virulence and population structure of UPEC, including the “source-sink” model81,82, which states that urovirulence is a result of repeated, *de novo* mutations that result in a trade-off of decreased fitness in the gut to increased fitness in the bladder, and the “coincidental pathogenesis” model14, which stipulates that extra-intestinal pathogenicity is a by-product of adaptation to the gut environment. While the source-sink model is supported mainly by the functional effects of mutations in the *fimH* gene83-86, several lines of evidence support the coincidental pathogenesis hypothesis. First, common extra-intestinal virulence genes have been found to affect the fitness of strains within the gut environment14. These virulence factors, such as hemolysin, type I fimbriae, and P fimbriae, are associated with persistence of E. coli in the gut of infants and adult women43,50. Additionally, in healthy women, dominant E. coli clones had higher urovirulence scores, defined as the presence of known urovirulence genes, than non-dominant clones, indicating that urovirulence factors helped mediate gut fitness31. This pattern is mirrored in resident and transient strains of *E. coli* in the gut, as resident strains were more likely than transient strains to present uropathogenic phenotypes87. Most convincingly, direct knockouts of urovirulence genes important in UTI progression have been found to affect gut fitness. For example, deletion of PAIs in the UPEC strain CFT073 reduces rate of intestinal colonization88. This pattern indicates that fitness in the bladder and fitness in the gut may be mediated by the same factors.

Secondly, *E. coli* strains that are dominant in the gut share a phylogenetic history with UPEC strains that dominate in the bladder. Persistent strains in the gut environment were statistically more likely to belong to the uropathogenic subgroup of clade B2, indicating a potential link between fitness in the gut, pathogenicity in the bladder, and clade membership14,89. Further, dominance of a B2 strain in the gut is correlated with both increased number of urovirulence factors in the dominant strain and reduced species richness in the gut habitat30,31. This suggests that more urovirulent strains, ones that with greater urovirulence gene content and pathogenicity-enhancing genomic context, are able to outcompete less urovirulent strains in the gut habitat, which may result in local extinction of those less virulent strains.

Finally, a number of genes have been found to have a presumed fitness cost when analyzed using *in vitro* models of the bladder habitat. In particular, the presence of PAIs in the UPEC strain CFT073 is linked to reduced growth rate in urine88, indicating that there may be genetic factors that are maintained in the population despite the fitness cost of these factors in the bladder environment. This is an indication of selection pressure in habitats outside of the bladder, which have maintained genes capable of pathogenesis in the bladder despite their fitness cost when grown in urine. Taken together these three lines of evidence suggest that UPEC virulence may be an accidental by-product of adaptation to the gut, as opposed to a phenotype selected for by adaptation to the bladder habitat.

**Future plans directions and unanswered questions**

Although the connection between gut populations of UPEC and UTI have been known for six decades, a number of questions remain unanswered regarding the connection between UPEC population structure, virulence factors, and the progression of UTIs, including:

* Is there an expansion of gut UPEC that coincides with the onset of a UTI episode?
* Does the relative abundance of known UPEC virulence genes change in the gut microbiome during a UTI episode?

Assessing these questions will aid our understanding of this important facultative pathogen by helping us to understand the population changes that occur in the gut during UTI episodes. Understanding these population changes will allow us to support or reject current models of virulence evolution, and help us better tailor therapeutic interventions prevent treatment failure and recrudescence.

**Experimental plans**

**Background:**

Uropathogenic *Escherichia* *coli* causes 80% of diagnosed urinary tract infections (UTIs)2, but are also known to live as commensal organisms in the human gut90. As yet, no investigation has defined the community structure of the gut microbiota or the relative abundance of UPEC in the gut during a UTI episode. Virulence factors that have been shown to be important for virulence in the bladder, such as hemolysin, P pilli, and enterobactin, have also been correlated with enhanced persistence over time in the gut78,89. Additionally, dominance within the subpopulation of *E. coli* in the gut has been correlated with the presence of known urovirulence genes30,31,54. With these factors in mind, ***I hypothesize that there is a bloom of uropathogenic E. coli in the gut that coincides with the colonization of the bladder by uropathogenic E. coli during the onset of an acute UTI.*** To test this hypothesis, I will compare the community structure of the gut microbiota during UTI and after recovery using a combination of metagenomic sequencing and targeted qPCR. Together, these methods will allow for an accurate measurement of the representation of UPEC in the gut microbiota during the onset of a UTI. Because the gut can act as a reservoir for UPEC4,51, understanding the community structure of this opportunistic pathogen in its reservoir will aid our understanding of UTI susceptibility and progression of the disease.

**Sub-aim1: Compare the community structure of the gut microbiota during acute UTI and after recovery.**

*Rationale:* While the human gut microbiome is generally stable through time55,91, changes in gut community structure between healthy and disease states have been identified92. Changes in the periurethral microbiota have also been found, as the prevalence of *E. coli* increases in the days preceding the onset of a UTI17. This is concordant with the rectal-perineal-urethral hypothesis, which states that the gut contains a reservoir of UPEC, which escape the gut and colonize the periurethral area before ascending into the bladder to cause a UTI51. Additionally, at least one population susceptible to UTIs, pregnant women93, undergoes shifts in the ecology of their gut microbiome during the course of pregnancy, including an expansion of proteobacteria94. Thus, ***I hypothesize that there is a greater representation of E. coli in the gut microbiota of a patient at the onset of a UTI than at times when the patient is healthy.*** To test this hypothesis, I will measure the relative abundance of *Escherichia* in fecal samples collected from patients during UTI and after their recovery. If overgrowth of the gut reservoir of UPEC coincides with the onset of a UTI, then the abundance of *Escherichia* will be higher in the patient during UTI than when the patient is healthy. This information will be very useful in explaining the relationship between microbiota and UTI susceptibility.

*Experimental methods:* In collaboration with Case Western University and their clinical facilities, women aged 18-41 years who present with symptoms of an uncomplicated cystitis will be open to enroll in a new cohort. Entrants will be required to meet certain criteria, including a lack of functional or anatomical abnormalities and time since antibiotic use, which will reduce the confounding factors in the analyses. Patients will supply three fecal samples, one at the time of entry, before the administration of antibiotics, one after completing the antibiotic regimen at 14 days after enrollment, and one at 28 days after enrollment. These samples will be labeled as UTI, Treated, and Recovered, respectively (Figure 2a) and will be stored in sterile, airtight containers at -20C until being returned to the clinic for further analysis, which are outlined below (Figure 2b). Urine will also be collected at enrollment in order to identify the bacteria causing the UTI. Fifty patients will be enrolled with the expectation that 30 will submit all the samples, be free from recurrent UTIs during the study, and have confirmed cases of UTI caused by UPEC. A cohort of 30 patients will give the study the power to detect an effect size of 1.40 using non-parametric t-tests (Figure 2c).

Genomic DNA from the fecal samples collected from cohort will be extracted and used as template to sequence the 16S rRNA gene using 454 pyrosequencing. Sequence data from this will then be used to estimate the community structure of the gut microbiota, as has been done previously91. Binning of the amplicons into phylotypes using the QIIME and RDP software packages will allow for the representation of those phylotypes in the gut microbiota to be measured at the genus level95. Statistical analysis will include application of the Shannon diversity index to identify intra- and inter-host differences in community structure, principle component analysis (PCA) to measure the tendency of the samples to cluster to their sample groups, and Mann-Whitney non-parametric t-tests to identify changes in the representation of *Escherichia* during the UTI and after recovery.

Antibiotic treatment has been shown to dramatically affect the composition of the gut microbiota. While the majority of taxons regenerate after 28 days

96-100, some genera may take months to reappear101. This will reduce the overall species richness of the gut microbiota, and may artificially reduce the divergence between the UTI and Recovered groups. As an additional control against this error, data on the gut community structure of healthy adult women available from the Human Microbiome Project (HMP) will be included in these analyses in order to identify differences between healthy microbiomes and microbiomes that may be altered after antibiotic treatment, as has been done in other studies94.

*Expected results:* Analysis using the Shannon diversity index will measure the differences between the sample groups and will show that the samples from the different groups will show more inter-group diversity than intra-group diversity, indicating that the groups have shared features of their community structure that differ between the groups. This analysis will be supported by PCA which will show that communities from the Recovered group will cluster with the data from the HMP, while the UTI group and the Treated group will each cluster separately. Finally, the Mann-Whitney t-test will show that the UTI group will have a statistically significant higher representation of *Escherichia* than the Recovered group, which is an indication that gut *E. coli* carriage is higher during UTI than after recovery. Taken together, these data will show that the gut microbiota during UTI is in an altered state distinct from a healthy state, and that the abundance of *Escherichia* is higher in this altered state than in the healthy state.

*Anticipated challenges:* Recurrence of a UTI episode within 6-12 months after an initial UTI caused by UPEC occurs in approximately 25-30% of cases. The chance of recurrence is greatest within the first 3 months following the initial UTI, and is often caused by the same UPEC strain as the one that caused the initial UTI102. The chance of another UTI episode caused by the same strain drops after 6 months. This indicates that there may be a short-lived reservoir of UPEC in the patient that survives antibiotic treatment and immune system clearance. This reservoir may be located in the gut, which would complicate analysis of the gut carriage of UPEC during an acute UTI episode. As a result, any patients who self-reported another UTI episode within the 6 months following enrollment in the cohort will be excluded from the final analyses. This self-reporting is cost-effective, non-invasive, and will reduce the confounding influence of recurrent UTIs on this analysis.

**Sub-aim 2: Compare the relative abundance of urovirulence genes in the gut microbiome during acute UTI and after recovery.**

*Rationale:* It is possible that there are members of the *Escherichia* that could have increased representation in the gut at the onset of a UTI episode due to host factors, such as TLR4 polymorphisms. As such, the relative abundance of the UPEC in the gut needs to be measured using techniques with greater specificity than metagenomic sequencing. Interestingly, the dominant strain of *E. coli* in the gut has, on average, a greater number of urovirulence genes than minor strains in the gut30,31,54; however, the relative representation of urovirulence genes in the gut microbiome during UTI and in a healthy state within the same patient has yet to be compared. While urovirulence genes can be found in non-pathogenic strains, they are much more represented in UPEC strains than other strains of bacteria76. Given the connection between urovirulence factors and gut fitness78,89 and their abundance of UPEC strains, ***I hypothesize that there is a greater relative abundance of known urovirulence genes in the gut microbiome of a patient at the onset of a UTI than when the person is healthy.*** While a bloom in the population of UPEC in the gut would increase the relative abundance of urovirulence genes, it is also possible that an increase in the relative abundance of *E. coli* in the gut microbiota is not needed for an increase in the relative abundance of UPEC urovirulence genes because an increased number of urovirulence genes in the dominant strain of *E. coli* is correlated with reduced *E. coli* strain diversity in the gut30,31. In this situation, a single strain of UPEC carrying many urovirulence genes may outcompete other *E. coli* species in the gut, resulting in their local extinction without changing the total abundance of *E. coli* in the gut. In this scenario, the relative abundance of urovirulence genes may increase without affecting the relative abundance of *E. coli* in the gut microbiota. No single set of urovirulence genes necessary and sufficient for pathogenesis has been defined32, therefore, many urovirulence genes must be assayed to measure the prevalence and abundance of UPEC. Thus, in order to measure the relative abundance of urovirulence genes in the gut microbiome, I will use a series of quantitative polymerase chain reactions (qPCRs) targeting known urovirulence genes and compare the results to qPCRs targeting housekeeping genes. If there is an increase in the relative abundance of urovirulence genes in the gut microbiome at the onset of a UTI, then the abundance of those genes will be higher at the onset of UTI relative to housekeeping genes common to all microbial taxa. This information is critical to understanding the population dynamics of UPEC in the gut, as the current definitions of UPEC rely extensively on the presence or absence of known urovirulence factors24,25.

*Experimental methods:* Genomic DNA extracted in sub-aim 1 will be used as template for qPCRs targeting known urovirulence genes from gene clusters listed in Table 1. Additional qPCRs will be performed that target a conserved region of the 16S rRNA gene common to all bacteria as well as a divergent region of the single copy housekeeping gene, *rpoB*, which is unique in *E. coli*.

Comparison of the qPCR results from the urovirulence genes to the qPCR results from the 16S rRNA and *rpoB* genes can be used to estimate the relative abundance of the urovirulence genes to the total gut microbiota population and the *E. coli* sub-population, respectively. Mann-Whitney non-parametric t-tests will be used to identify statistically significant differences between sample groups in urovirulence gene abundance relative to the total microbiota and to E. coli specifically.

*Expected results:* In healthy adults, *E. coli* are found at an abundance of >0.1% in 15% of microbiomes and are detectable in a total of 61% of fecal samples91. Thus, given the low representation of *E. coli* in the gut, I expect the abundance of UPEC urovirulence genes to be very low compared to the abundance of the bacterial 16S rRNA genes in the gut microbiota. However, in samples collected at the onset of a UTI episode, I expect that the proportion of urovirulence genes relative to both the *E. coli* *rpoB* gene and the bacterial 16S gene to be significantly higher in the UTI sample set as compared to the Recovered set. Taken together, these data will define the reservoir of UPEC urovirulence genes that exist in the gut microbiome during the onset of a UTI episode.

*Anticipated challenges*: Given the low abundance of *E. coli* in the healthy genome, standard qPCR may not be able to detect urovirulence genes accurately given its technical limitations, such as off-target priming and analog output. Digital PCR103 has been used in the place of qPCR to detect and quantify the presence of pathogens, such as Human Immunodeficiency Virus due to its lower limit of detection and digital output. If qPCR fails to identify the relative abundance of targeted urovirulence genes, then Digital PCR can be used to determine the abundance of urovirulence genes or confirm their absence. This information can be substituted for the relative abundance of known urovirulence genes, and instead be used to quantify the number of copies of known urovirulence genes per copy of the *E. coli rpoB* gene or 16S rRNA gene. Both qPCR and Digital PCR will enable the comparison of urovirulence gene abundance at the onset of UTI to the abundance at after recovery, providing a necessary insight into the repertoire of genes available to UPEC.

**Conclusions:** Description of the gut community structure and abundance of urovirulence genes at the onset of acute UTI will aid in our understanding of the reservoir of UPEC that exists in the gut. Population dynamics in this reservoir may mediate susceptibility to acute UTI; however, relatively little attention has been paid to this sub-population of UPEC. The identification of a bloom in the gut population of *E. coli* and an increase in the representation of urovirulence genes in the gut microbiome that precedes the onset of a UPEC UTI would be significant support to the rectal-perineal-urethral hypothesis that UPEC originating in the gut are responsible for UTIs. This hypothesis is a significant insight into disease progression – and an important step in learning to control the diseases.

**Literature Cited**

1. Foxman, B. & Brown, P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. *Infect. Dis. Clin. North Am.* **17**, 227–241 (2003).

2. Foxman, B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am. J. Med.* **113 Suppl 1A**, 5S–13S (2002).

3. Foxman, B. *et al.* Risk factors for second urinary tract infection among college women. *American Journal of Epidemiology* **151**, 1194–1205 (2000).

4. Hooton, T. M. Recurrent urinary tract infection in women. *International Journal of Antimicrobial Agents* **17**, 259–268 (2001).

5. Vollmerhausen, T. L. *et al.* Population structure and uropathogenic virulence-associated genes of faecal Escherichia coli from healthy young and elderly adults. *Journal of Medical Microbiology* **60**, 574–581 (2011).

6. Ronald, A. The etiology of urinary tract infection: traditional and emerging pathogens. *Am. J. Med.* **113 Suppl 1A**, 14S–19S (2002).

7. Foxman, B. The epidemiology of urinary tract infection. *Nature Publishing Group* **7**, 653–660 (2010).

8. Stamm, W. E. Scientific and clinical challenges in the management of urinary tract infections. *Am. J. Med.* **113 Suppl 1A**, 1S–4S (2002).

9. LITWIN, M. *et al.* UROLOGIC DISEASES IN AMERICA PROJECT: ANALYTICAL METHODS AND PRINCIPAL FINDINGS. *J. Urol.* **173**, 933–937 (2005).

10. Russo, T. & Johnson, J. R. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes and Infection* **5**, 449–456 (2003).

11. Schappert, S. M. & Rechtsteiner, E. A. Ambulatory medical care utilization estimates for 2007. *Vital Health Stat 13* 1–38 (2011).

12. Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. Urinary tract infection: self-reported incidence and associated costs. *Ann Epidemiol* **10**, 509–515 (2000).

13. Hooton, T. M. Clinical practice. Uncomplicated urinary tract infection. *N. Engl. J. Med.* **366**, 1028–1037 (2012).

14. Le Gall, T. *et al.* Extraintestinal Virulence Is a Coincidental By-Product of Commensalism in B2 Phylogenetic Group Escherichia coli Strains. *Mol. Biol. Evol.* **24**, 2373–2384 (2007).

15. Denamur, E., Picard, B. & Tenaillon, O. Population genetics of pathogenic Escherichia coli. *Bacterial Population Genetics in Infectious Disease* 269–286 (2010).

16. Stamey, T. A., Timothy, M., Millar, M. & Mihara, G. Recurrent urinary infections in adult women. The role of introital enterobacteria. *Calif Med* **115**, 1–19 (1971).

17. Czaja, C. A. *et al.* Prospective Cohort Study of Microbial and Inflammatory Events Immediately Preceding Escherichia coli Recurrent Urinary Tract Infection in Women. *J. Infect. Dis.* **200**, 528–536 (2009).

18. Hannan, T. J. *et al.* Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic Escherichia coli bladder infection. *FEMS Microbiol Rev* **36**, 616–648 (2012).

19. Svanborg, C. *et al.* Uropathogenic Escherichia coli as a model of host-parasite interaction. *Current Opinion in Microbiology* **9**, 33–39 (2006).

20. Dobrindt, U., Chowdary, M. G., Krumbholz, G. & Hacker, J. Genome dynamics and its impact on evolution of Escherichia coli. *Med Microbiol Immunol* **199**, 145–154 (2010).

21. Darfeuille-Michaud, A. Adherent-invasive Escherichia coli: a putative new E. coli pathotype associated with Crohn's disease. *International Journal of Medical Microbiology* **292**, 185–193 (2002).

22. Nataro, J. P. & Kaper, J. B. Diarrheagenic Escherichia coli. *Clinical microbiology reviews* (1998).

23. Köhler, C.-D. & Dobrindt, U. What defines extraintestinal pathogenic Escherichia coli? *Int. J. Med. Microbiol.* **301**, 642–647 (2011).

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

25. Russo, T. A. & Johnson, J. R. Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of Escherichia coli: ExPEC. (2000).

26. Escobar-Páramo, P. *et al.* A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in Escherichia coli. *Molecular biology and …* (2004).

27. Escobar-Páramo, P. *et al.* Large-Scale Population Structure of Human Commensal Escherichia coli Isolates. *Applied and …* (2004).

28. Touchon, M. *et al.* Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet.* **5**, e1000344 (2009).

29. Zhang, L., Foxman, B. & Marrs, C. Both urinary and rectal Escherichia coli isolates are dominated by strains of phylogenetic group B2. *Journal of Clinical Microbiology* **40**, 3951–3955 (2002).

30. Moreno, E. *et al.* Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. *Journal of Clinical Microbiology* **46**, 2529–2534 (2008).

31. Moreno, E. *et al.* Structure and urovirulence characteristics of the fecal Escherichia coli population among healthy women. *Microbes Infect.* **11**, 274–280 (2009).

32. Picard, B. *et al.* The link between phylogeny and virulence in Escherichia coli extraintestinal infection. *Infection and Immunity* **67**, 546–553 (1999).

33. Nickel, J. C. *Management of urinary tract infections: historical perspective and current strategies: Part 1--Before antibiotics.* *J. Urol.* **173**, 21–26 (2005).

34. Ferry, S. A., Holm, S. E., Stenlund, H., Lundholm, R. & Monsen, T. J. The natural course of uncomplicated lower urinary tract infection in women illustrated by a randomized placebo controlled study. *Scand. J. Infect. Dis.* **36**, 296–301 (2004).

35. Gupta, K., Hooton, T. M. & Stamm, W. E. Isolation of fluoroquinolone-resistant rectal Escherichia coli after treatment of acute uncomplicated cystitis. *J. Antimicrob. Chemother.* **56**, 243–246 (2005).

36. Karlowsky, J. A., Hoban, D. J., Decorby, M. R., Laing, N. M. & Zhanel, G. G. Fluoroquinolone-resistant urinary isolates of Escherichia coli from outpatients are frequently multidrug resistant: results from the North American Urinary Tract Infection Collaborative Alliance-Quinolone Resistance study. *Antimicrob. Agents Chemother.* **50**, 2251–2254 (2006).

37. Schwartz, D. J., Chen, S. L., Hultgren, S. J. & Seed, P. C. Population Dynamics and Niche Distribution of Uropathogenic Escherichia coli during Acute and Chronic Urinary Tract Infection. *Infection and Immunity* **79**, 4250–4259 (2011).

38. Walters, M. S. *et al.* Kinetics of uropathogenic Escherichia coli metapopulation movement during urinary tract infection. *mBio* **3**, (2012).

39. Mulvey, M. A. *et al.* Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. *Science* **282**, 1494–1497 (1998).

40. Anderson, G. G. Intracellular Bacterial Biofilm-Like Pods in Urinary Tract Infections. *Science* **301**, 105–107 (2003).

41. Justice, S. S. *et al.* Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1333–1338 (2004).

42. Wright, K. J., Seed, P. C. & Hultgren, S. J. Development of intracellular bacterial communities of uropathogenic Escherichia colidepends on type 1 pili. *Cell. Microbiol.* **9**, 2230–2241 (2007).

43. Nowrouzian, F. *et al.* Escherichia coli in Infants' Intestinal Microflora: Colonization Rate, Strain Turnover, and Virulence Gene Carriage. *Pediatr Res* **54**, 8–14 (2003).

44. Sears, H. J., Brownlee, I. & Uchiyama, J. K. Persistence of individual strains of Escherichia coli in the intestinal tract of man. *Journal of Bacteriology* **59**, 293–301 (1950).

45. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the …* (2011).

46. Sears, H. J. & Brownlee, I. Further observations on the persistence of individual strains of Escherichia coli in the intestinal tract of man. *Journal of Bacteriology* **63**, 47–57 (1952).

47. Bettelheim, K. A., Faiers, M. & Shooter, R. A. SEROTYPES OF ESCHERICHIA COLI IN NORMAL STOOLS. *The Lancet* **300**, 1224–1226 (1972).

48. Caugant, D. A., Levin, B. R. & Selander, R. K. Genetic diversity and temporal variation in the E. coli population of a human host. *Genetics* **98**, 467–490 (1981).

49. Schlager, T. A., Hendley, J. O., Bell, A. L. & Whittam, T. S. Clonal diversity of Escherichia coli colonizing stools and urinary tracts of young girls. *Infection and Immunity* **70**, 1225–1229 (2002).

50. Schlager, T. A. *et al.* Variation in Frequency of the Virulence‐Factor Gene in Escherichia coliClones Colonizing the Stools and Urinary Tracts of Healthy Prepubertal Girls. *J. Infect. Dis.* **188**, 1059–1064 (2003).

51. Yamamoto, S. *et al.* Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. *J. Urol.* **157**, 1127–1129 (1997).

52. Grüneberg, R. N. Relationship of infecting urinary organism to the faecal flora in patients with symptomatic urinary infection. *Lancet* **2**, 766–768 (1969).

53. Stapleton, A., Hooton, T. M., Fennell, C., Roberts, P. L. & Stamm, W. E. Effect of secretor status on vaginal and rectal colonization with fimbriated Escherichia coli in women with and without recurrent urinary tract infection. *J. Infect. Dis.* **171**, 717–720 (1995).

54. Moreno, E. *et al.* Relationship between Escherichia coli strains causing urinary tract infection in women and the dominant faecal flora of the same hosts. *Epidemiol. Infect.* **134**, 1015–1023 (2006).

55. Schloissnig, S. *et al.* Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2012).

56. Turck, M. & Petersdorf, R. G. THE EPIDEMIOLOGY OF NON-ENTERIC ESCHERICHIA COLI INFECTIONS: PREVALENCE OF SEROLOGICAL GROUPS\*. *J. Clin. Invest.* **41**, 1760–1765 (1962).

57. Plos, K. *et al.* Intestinal carriage of P fimbriated Escherichia coli and the susceptibility to urinary tract infection in young children. *J. Infect. Dis.* **171**, 625–631 (1995).

58. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449**, 835–842 (2007).

59. Hultgren, S. J., Porter, T. N., Schaeffer, A. J. & Duncan, J. L. Role of type 1 pili and effects of phase variation on lower urinary tract infections produced by Escherichia coli. *Infection and Immunity* **50**, 370–377 (1985).

60. Zhou, G. *et al.* Uroplakin Ia is the urothelial receptor for uropathogenic Escherichia coli: evidence from in vitro FimH binding. *Journal of Cell Science* **114**, 4095–4103 (2001).

61. Lane, M. C. & Mobley, H. Kidney International - Abstract of article: Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic Escherichia coli (UPEC) in the mammalian kidney. *Kidney international* (2007).

62. Castelain, M., Sjöström, A. E., Fällman, E., Uhlin, B. E. & Andersson, M. Unfolding and refolding properties of S pili on extraintestinal pathogenic Escherichia coli. *Eur. Biophys. J.* **39**, 1105–1115 (2010).

63. Das, M. *et al.* Hydrophilic domain II of Escherichia coli Dr fimbriae facilitates cell invasion. *Infection and Immunity* **73**, 6119–6126 (2005).

64. Nielubowicz, G. R. & Mobley, H. L. T. Host–pathogen interactions in urinary tract infection. *Nature Publishing Group* **7**, 430–441 (2010).

65. Wiles, T. J., Kulesus, R. R. & Mulvey, M. A. Origins and virulence mechanisms of uropathogenic Escherichia coli. *Experimental and Molecular Pathology* **85**, 11–19 (2008).

66. Johnson, T. J., Wannemuehler, Y. M. & Nolan, L. K. Evolution of the iss gene in Escherichia coli. *Appl. Environ. Microbiol.* **74**, 2360–2369 (2008).

67. Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**, 215–237 (2003).

68. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13950–13955 (2005).

69. Medini, D. *et al.* Microbiology in the post-genomic era. *Nature Reviews Microbiology* (2008).doi:10.1038/nrmicro1901

70. Rasko, D. A. *et al.* The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates. *Journal of Bacteriology* **190**, 6881–6893 (2008).

71. Brzuszkiewicz, E. *et al.* How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic Escherichia coli strains. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12879–12884 (2006).

72. Welch, R. A. *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17020–17024 (2002).

73. Chen, S. L. *et al.* Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5977–5982 (2006).

74. Norinder, B. S., Köves, B., Yadav, M., Brauner, A. & Svanborg, C. Do Escherichia coli strains causing acute cystitis have a distinct virulence repertoire? *Microbial Pathogenesis* **52**, 10–16 (2012).

75. Poey, M. E., Albini, M., Saona, G. & Laviña, M. Virulence profiles in uropathogenic Escherichia coli isolated from pregnant women and children with urinary tract abnormalities. *Microbial Pathogenesis* **52**, 292–301 (2012).

76. Johnson, J. R., Delavari, P., Kuskowski, M. & Stell, A. L. Phylogenetic distribution of extraintestinal virulence-associated traits in Escherichia coli. *J. Infect. Dis.* **183**, 78–88 (2001).

77. Marrs, C. F., Zhang, L. & Foxman, B. Escherichia coli mediated urinary tract infections: are there distinct uropathogenic E. coli (UPEC) pathotypes? *FEMS Microbiol. Lett.* **252**, 183–190 (2005).

78. Nowrouzian, F. L., Wold, A. E. & Adlerberth, I. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **191**, 1078–1083 (2005).

79. Grozdanov, L. *et al.* Analysis of the Genome Structure of the Nonpathogenic Probiotic Escherichia coli Strain Nissle 1917. *Journal of Bacteriology* **186**, 5432–5441 (2004).

80. Boyd, E. F. & Hartl, D. L. Chromosomal regions specific to pathogenic isolates of Escherichia coli have a phylogenetically clustered distribution. *Journal of Bacteriology* **180**, 1159–1165 (1998).

81. Sokurenko, E. V. *et al.* Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. *Proceedings of the …* (1998).

82. Sokurenko, E. V., Gomulkiewicz, R. & Dykhuizen, D. E. Source–sink dynamics of virulence evolution. *Nature Reviews Microbiology* **4**, 548–555 (2006).

83. Sokurenko, E. V., Courtney, H. S., Ohman, D. E., Klemm, P. & Hasty, D. L. FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. *Journal of Bacteriology* **176**, 748–755 (1994).

84. Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. & Hasty, D. L. Quantitative differences in adhesiveness of type 1 fimbriated Escherichia coli due to structural differences in fimH genes. *Journal of …* (1995).

85. Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7**, 191–195 (1999).

86. Weissman, S. J. *et al.* Differential stability and trade-off effects of pathoadaptive mutations in the Escherichia coli FimH adhesin. *Infection and Immunity* **75**, 3548–3555 (2007).

87. Wold, A. E., Caugant, D. A., Lidin-Janson, G., de Man, P. & Svanborg, C. Resident colonic Escherichia coli strains frequently display uropathogenic characteristics. *J. Infect. Dis.* **165**, 46–52 (1992).

88. Diard, M. *et al.* Pathogenicity-Associated Islands in Extraintestinal Pathogenic Escherichia coli Are Fitness Elements Involved in Intestinal Colonization. *Journal of Bacteriology* **192**, 4885–4893 (2010).

89. Nowrouzian, F. L., Adlerberth, I. & Wold, A. E. 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 Infect.* **8**, 834–840 (2006).

90. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal Escherichia coli. *Nature Reviews Microbiology* **8**, 207–217 (2010).

91. Consortium, T. H. M. P. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).

92. Blumberg, R. & Powrie, F. Microbiota, Disease, and Back to Health: A Metastable Journey. *Science Translational Medicine* **4**, 137rv7–137rv7 (2012).

93. Law, H. & Fiadjoe, P. Urogynaecological problems in pregnancy. *J Obstet Gynaecol* **32**, 109–112 (2012).

94. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012).

95. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).

96. Trinidad, N. J., Ayres, J. S. & Vance, R. E. Lethal inflammasome activation by a multidrug-resistant pathobiont upon antibiotic disruption of the microbiota. *Nature Medicine* **18**, 799–806 (2012).

97. Ubeda, C. *et al.* Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120**, 4332–4341 (2010).

98. Hill, D. A. *et al.* Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. *Mucosal Immunol* **3**, 148–158 (2010).

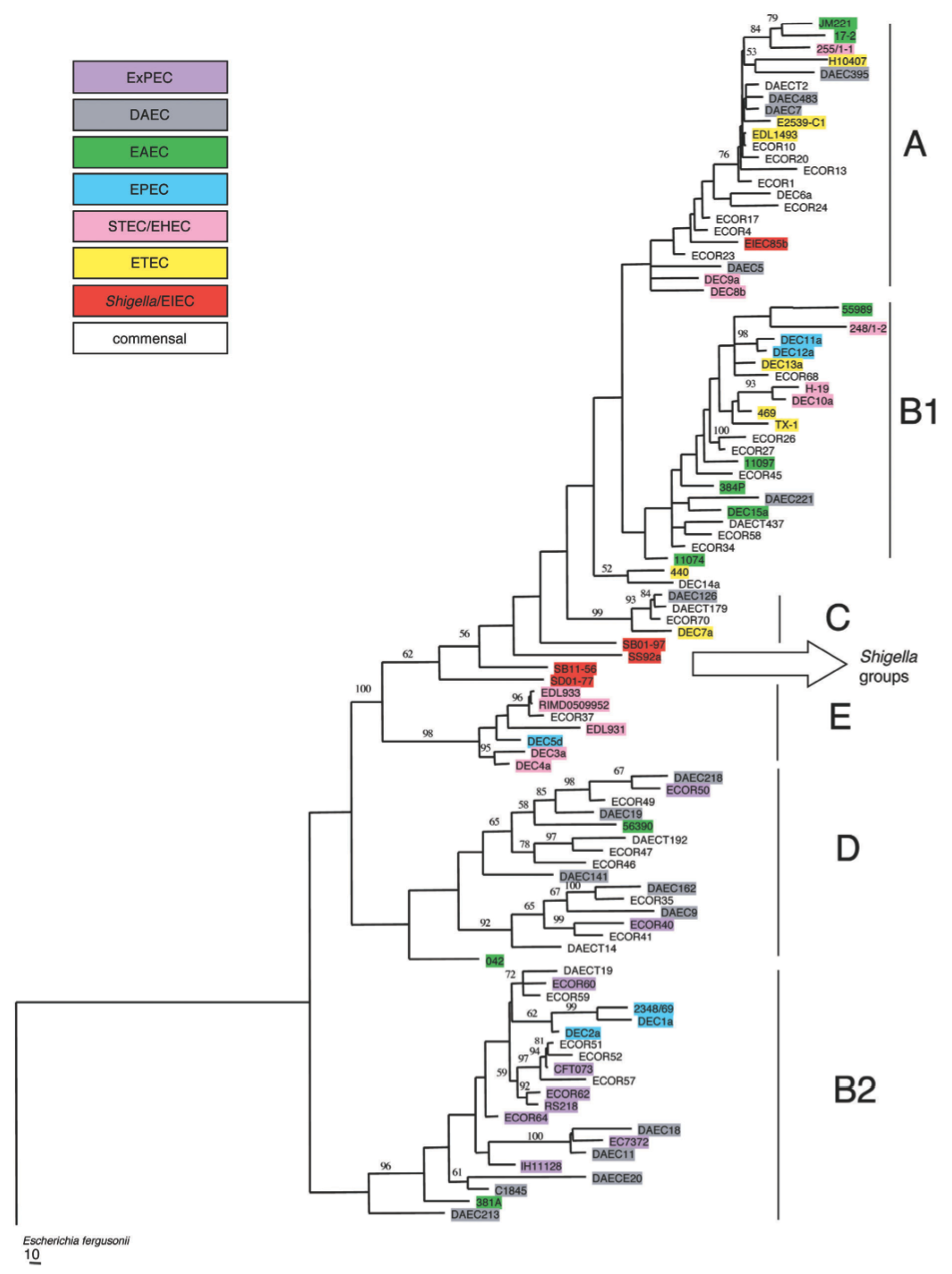
99. Croswell, A., Amir, E., Teggatz, P., Barman, M. & Salzman, N. H. Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric Salmonella infection. *Infection and Immunity* **77**, 2741–2753 (2009).

100. Sekirov, I. *et al.* Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infection and Immunity* **76**, 4726–4736 (2008).

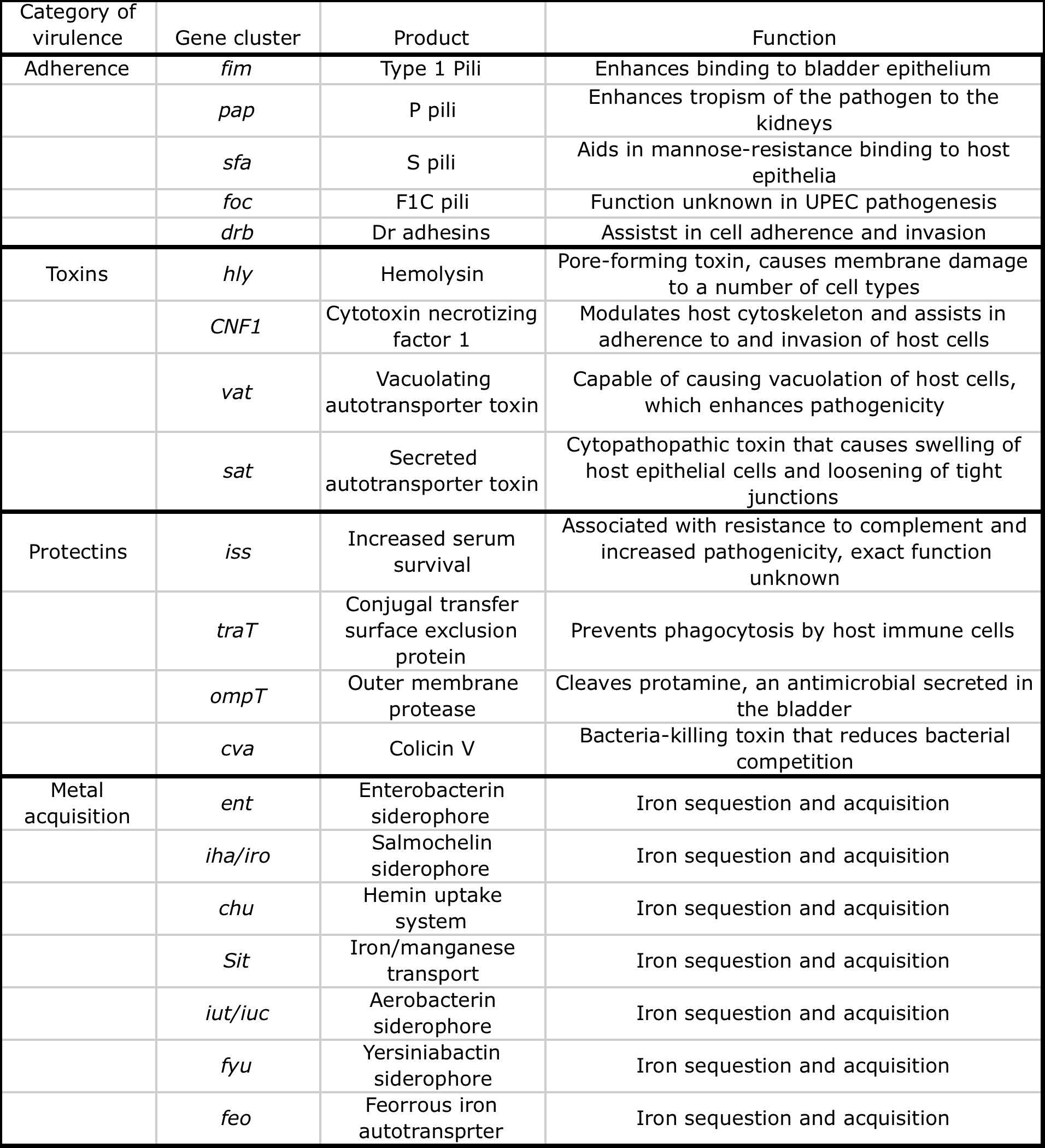
101. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *Plos Biol* **6**, e280 (2008).

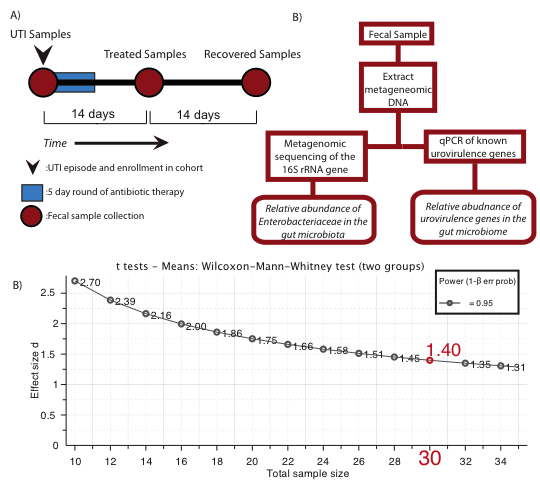
102. Foxman, B. *et al.* Virulence characteristics of Escherichia coli causing first urinary tract infection predict risk of second infection. *J. Infect. Dis.* **172**, 1536–1541 (1995).

103. Vogelstein, B. & Kinzler, K. W. Digital PCR. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9236–9241 (1999).

**Figure 1.** Semistrict consensus tree built using six housekeeping genes (trpA, trpB, pabB, putP, icd, and polB) and the parsimony method. Bootstrap values above 50% are indicated above the nodes. Vertical bars and labels indicate major lineages of *Escherichia coli*. UPEC isolates are located exclusively in the B2 and D clades and are absent from the others. Adapted from Escobar –Páramo, 2004b.

**Table 2.** Virulence factors found in UPEC can be categorized into 4 groups based on their general function. List adapted from Wiles *et al.*, 2012, Nielubowicz and Mobley, 2010, Kohler *et al.*, 2011, Luo et al. 2012.





**Figure 2. A)** Diagram of the sample collection procedure. Three samples will be collected, one at enrollment, one at 14 days post enrollment, and one at 28 days post enrollment. A standard five day course of antibiotics will be given between the first two collections to treat the urinary tract infection (UTI). **B)** Outline of experimental design for sub-aims. **C)** Power analysis curve constructed using G\*Power. The power, or one minus the rate of a Type II error, of a Mann-Whitney t-test was plotted using the total sample size and effect size, d, to calculate the effect size of a cohort consisting of 30 patients.