WHAT MAKES AN ORGANISM PATHOGENIC? A STUDY OF THE GUT COMMENSAL AND BLADDER PATHOGEN ESCHERICHIA COLI

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

Approximately 50% of women will suffer a urinary tract infection (UTI) by the age of 32{Foxman:2003uo} and up to 15 million cases of UTI occur each year{Foxman:2002va}. 20-30% of these women will suffer a recurrent UTI within three to four months following an initial UTI {Foxman:2000wu, Foxman:2002va, Hooton:2001dp}. The rates of UTI increase in the immunosuppressed, including the elderly and children, and may result in significant complications, including renal scarring, septicemia, and pyelonephritis{Vollmerhausen:2011fe, Foxman:2002va}. UTIs are commonly acquired in the community, but are also the most common nosocomial infection{Ronald:2002tn, Foxman:2003uo, Foxman:2010bx}. UTIs are responsible for over 8.4 million doctors visits in 2007 and more than $2.5 billion in direct costs per annum, and have been increasing in costs{Stamm:2002vf, LITWIN:2005cr, Russo:2003bc, Schappert:2011te, Foxman:2000tl, Foxman:2002va}. Clinical manifestations of UTIs include dysuria, foul-smelling or cloudy urine, fever, and flank pain{Hooton:2012hb, Stamm:2002vf, Schappert:2011te}. 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 a catheter or other instrument installed; all other cases of UTIs are considered complicated{Foxman:2010bx, Hooton:2012hb}. This review will focus on uncomplicated UTIs that are acquired in the community.

Nearly 80% of community acquired UTIs are caused by uropathogenic *E. coli* (UPEC){Stamm:2002vf}, while the remaining 20% are mainly caused by other Enterobacteriaceae, such as *Klebsiella* and *Enterobacter*, as well as Gram-positive rods, such as *Staphylococcus*{Ronald:2002tn}. Despite its prevalence and pathogenicity, UPEC is considered a facultative pathogen, as are many other types of *E. coli*{LeGall:2007bq, Denamur:2010ug}. Facultative pathogens, such as the deadly 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 disease{Denamur:2010ug}. UPEC is known to exist in the human gut as a commensal bacterium, but it 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 periurethra and vagina without causing disease, and these habitats have been suggested to be potential reservoirs that are capable of invading the urinary tract{Stamey:1971wz, Czaja:2009jx}.

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 expression, epithelial cell invasion, toxin production, and population bottlenecks, in addition to its obvious use as a model for uropathogenicity{Hannan:2012jh, Svanborg:2006it, Dobrindt:2010fe}. 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 the 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 infections{Svanborg:2006it}.

**The population structure of UPEC – a tale of two homes**

*Classification of Escherichia coli isolates*

*Escherichia coli* are associated with a number clinical conditions, each caused by collection of *E. coli* strains harboring different repertoires of gene sets and virulence factors, and, therefore, can been 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){Nataro:1998uo, DarfeuilleMichaud:2002dx, Kaper:2004bm: Kohler:2011cn}. 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 {Russo:2000vr}. 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 from Escobar here). Currently, four main clades of *E. coli* have been described, A, B1, B2, and D, along with two smaller clades, C and E{EscobarParamo:2004to, EscobarParamo:2004up, Touchon:2009kw}. ExPEC fall predominately into clade B2, and to a lesser extent D and are generally absent from other clades{EscobarParamo:2004to} and the majority of urine isolates of E. coli are from clade B2 {Zhang:2002wo, Moreno:2008eg, Moreno:2009jc}. Clade B2 can be further subdivided into 9 sub-clades, of which several are correlated with increased urpathogenicity{LeGall:2007bq}. Although there appears to be a connection between phylogeny and virulence, UPEC strains have been isolated from clades A, B1, B2, and D {Picard:1999uk, EscobarParamo:2004to}. 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 potential{LeGall:2007bq, EscobarParamo:2004up, EscobarParamo:2004to, Picard:1999uk}.

*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 symptoms{Nickel:2005ke}, 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 weeks{Ferry:2004th}. 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 resistance{Gupta:2005jo} and given rise to multidrug resistant isolates{Karlowsky:2006jt}. 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 epithelium{Hannan:2012jh} and result in a drift to clonality within the bladder{Schwartz:2011cy, Walters:2012hq}. These bottlenecks recur many times during the cyclical course of UTI. 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 UTIs{Mulvey:1998wv, Anderson:2003kb, Justice:2004gx}. Although IBCs allow for significant clonal expansion of UPEC{Anderson:2003kb}, formation of the IBCs occurs at a very low rate, with only 50-700 IBCs persisting at 6h after inoculation of 107 UPEC bacteria{Schwartz:2011cy}. 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 factors (Figure 2){Hannan:2012jh}. Formation of these IBCs requires known virulence factors, including the adhesin *fimH*{Wright:2007ha}. 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 UTI{Hannan:2012jh}. 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 UTI{Hannan:2012jh}. 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.

The population structure of UPEC changes during recurrent episodes of UTI, which are commonly defined as additional UTI that occur within 6 months of an initial UTI episode{Foxman:2000wu, Foxman:2010bx, Hooton:2012hb}. UTIs may recur through a number of ways, including recrudescence via treatment failure{Gupta:2005jo}, re-emergence of the strain from quiescent intra-cellular reservoirs{Mysorekar:2006ja}, or re-invasion of the bladder by the same or different strains of UPEC. Estimates of the percentage of recurrent UTIs caused by this reinvasion by different strains vary greatly between studies from 16% to 82% {Czaja:2009jx, Ejrnaes:2006vr, Foxman:2000wu, Foxman:1995ug, Ikaheimo:1996ue, Karkkainen:2000iv, Koljalg:2009ud, Kunin:1970wq, Luo:2012bm, McGeachie:1966us, Pryles:1965wj, Russo:1995ts, SkjotRasmussen:2011ce, Winberg:1974dj}. This range may result from differences in the demographics of the cohort, urine collection methods, definition of significant bactiuria, definition of symptoms, treatment regimes, length of study, and, perhaps most importantly, strain typing methods which differ between the studies. As expected, recurrences that occur closer to the initial UTI are more likely to be caused by the same strain{Ikaheimo:1996ue}. Interestingly, virulence factors have been shown to affect the population structure of UPEC during recurrent UTI episodes, and strains with greater urovirulence scores, as measured by the presence of urovirulence genes, have been shown to be more likely to persist and cause subsequent UTIs while strains with lower urovirulence scores are more likely to be replaced during reinvasion {Luo:2012bm}. The function of these virulence factors also appears to be important, as some virulence factors are more associated with increased rates of recurrence than others{Foxman:2000wu, Foxman:1995ug, Ikaheimo:1996ue, Karkkainen:2000iv, SkjotRasmussen:2011ce, Luo:2012bm}. This is additional evidence supporting a link between urovirulent gene content in UPEC and population dynamics in the bladder.

*Escherichia coli population structure in the gut*

*E. coli* are some of the first bacteria to colonize the gut{Nowrouzian:2003bs, Sears:1950uk}, although they become less abundant as the gut microbiome matures{Koenig:2011up}. In adults, the gut population of *E. coli* is comprised of a dominant strain that accounts for the majority *E. coli* in the gut, and a handful of minor strains that contribute the remainder{Sears:1950uk, Sears:1952vw, Bettelheim:1972dq, Caugant:1981tm, Moreno:2009jc, Schlager:2002we, Schlager:2003hh}. 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 month{Sears:1950uk, Sears:1952vw, Caugant:1981tm, Bettelheim:1972dq}; 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 weeks{Schlager:2002we, Schlager:2003hh}.

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 typing{Moreno:2008eg, Moreno:2009jc}. 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 periurethra is known to occur during the days preceding a UTI{Czaja:2009jx}. Given the connection between the gut and the periurethra, 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 UTIs{Yamamoto:1997wk, Gruneberg:1969wo, Stamey:1971wz}. In this model, UPEC originate in the gut habitat and are colonize to periurethra 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 periurethra, or both. During acute UTI, several studies have shown that the strains isolated from the urine are found to be the dominate 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 cases{Gruneberg:1969wo, Yamamoto:1997wk, Moreno:2006ji, Moreno:2008eg, Czaja:2009jx}. These findings are in contrast to two studies indicating that the UPEC strain causing the UTI may not be found regularly in the feces or periurethra{Schlager:2002we, Schloissnig:2012hx}, 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 gut{Turck:1962do} while the “special pathogenicity” hypothesis stipulates that the presence of virulence factors mediate the success of bladder colonization{Plos:1995wd}. 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 associated{Moreno:2006ji, Moreno:2008eg, Moreno:2009jc}. 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 colonization{Moreno:2006ji, Moreno:2008eg}. 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 periurethra 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 periurethra in the same week{Schlager:2002we, Schlager:2003hh}. 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 – multi-purpose tools**

*UPEC genotypes are varied, but structured*

The pangenome of a species, defined as the collection of all genes found in at least one strain of the species, is comprised of the core genome, 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 species{Tettelin:2005jg, Medini:2008gi}. The composition of a bacterial pangenome has been shown to affect the evolution of virulence within a bacterial species, as well as the abundance of virulence factors present in the bacteria{Dobrindt:2010fe}. 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* strains{Rasko:2008bx, Touchon:2009kw}. The UPEC genomes that have been sequenced thus far, such as the model strains 536{Brzuszkiewicz:2006cu}, CFT0073{Welch:2002bj}, and UTI89{Chen:2006wz}, 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 UPEC{Rasko:2008bx, Touchon:2009kw, Dobrindt:2010fe}. Despite the number of unique genes, members of the UPEC group have a greater genomic similarities and are more genetically distinct, as a group, than other pathovars{Rasko:2008bx}, which indicates that 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 factors{Picard:1999uk, Norinder:2012fq, Touchon:2009kw}. 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 known virulence factors and clade membership of UPEC strains {Poey:2012be}. 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 genes {Johnson:2001cl}. This pattern mirrors the homologous recombination in core genes, which has been shown to be high in *E. coli* {Touchon:2009kw}, 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 pathogenesis{Picard:1999uk}, 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 pathogenicity{Picard:1999uk}.

*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 themselves{Picard:1999uk, Marrs:2005ty}, 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 gene {EscobarParamo:2004up, Nowrouzian:2005uu}. In many cases, non-pathogenic and pathogenic bacteria contain similar sets of virulence factors{Kohler:2011cn}, as can be seen in commensal probiotic strain *E. coli* Nissle 1917 and the uropathogenic bacterium *E. coli* CFT073 {Grozdanov:2004bd}, 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 D{Boyd:1998ub}. Furthermore, genomic hybridization shows a correlation between the presence and absence of specific genr content and the phylogenetic history of the core-genome of B2 isolates, indicating the co-evolution of the accessory and core genomes{LeGall:2007bq} possibly through a process of “fine-tuning” {EscobarParamo:2004to}. 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 phenotype{LeGall:2007bq}

*Urovirulence factors: swords or plowshares?*

A number of models have been proposed to explain the evolution of virulence and population structure of UPEC, including the “source-sink” model{Sokurenko:1998ww, Sokurenko:2006vm}, 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” model{LeGall:2007bq}, 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* gene{Sokurenko:1994to, Sokurenko:1995ul, Sokurenko:1999wg, Weissman:2007bb}, 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 environment{LeGall:2007bq}. 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 women{Nowrouzian:2003bs, Schlager:2003hh}. Additionally, in adult healthy women, dominant E. coli clones had higher urovirulence scores than non-dominant clones, indicating that urovirulence factors helped mediate gut fitness{Moreno:2009jc}. This pattern is mirrored in persistent and transient strains of *E. coli* in the gut, as persistent strains were more likely than transient strains to present uropathogenic phenotypes{Wold:1992tg}. 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 CFT0073 reduces rate of intestinal colonization{Diard:2010fr}. 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 membership{Nowrouzian:2006bu, LeGall:2007bq}. 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 habitat{Moreno:2008eg, Moreno:2009jc}. 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 CFT0073 is linked to reduced growth rate in urine{Diard:2010fr}, 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 urovirulence 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?
* Are there genomic differences between UPEC strains that persist between recurrent episodes and those that are replaced via reinfection? Are persistent or replaced strains more likely to be found in the gut habitat prior to infection?

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.