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 and incidence{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}. 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}. Asymptomatic bactiuria, in which bacteria exist in the bladder without causing clinical manifestations, can be caused by UPEC{Salvador:2012bc}. 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 exist in a host without causing disease{Denamur:2010ug}.

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 in 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) capable of causing urinary tract infections (UTIs) {Russo:2000vr}. In addition to phenotype, *E. coli* can also be categorized according to their phylogenetic history. Interestingly, pathogenic potential, genomic content, and phylogenetic history are not always 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}.

Currently, the most widely used method of grouping strains into a phylogenetic structure relies on Multi-Locus Sequence Typing (MLST).

*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 chemotherapy. Before the rise of antibiotic use, bacteria were known to reside in the bladder for long periods of time, despite palliative care to remove symptoms{Nickel:2005ke}, which indicates that a large portion of people are unable to clear the bacterial infection by themselves. 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. Because of this widespread antibiotic therapy, such as treatment 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 treatment 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} which result in a drift to clonality in UPEC in the bladder{Schwartz:2011cy, Walters:2012hq}. These bottlenecks may 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{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 -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 UPEC. Recurrent UTI strains are not always caused by the same strains responsible for the initial UTI, indicating that reinvasion by unique strains may occur. Estimates of the percentage of recurrent UTIs caused by this reinvasion by unique strains vary greatly between studies{Russo:1995ts, Ikaheimo:1996ue, Czaja:2009jx, Jantunen:2001td, Luo:2012bm, Ejrnaes:2006vr}. 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. 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}. This is more evidence supporting a link between urovirulent gene content in UPEC and population dynamics in the bladder

*Escherichia coli population structure in the gut and perineum*

The gut populations of *E. coli* are surprisingly simple. The number of *E. coli* strains in the gut of women experiencing a UTI (~3) does not differ significantly from the number of *E. coli* strains in the guts of healthy women (2.5) as determined by PCR typing{Moreno:2008eg, Moreno:2009jc}. 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 are able to outcompete less urovirulent strains in the gut habitat, which may result in local extinction of those less virulent strains. This pattern mirrors the population dynamics that occur in the bladder during UTI, indicating a shared functional effect on population structure by virulence factors in both habitats.

*Transmission of UPEC between the gut and bladder*

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*{Gruneberg:1969wo, Yamamoto:1997wk, Moreno:2006ji, Moreno:2008eg}.

**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 (reviewed in Dobrindt *et al.* 2010). 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 and less intra-group diversity. 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 processes of 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, genomic analysis of virulence factors presence and its correlation to fitness has shown that the genetic context of the virulence factors is important in determining its functional effect within UPEC strains{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 genomic architecture, 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 genomic 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 “coincidental pathogenesis” model{LeGall:2007bq}, which stipulates that extra-intestinal pathogenicity is a by-product of adaptation to the gut environment. 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{Nowrouzian:2003bs}. 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}. Additionally, in patients with UTI, strains with greater numbers of urovirulence factors were associated with reduced species richness in the gut habitat and were more likely to belong to clade B2 {Moreno:2008eg}. This suggests that urovirulence factors may facilitate the clonal expansion of UPEC strains in the gut at the cost of competing *E. coli* strains, which results in a reduction of species diversity, and that this propensity for urovirulence may be mediated, in part, by the phylogenetic history of a strain.

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**

The population dynamics of