**Uropathogenic *Escherichia coli*: Population Structureand the Evolution of Virulence**

Approximately 50% of women will suffer a urinary tract infection (UTI) at some point in their life, and 20-30% of these women will suffer a recurrent UTI within three to four months (Hooton papers and Foxman 2002). The rates of UTI increase in the immunosuppressed, including the elderly and children, and may result in significant complications, including renal scarring, septicaemia, and pyelonephritis (Foxman 2002). Approximately 80% of community acquired UTIs are caused by uropathogenic *E. coli* (UPEC), a Gram-negative gammaproteobacteria (Foxman 2002, Foxman 2010). UPEC has been linked to several outbreaks in recent years, and, of particular concern, antibiotic resistance within UPEC has begun to spread intercontinentally, resulting in increased morbidity and mortality (Nicolas-Chanoine, Rogers 2010). This spread of multi-drug resistant, highly virulent strains of UPEC, as well as the rate of infection in the U.S., necessitate an understanding of the underlying population dynamics of this pathogen.

While UPEC is clinically important, it is also an excellent model system to study the evolution of virulence in opportunistic pathogens (Dobrindt 2005 Dobrindt 2010, Tenaillon 2010, Hannan 2012). 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 2012). 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 (Welch 2002, Chen 2006, Luo 2009, Avasthi 2011). Despite these benefits, and in contrast to other forms of pathogenic *E. coli*, the genome dynamics and population structure of UPEC remain largely unexplored. Additionally, 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 and bioinformatic tools now enable high-resolution descriptions of bacterial population structures using genomic analyses, enabling research into these unexplored areas.

**Pathogenomics can be used to Study UPEC Adaptation and Population Structure**

Researchers in the field of pathogen genomics (pathogenomics) aim to understand how changes in the genomic architecture of pathogens result in changes in virulence to host organisms. Successful research into bacterial virulence has been accomplished through integration of tools of microbial research with the insights available from sequence information. Through this combination, researchers have been able to build a model framework around the differential phenotypes that exist between strains of pathogens. From this foundation, researchers are able to describe how a particular collection of genes within a particular organism can cause disease, determine pathogen fitness and niche specialization, and modulate the host-pathogen interface – all important considerations in understanding the evolution of opportunistic pathogens like UPEC. Much of the recent pathogenomics research has fallen into two main categories: (i) the identification of genes important for virulence, drug-resistance, or immunogenicity and (ii) the development and use of typing tools to describe population dynamics (Pallen and Wren). In UPEC, pathogenomic analyses have focused extensively on the first category, while the second has been investigated less thoroughly.

*Pathogenomic Analysis of Pathoadaptation*

Horizontal gene transfer (HGT) is a major mechanism of innovation in bacterial evolution (Innovation Paper) and has been shown to affect the evolution of virulence in opportunistic pathogens (Canchaya, etc); however, HGT is not the only mechanism by which bacteria increase their fitness and virulence. Other mechanisms, such as pathoadaptation, have been shown to have effects on bacterial pathogenicity (Chen 2009, Sokurenko 2009, Sokurenko 1998, Iebba 2012). Pathoadaptation can be defined as alterations in existing genes that occur as a result of specialization to new host-associated niches (Sokurenko 1999, Pallen Review, Maurelli Review, Rakin Review). These pathoadaptive changes affect the population structure of pathogens, such as UPEC, by increasing the abundance of bacteria in the novel niche (Sokurenko 1999, 2009). The functional effects of pathoadaption are numerous and may include increased host-defense evasion, greater invasion or transmission potential, or a novel ability to modify the host-environment among many others and often results in increased fitness at the expense of the host. Pathoadaptive changes occur in already existing genes through the processes of gene loss, gene mutation, or changes in gene expression and are common in bacteria that shift into a novel niche (Sokurenko 1998, Wren Review). These pathoadaptive changes have been identified in UPEC, and often result from modifications to a repertoire of virulence genes through processes of gene-loss (Salvadore 2012), refinement of existing pathogenicity factors (Chen 2009, Iebba 2012), or changes in expression of virulence genes through mechanisms such as phase-variation (Pallen and Wren Review, Maurelli Review, Rakin Review, Groznadov 2012).

A number of UPEC strains have been sequenced, including the standard model strains CFT073 (Welch 2002) and UTI89 (Chen 2006), in addition to a very wide array of sequence data available regarding the prevalence of different virulence factors involved in uropathogenicity. The majority of differences between strains of *E. coli* are found in the pan-genome, which is comprised of all genes that exist in the genomes of all strains of that species (Medini paper, review, Welch 2002, Chen 2006, Dobrindt 2005). The pan-genome of *E. coli*, which is still growing with each new genome sequenced,consists of over 10,000 genes and is nearly five times larger than the core genome shared between all strains, indicating a highly reticulated population structure (Rasko, Touchon Chaudhuri 2012). In the last decade a number of comparative genomics investigations identified a suite of virulence factors and molecular mechanisms involved in uropathogenicity, pathoadaptation, and niche adaptation in UPEC strains (Mysorekar, Reigstaad, Lloyd 2009, Dobrindt 2010, Nielbuwicz, Norinder, Bauchart, Schneider, Salvador, Grobdanov, Johnson).

While research into pathoadaptation and virulence factors is ubiquitous, specific definitions of these concepts are difficult to articulate (Pallen Review). The current definition of pathoadptation can be succinctly described as genetic or genomic alterations in a pathogen that result from specialization to a new niche through the modification of existing genes; however, whether pathoadaptation necessitates an increase in virulence is unclear (Sokurenko 1999, Pallen Review, Rakin review). The definition of a virulence factor requires more finesse and is generally more dynamic depending on the context of host organism, non-pathogenic habitats, and the pathogen of study; however, for the purposes of this paper, a virulence factor is described as a gene or gene product that increases fitness of a host-associated microbe at the host's expense (pallen review). The complexities in the definitions are excellent analogies for the complexities that exist in the field of research into UPEC pathogenesis and population structure. As technological capabilities have increased, anthropocentric, single-gene research has been replaced by genome-level analysis that incorporates the ecological context of UPEC into the interpretation of data. Given the dynamic and multifarious nature of pathoadaptation, genome-level analyses, rather than a focus on individual genetic components of a virulence factors, are necessary to completely understand pathogen evolution. These pathogenomic analyses subsume genetic changes in single virulence genes into a model of genome-wide genetic alterations and chromosomal dynamics and provides additional context necessary to fully articulate the complex relationships that exist between hosts and pathogen populations.

**Population Dynamics of UPEC during UTI: Questions Remain**

Currently, pathogenic *E. coli* have been categorized according to their pathology and genomic content into a number of different clades (Nataro 1998, Boudeau 1999, Darfeuille-Michaud 2002, Kaper 2004, Russo and Johnson 2000 and Chaudhuri 2012). These categories of *E. coli* are delimited by differences in the genomic content of the strains, which differ markedly between clades (Rasko 2008, Touchon 2009, Chaudhuri 2012). Within the UPEC clade, differences in genomic content are also found within the pan-genome, although allelic variation within core genes is also present (Welch, Chen 2006, Dobrindt 2005). These genomic differences have been used to identify the population and phylogenetic structure of UPEC between patients (Rijavec 2006, Poey 2012, Gilbreel 2012, Salvador 2012), however, these analyses have relied on multilocus-sequence typing, which has been shown to lack resolution in species with large pan-genomes, such as UPEC (Medini 2006, Achtman 2008). While serotyping and multi-locus sequence typing (MLST) have long been used to classify and subcategorize pathogens, these technologies have been found to have limitations in accurately describing the *E. coli* populations (Figure 1)(Medini 2006, Achtman 2008). Both 16S ribotyping and MLST offer greater resolution between bacteria than serotyping, their resolution is still limited when compared to the resolution offered by analysis of single-nucleotide polymorphisms (SNPs). These limitations result, in part, to the high levels of recombination found in many bacterial pathogens, which obscures their phylogenetic history and population dynamics (Gogarten 2002, Koonin 2003, Doolittle 2006). 16S ribotyping and MLST tools have offered significant insight into the global population structure of UPEC, however, despite these many accomplishments, a number of important areas of research into the population structure of UPEC remain unexplored, most notably pathogenomic analyses of population structure of UPEC before, during, and after UTIs.

*The Population Structure of UPEC is Affected by the Infection Cycle*

As to be expected, the dynamics of UPEC population structure during UTI is directly affected by the progression of the infection. The current model of UTI progression is complicated and consists of a number of invasion events that restrict the population diversity, thus affecting the underlying population structure (Figure 2) (Schwartz 2011). UPEC that invade the bladder are thought to originate in the gastrointestinal tract (Moreno 2006 and 2008, Yamamoto 1997), although direct evidence for this phenomenon has not been provided (hannan 2012). Once UPEC are in the lumen of the bladder, Type 1 pili tipped with a FimH adhesin bind to mono-mannosylated ligands present on the bladder epithelium known as uroplakins (Zhou 2001). Following adherence, UPEC subsequently invade the epithelial cell and establishes a clonal community called an intracellular bacterial community (IBC) in a *fimH* dependent manner (Mulvey 1998, Anderson 2003). After maturation of the IBC, the clonal UPEC bacteria flux out of the urothelium, killing the host cell and invade new epithelial cells. Continuation of this cycle results in chronic cystitis and occurs if bacterial titers are high enough in the initial acute phase of the UTI (Schwartz 2011). Alternatively, quiescent intracellular reservoirs (QIRs) may develop if the UPEC gain entry into the underlying epithelium below the superficial facet cells lining the bladder (Mysorekar 2006). In such cases, UPEC may exist in a dormant state and emerge at a later time to cause a recurrent UTI (Schilling 2001, Mysorekar). During infection progression, a combination of population bottlenecks during invasion and IBC formation, founder effects during recurrent UTIs, and migration patterns between the gut and bladder habitats have significant effects on the population structure of UPEC (Schwartz 2011, Walters 2012). The complex nature of the UPEC infection cycle has resisted description, and, currently, there exists only one dominant model describing the overall effects of UTI on the population structure of UPEC - the source-sink model.

*The Source-Sink Model of UPEC UTI*

The source-sink model of population dynamics is an explanation of observed population dynamics and migration patterns between niches in which certain "source" habitats support the population of connected "sink" habitats (Pulliam 1988). In this model, populations that exist in source habitats experience growth due to increased birth rates relative to death rates. This population growth results in an increased level of emigration to other habitats as compared to the habitat's rate of immigration. Sink habitats, on the other hand, experience greater death rates than birth rates, and must have their populations supplemented by rates of immigration that are higher than their rates of emigration. Although this model was originally applied to macroscopic ecology, the model was adapted to explain the population dynamics of several pathogens, including UPEC (Sokurenko 1998, 2004, 2006, 2009). In this model of bacterial pathogenicity, virulence factors that are adapted to increase fitness in one environment cannot be optimally adapted for a different environment (Sokurenko 1998, 2006). When bacteria invade a new niche, the population that persists develops pathoadaptive mutations that increase fitness in the novel environment, which results in a concomitant loss of fitness in the old niche. However, if the novel niche is particularly stringent, short-lived, or if the invasive population is small, then pathoadaption after invasion is unlikely, and the population soon goes extinct (Figure 3). As such, in these cases, pathoadaptation is expected to occur before invasion of a novel niche, the sink habitat, and expected to exist at low frequency in the old niche, the source habitat, as a result of the loss of fitness that occurs due to pathoadaptation to the novel sink habitat. If transmission between niches is possible, then a population pathoadapted to the sink habitat may migrate back to the source niche, however, these events are predicted to be rare (Sokurenko 2006).

The source-sink model has been proposed as an explanation of the population dynamics of UPEC UTI in humans. In this model, the gut habitat is considered the source habitat while the bladder has been determined to be a sink habitat, most likely due to the differences between nutrient availability, presence of host defenses, and competition for niche space (Alteri 2012). Pathoadaptation is likely to occur before *E. coli* strains invade the bladder, given the small numbers of invading bacteria and inhospitable nature of the bladder habitat (Sokurenko 2006). After migration to the bladder, UPEC persistence is short-lived due to natural clearance of the bacteria by the innate and adaptive immune system or therapeutic intervention (Foxman 2010), resulting in a much higher death rate than growth rate, thus satisfying the definition of the source-sink model. In this model, recurrent UTIs are caused by recurrent colonization of the bladder with different strains of UPEC that have resulted from separate *de novo* pathoadaptation processes (Sokurenko 2007).

Support for the source-sink model of UPEC UTI relies mainly on evidence of pathoadaptation in the *fimH* gene and its role in niche differentiation and extra-intestinal colonization. The *fimH* gene has been shown to be critical for tropism to the bladder (Hung 2002) and invasion of uroepithelial cells by binding to mono-mannosylated uroplakin receptor UPA1A (Zhou 2001), but the *fimH* adhesin has also been suggested as an important factor in colonization of the gastrointestinal tract by binding to D-mannose moieties on mucosal glycoproteins found on many types of cells (Sokurenko 1994). Subsequent *in vitro* functional investigations found that the polymorphisms in *fimH* resulted in altered binding affinities for different ligands expressed in differential tissues (Sokurenko paper 1995 and 1998). These changed binding affinities come at a cost, however. Although *E. coli* carrying these mutations in *fimH* have increased binding affinities to ligands expressed in one location (either the GI tract or the bladder), the mutations have been found to decrease the binding affinity for ligands expressed in the other body habitat (Sokurenko paper 1998). For example, several point mutations in the *fimH* gene increase the binding affinity of type I pili to the mono-mannosylated uroplakin proteins in the bladder, however, these mutations also increase the susceptibility of type I pili to inhibition by compounds found in the saliva of mammalian hosts, which may reduce the bacterium's capability to colonize the GI tract (Sokurenko 1998). As a result, these pathoadaptive mutations in UPEC exist at a low frequency in the gut populations of UPEC as a result of negative selection against the reduced fitness of the mutation, as predicted by the souce-sink model (Sokurenko 2006, 2007, 2009). This model also predicts that, for these pathoadapative mutations, the bladder environment functions as an evolutionary dead-end as a result of the increased instability of the mutation in the gut and the reduced potential for fecal-oral transmission, thus further reducing the likelihood of persistence of the pathoadaptive mutation in the global population of *E. coli* (Sokurenko 1998, 1999, 2006, 2007). Additionally, the presence of footprints of positive-selection in *fimH* have been suggested as evidence for the role of *fimH* in niche differentiation, as these types of point mutations are associated with entry into a novel niche (Sokurenko 2004). Additional support for this claim of niche adaptation comes from the greater haplotype diversity found in *fimH* genes from *E. coli* strains isolated from urinary tract samples than *E. coli* strains isolated from fecal samples, which may indicate repeated adaptation to the bladder following many *de novo* mutations (Sokurenko 2006). These data show that polymorphisms in the *fimH* gene are associated with functional differences in different body habitats, resulting in altered fitness and population persistence. These data are a clear indication that evolution of virulence and population dynamics are intrinsically linked during UPEC UTI and have been used as support for the source-sink model.

*Flaws in the Souce-Sink Model of UTI*

At its core, the source-sink model of UPEC UTI relies on the assumption that pathoadaptation to one environment necessitates a reduction in the fitness in other, dissimilar environments. While it is true that optimization of a continuous character trait (such as the dimensions of the beaks on Darwin's finches) responsible for multiple tasks (such as crushing insects, capturing nectar, or cracking seeds) requires trade-offs between optimal design between those tasks (Shoval 2012), on a molecular level, optimal fitness in multiple tasks (such as binding to two different mucosal surfaces) could be achieved through a process of compensatory mutations. The current model of source-sink population dynamics neglects the role of compensatory mutations, which may lessen the fitness cost of pathoadaptation in microbes that inhabit multiple environments. Compensatory mutations are often found in bacteria that have developed antibiotic resistance, as initial antibiotic resistance mutations may have a steep fitness cost (Levin 2000). These compensatory mutations are thought to occur very quickly (Sousa 2012) and can occur even in the absence of antibiotics (Levin 2000), indicating the compensatory mutation may exist even in the absence of a strong selection pressure for compensation. While mention of compensatory mutations is made briefly in literature detailing the source-sink model of UPEC UTI (Sokurenko 2000, Weissman 2006), the role of compensatory mutations in the abrogation of fitness costs of pathoadaptation has yet to be fully explored. Because compensatory mutations may occur in a number of genes (Soursa 2012), a genomics approach is best suited to identify recurring compensatory mutations. If pathoadaptation to the bladder does come at a fitness cost to bacterial capability to colonize the gut, then an abundance of compensatory mutations restoring fitness in the gut may explain the ability for clonal populations of UPEC to dominate in both the gut and bladder habitats.

Additionally, the stringency of selection for tasks in multiple environments is highly context dependent. For example, while pathoadaptation of theFimH adhesin to bind ligands in the bladder has been found to reduce the ability of FimH to perform its other task of binding ligands present in the gastrointestinal tract *in vitro* (Sokurenko 1995, Schembri 2000), knockout of the entire Type 1 pili apparatus, of which *fimH* is an integral part (Hultgren 1985), does not affect gut colonization by *E. coli* *in vivo* (Bloch 1992, McCormick Paper). This indicates that, even if pathoadaptation of *E. coli fimH* to the task of binding to the bladder ligands was made optimal and completely eliminated ability to perform other tasks, there are still ecological contexts in which the fitness cost expected to occur through the process of pathoadaptation would be removed. As a result, while the source-sink model of UPEC UTI may be theoretically sound, the biological relevancy of the model deserves further research.

According to the source-sink model of UTI by UPEC, the bladder will be invaded and colonized by separate strains of *E. coli* that have developed separate, though possibly recurrent, *de novo* mutations that increase fitness in the bladder environment; however, these invasive, pathoadapted *E. coli* clones eventually go extinct in the bladder due to the transient nature of UTIs and the reduced fitness of the pathoadaptive strains in other habitats (Sokurenko 2004 paper, 2006 review, 2006 paper, 2009 paper). As a result, virulence factors mediating uropathogenicity do not persist long enough to develop non-synonymous mutations and are expected to exist at low frequencies in the source populations of *E. coli* residing in the gut (Sokurenko 2009). However, recently obtained evidence from the Hultgren lab indicates that at least two predictions postulated by the source-sink model of UPEC UTI may not be biologically accurate (paper in review). These predictions are that: (i) pathoadaptive mutations increasing fitness in the bladder exist at low frequency in the gut and (ii) pathoadaptive mutations increasing fitness in the bladder environment result concomitant decrease fitness in the gut. In this investigation, *E. coli* strains were isolated from urine and rectal swab samples from four patients across multiple time-points during recurrent UTIs and subjected to multi-locus sequence typing and whole-genome sequencing in order to identify the clonality of the strains. This research shows that, during recurrent UTIs, the dominant strain in populations of *E. coli* in the distal colon are clonal matches to UPEC found in the bladder. These data are evidence that virulence factors can exist in high frequency in the gut, in contradiction to the expected allele frequency predicted by the current source-sink model. The strains also remained constant across time-points, except in one patient. In this particular patient, the previously dominant strain was supplanted by a new strain containing different SNPs and gene content in both the gut and bladder habitats, indicating that a new strain was more fit in both environments, in direct contradiction to the fitness trade-off that is predicted to occur during pathoadaptation to the bladder environment. These data were confirmed using MLST analysis and the fitness effects were further elucidated in competition experiment in animal models of both gut and bladder persistence. Taken together, these data indicate that the source-sink model may not apply to all, or even most, of human UTIs caused by UPEC, and, thus, deserves further investigation in order to resolve the discrepancy between the theoretical model and empirical evidence.

**Concluding Remarks**

The dynamics of pathogen population structure have been found to influence the spread of drug-resistance as well as the evolution of virulence in many important infectious diseases (Courcher, Harris, Holt, Leiberman) and is likely a factor in the UPEC virulence as well. Currently, research into the population structure of UPEC has focused mainly on investigations into the source-sink model using molecular techniques such as MLST or single-locus analysis of the virulence factors, which have been shown to lack resolution in many cases (Achtman 2012). Additionally, although the process of recurrent UTIs offers a unique opportunity to test the biological relevancy of the source-sink model, research has focused extensively on acute infections. As yet, the diversity and population structure of UPEC in the bladder and the gut have yet to be thoroughly explored using next generation sequencing technologies. These technologies offer increased resolution between strains and greater robustness to confounding effects of recombination (Next-gen sequencing technology paper), and can be used to study this model system of an opportunistic pathogen.

Knowing general mechanisms of evolution and population dynamics allows prediction of the evolutionary trajectory of pathogens through morphospace, which is defined as the collection of all possible phenotypes for a particular organism (Shoval 2012). Knowledge of this trajectory ultimately enables researchers to understand, and correct, human activities that increase the virulence, transmission, and drug-resistance of clinically important pathogens.