**Studying Pathogens at the Genomic Level**

*Pathoadaption*

A central, and contentious, question in the investigation of microbial pathogens can be simply stated: What makes a commensal microbe different from a pathogen? Chief among the mechanisms suggested to mediate these differences is the process of pathogenicity adaptive (pathoadaptive) mutations (sokurenko 1999). 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). Pathoadaptive changes often result in increased fitness of the pathogen in their new habitat at the expense of the host (Pallen and Wren review). 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. 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 often result from modifications to a repertoire of virulence genes in the microbe through the gene-loss, refinement of existing pathogenicity factors, or changes in expression of virulence genes through mechanisms such as phase-variation (Pallen and Wren Review, Maurelli Review, Rakin Review).

Pathoadaptation through regulated gene-loss has been identified in a number of bacterial pathogens, most notably in *Shigella* species (Blackholes paper 2007). In *Shigella* species, loss of the anti-virulence gene *cadA*, encoding a lysine decarboxylase has been found to dramatically increase the enterotoxigenic effects of the pathogen by removing a direct inhibitor of plasmid-encoded enterotoxins (Blackholes paper 1998). Interestingly, pathoadaptation through gene-loss does not always result in increased virulence or damage to the host. In asymptomatic bactiuria (ABU), *E. coli* stably colonize the bladder and persist at high titers without causing inflammation of the bladder, pyelonephritis, or tissue damage. Analysis of known uropathogenicity genes in strains isolated from patients with ABU shows that strains causing ABU have a number of nonsense or nullomorphic mutations that reduced the overall pathogenic capability of the strains (Salvador 2012). Additionally, these ABU strains harbored smaller genomes containing multiple deletions of uropathogenic genes. These losses of pathogenic potential resulted in a much greater carriage of these ABU strains in patients through time and are evidence of pathoadaptive changes to the *E. coli* strains to the bladder habitat.

In many cases, non-pathogenic and pathogenic bacteria contain similar sets of virulence factors, as can be seen in commensal probiotic strain *E. coli* Nissle 1917 and the uropathogenic bacterium *E. coli* CFT073 (Grozdanov 2012), however, despite their similar genomic architecture, these strains have widely different pathogenic potential. Often, these different phenotypes may occur through pathoadapative changes in the regulation of virulence gene expression. Regulation of virulence factors acquired through horizontal gene transfer (HGT) is intricate and dependent on interactions between the newly acquired and existing genetic structure of the pathogen (Escobar 2004); however, these interactions provide ample resources for niche adaptation and pathoadaptation.

A final mechanism of pathoadaptive changes occurs through the non-synonymous mutations in the coding regions of existing genes. Through this process, non-synonymous changes in existing genes results in increased fitness in the novel niche by altering or refining the function of the affected gene. The initial, and perhaps most persuasive, evidence for this pathoadaptive model of pathogen evolution comes from studies of the *fimH* gene in uropathogenic *E. coli* (UPEC). The *fimH* gene encodes an adhesin at the tip of type 1 pili that is involved in adherence and invasion of uroepithelial cells lining the bladder (Hultgren 1985 and Mulvey 1998). Analysis of the *fimH* gene revealed a number of non-synonymous mutation hotspot that were found to be functionally important (Snyder 2006, Weissman 2006, Chen 2009). The distribution of evidence for positive selection between isolates collected from fecal and urine samples, indicating a putative role in niche differentiation between the gut and bladder environments (Sokurenko 2004 Chen 2009). This hypothesis is further supported by functional assays that show inversely proportional differences in binding affinity between ligands found in the gastrointestinal tract and those of the bladder environment (Sokurenko 1995, Schembri 2000). These studies, among many others, indicate the significance of non-synonymous mutations in the adaptation of pathogenic bacteria to their environments.

While the concepts of pathoadaptation and virulence factors specific definition of these concepts are difficult to articulate (Pallen Review). The current definition of pathoadptation can be succinctly described as genetic or genomic alterations that result from specialization to a new niche through the modification of existing genes (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 microbial pathogenesis. As technological capabilities have increased, anthropocentric, single-gene research has been replaced by genome-level analysis that incorporates the ecological context of pathogens 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 pathogens’ virulence factors, are necessary to completely understand pathogen evolution. This genome-wide view of pathogenicity subsumes 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, pathogens, and the ecosystem at large. This new avenue of research has been labeled pathogenomics.

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

Researchers in this new field of pathogenomics aim to understand how changes in the genomic architecture of pathogens result in changes in the virulence to host organisms. This is accomplished by the integration of tools of microbial research with the insights available from sequence information. Through this combination, researchers are able to build a model framework around the differential phenotypes that result from genomic changes in vitally important microbial pathogens. From this foundation, researchers are able to describe, in fascinating detail, 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 - the essence of pathogenomics. Much of the recent pathogenomics research has fallen into two main categories: (i) the development and use of typing tools to describe current and historical population dynamics and (ii) the identification of genes important for virulence, drug-resistance, or immunogenicity. These two areas of research are fundamentally necessary to understanding how pathogens cause disease in human hosts, and, perhaps more importantly, aid our understanding of how human hosts affect the evolution of pathogens and their virulence mechanisms.

While serotypes, serogroups, and multi-locus sequence types have long been used to classify and subcategorize pathogens, these technologies have been found to have limitations in accurately describing the pathogen populations (Chadhuri 2012). 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 2003, Koonin 2003) and the discordance found in gene content between sister strains of the same species, which blurs the distinction between pathogenic and commensal (Tettilini 2005, Medini 2006). The advent of next generation sequencing has sparked a surge of new insights into the ways that pathogens interact with their host environment and each other, leading to a deeper understanding of their intricate population structures. Second and third generation sequencing technologies now offer an array of tools to sequence, compare, and analyze large numbers of bacterial genomes on an incredible scale (Loman 2012). Within the last year, these technologies have been used in pathogenomic studies to describe the phylogeny and dissemination of the pathogens *Salmonella Typhimurium* (Okoro et al. 2012) and *Shigella sonnei* (Holt et Al 2012). These "historical" phylogenies are accurate calculations of the diversity and divergence within these two pathogens, and have resulted in new knowledge regarding the origin, and potential distribution, of both highly virulent and mutli-drug resistant strains. In addition to their high-throughput data generation capabilities, new sequencing technologies also offer increased levels of resolution that can be used to identify important features of bacterial population dynamics. Recently, next generation sequencing and subsequent analyses of *Chlamydia trachomatis* were used to show that previous phylogenies based on single genes were inaccurate (Harris 2012). These analyses also showed that a fundamental assumption regarding the genomic dynamics of *C. tracomatis*, namely an absence of recombination, were flawed and not biologically accurate. These recent examples show how effective current sequencing analyses can be used to accurately, and with great detail, describe the dynamics of population genetics in clinically-relevant pathogens.

In addition to elucidating population dynamics of pathogenic bacteria, high-throughput sequencing has been used to describe the effects of within-host evolution on the genes responsible for bacterial virulence and persistence, which are both important considerations in understanding the effects of pathoadaptation. Recently, 112 genomes of the prominent bacterial pathogen *Burkholderia dolosa* were analyzed for adaptive mutations that occurred as the bacterium persisted in chronically infected patients (Lieberman 2012). This information was used to identify genes under the effects of convergent evolution, a strong indication of their importance in pathogenicity. Pathogenomics has also been used to identify the adaptive mutations of pathogenic bacteria that have been exposed to therapeutic interventions. In a study of *Streptococcus pneumoniae,* Croucher *et al.* elucidated the contribution of horizontal and vertical adaptive mutations to bacterial diversification and pathogenicity following the deployment of a vaccine (Croucherr 2012). These studies show how evolution of virulence can be tracked using pathogenomics and used to identify regions important to the adaptation of pathogens after changes in their population structure.

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. Understanding the effects of microbial ecology and population structure on the evolution of the host-pathogen interface is critical in determining effective therapeutics, as can be seen in recent research investigating the spread of drug-resistance in pathogenic *Plasmodium* species responsible for malaria (Harrington, Bell 2012). Additionally, novel therapeutics can be developed with the assistance of pathogenomic analysis through the process of reverse vaccinology, which has already yielded significant advancements in the prevention of infectious disease (Medini 2006, Pallen review). Although pathogenomics has been used to investigate the breadth of pathogenic species from viruses to bacteria to eukaryotes, the bacterium *Escherichia coli* offers unique advantages as a model organism for pathogenomic analysis to further understand the basic biological principles that govern the processes of pathoadaptation and the evolution of virulence.

*Pathogenomic analysis of* Escherichia coli

*Escherichia coli* are associated with a number clinical conditions, each caused by a separate clade of *E. coli* harboring different repertoires of gene sets and virulence factors, and, as such, have been categorized according to their pathology and genomic content. *E. coli* that cause disease in the gastrointestinal tract are grouped together and include enteropathogenic *E. coli* (EPEC), capable of causing adherent and effacing lesions in the gastrointestinal tract, enterohaemorrhagic *E. coli* (EHEC) that elaborate Shiga toxin and cause haemolytic uremic syndrome, Enteroaggregative *E. coli* (EAEC) and diffuse-adherent *E. col*i (DAEC) which differ in the organization during attachment to the epithelium, enteroinvasive *E. coli* (EIEC) which are capable of invading the gut epithelium, and adherent invasive *E. coli* (AIEC) which are associated with Crohn's Disease (Nataro 1998, Darfeuille-Michaud 2002, Kaper 2004, Chaudhuri 2012). A separate group consists of extra-intestinal pathogenic *E. coli* (ExPEC) and includes strains capable of causing neonatal meningitis (NMEC) and uropathogenic *E. coli* (UPEC) capable of causing urinary tract infections (UTIs) (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). 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). 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 (Rasko, Touchon Chaudhuri 2012). Additionally, virulence genes, often associated with mobile factors that make up the pan-genome (Medini 2006), are present in *E. coli* genomes in a clade specific manner (Dobrindt 2005, others). These factors, the size and plasticity of the pan-genome, the range of pathotypes, and the phylogenomic structure, combine to make *E. coli* a unique model system to interrogate the development of virulence in a genomic context. In particular, UPEC offers a unique opportunity to study the genomic differences that separate opportunistic pathogens from commensal strains (Dobrindt 2005).

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. In the last decade a number of comparative genomics investigations identified a suite of virulence factors and molecular mechanisms involved in uropathogenicity and niche adaptation (Mysorekar, Reigstaad, Lloyd 2009, Dobrindt 2010, Bauchart, Schneider, Johnson). These studies have laid a strong foundation for pathogenomic interrogations into the development of pathogenicity, however, despite these many significant accomplishments, a number of important areas of research remain unexplored, most notably pathogenomic analyses of population structure of UPEC before, during, and after UTIs. As was shown earlier, pathogenomics can be used to identify important virulence factors and processes of evolution, and can also be used to elucidate the population dynamics of pathogenic organisms. This is critically important in pathogens that have complex infection cycles, such as UPEC.

**Population Dynamics of UPEC of UTI: Waiting to Be Explored**

*The Population Structure of UPEC*

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 1) (Schwartz 2011). UPEC that invade the bladder are thought to originate in the gastrointestinal tract, 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 invading 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). 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 (Figure 2). 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 occure before *E. coli* strains invade the bladder, given the small numbers of invading bacteria and inhospitable nature of the bladder habitat (Sokurenko 2006). Pathoadaptation in a population of UPEC residing in the source habitat of the gut become capable of persisting in the unstable sink habitat of the bladder, through modification of a variety of existing virulence factors. After migration to the bladder, UPEC persistence is short-lived due to natural clearance of the bacteria 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. Support for the source-sink model of UTI has come mainly from analysis of the frequencies and functional effects of virulence gene alleles (Sokurenko 2006 review).

Despite the large array of virulence factors that have been shown to affect urovirulence, support for the source-sink model of UPEC UTI relies mainly on 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 GI tract by binding to D-mannose moieties on mucosal glycoproteins found on many types of cells (Sokurenko 1994). Subsequent 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 (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.

*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 phenotype (such as the dimensions of the beaks of 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). While mention of compensatory mutations is mentioned 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 (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* after streptomycin treatment (McCormick Paper). This indicates that, even if pathoadaptation of *E. coli fimH* to the task of binding to the bladder ligands was reached 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 nullified. 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 three predictions postulated by the source-sink model of UPEC UTI may not be biologically accurate (paper in submission). These predictions are that: (i) pathoadaptive mutations increasing fitness in the bladder exist at low frequency in the gut, (ii) pathoadaptive mutations increasing fitness in the bladder environment concomitant decrease fitness in the gut, and (iii) recurrent colonization of the gut occurs through invasion by new clones of UPEC with separate pathoadaptations. In this research, *E. coli* strains were isolated from urine and rectal swab samples from four patients across multiple time-points related to 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. Additionally, the strains also remained constant across time-points, except in one patient. In this particular patient, a serotype switch occurred 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. Finally, in the remaining patients, the clonal nature of each instance recurrent UTIs indicates that *de novo* 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**

During their lifetimes, approximately 80% of women will suffer from a UTI, and between 20% and 30% of these women will experience a recurrent UTI shortly thereafter (Foxman 2010). 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, 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. 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 for the establishment of virulence in an opportunistic pathogen. These studies will not only inform on a clinically important pathogen, but also reveal the effects of population structure on the evolution of virulence in opportunistic pathogens that exist in multiple habitats, such as *Pseudomonas aeruginosa* and *Streptococcus pneumophilia.*