

## **Uropathogenic *Escherichia coli*: Population Structure and 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<sup>1-3</sup>. 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<sup>3</sup>. Approximately 80% of community acquired UTIs are caused by uropathogenic *E. coli* (UPEC), a Gram-negative gammaproteobacteria<sup>3,4</sup>. 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<sup>5,6</sup>. 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<sup>7-10</sup>. 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<sup>10</sup>. 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<sup>11-14</sup>. 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<sup>15</sup>. 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<sup>16–18</sup> and has been shown to affect the evolution of virulence in opportunistic pathogens<sup>19,20</sup>; 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<sup>21–24</sup>. Pathoadaptation can be defined as alterations in existing genes that occur as a result of specialization to new host-associated niches<sup>23,25–27</sup>. These pathoadaptive changes affect the population structure of pathogens, such as UPEC, by increasing the abundance of bacteria in the novel niche<sup>22,28</sup>. The functional effects of pathoadaptation 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<sup>15,23,25</sup>. These pathoadaptive changes have been identified in UPEC, and often result from modifications to a repertoire of virulence genes through processes of gene-loss<sup>29</sup>, refinement of existing pathogenicity factors<sup>21,24</sup>, or changes in expression of virulence genes through mechanisms such as phase-variation<sup>30,31</sup>.

A number of UPEC strains have been sequenced, including the standard model strains CFT073<sup>11</sup> and UTI89<sup>12</sup>, 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 accessory genome, which is comprised of all genes that exist in at least one, but not all, of the genomes that species<sup>7,11,12,32</sup>. The accessory 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, which is the collection of genes shared between all strains, indicating a highly reticulated population structure<sup>9,33,34</sup>. 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<sup>8,20,35–41</sup>.

While research into pathoadaptation and virulence factors is ubiquitous, specific definitions of these concepts are difficult to articulate<sup>15</sup>. The current definition of pathoadaptation 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<sup>15,23,25,26</sup>. 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<sup>15</sup>. 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.

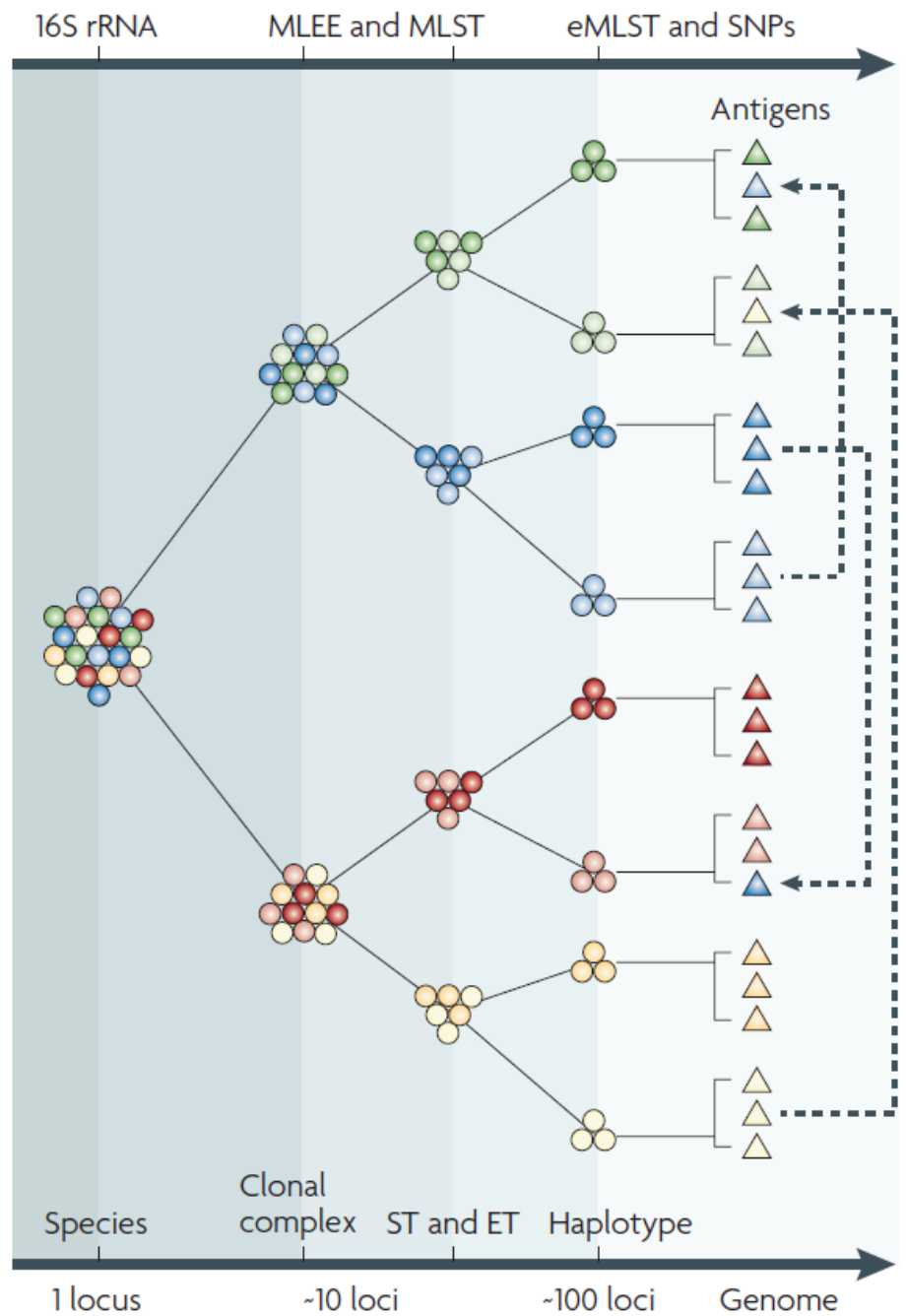
### **Population Dynamics of UPEC during UTI: Questions Remain**

Currently, pathogenic *E. coli* have been categorized according to their pathology and genomic content into a number of different clades<sup>34,42–46</sup>. These categories of *E. coli* are delimited by differences in the genomic content of the strains, which differ markedly between clades<sup>9,33,34</sup>. Within the UPEC clade, differences in genomic content are also found within the pan-genome, although allelic variation within core genes is also present<sup>11,12</sup>. These genomic differences have been used to identify the population and phylogenetic structure of UPEC between patients<sup>29,47–49</sup>, 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<sup>32,50</sup>. 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)<sup>32,50</sup>. 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)<sup>51</sup>. These limitations result, in part, to the high levels of recombination found in many bacterial pathogens, which obscures their phylogenetic history and population dynamics<sup>18,52,53</sup>. 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.

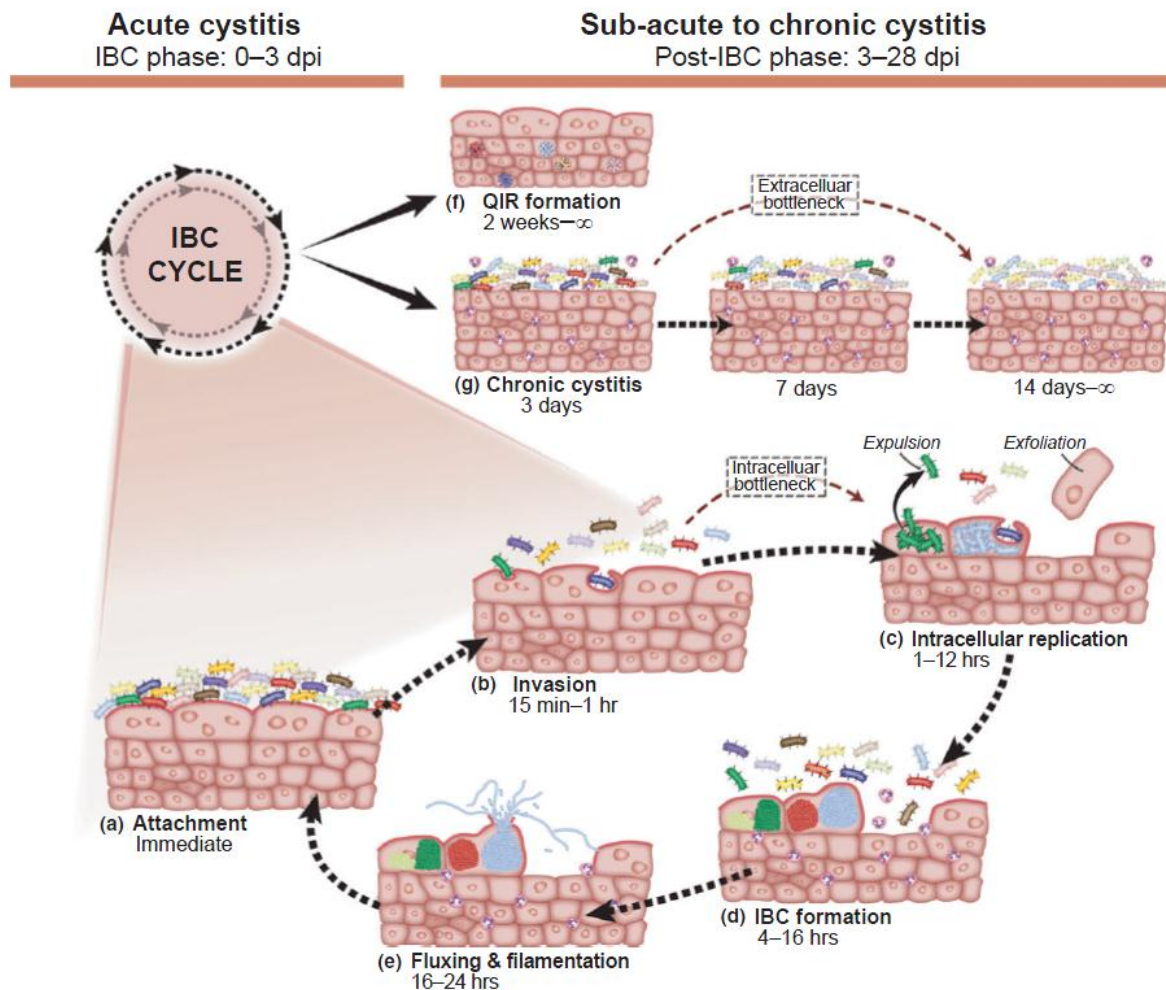
### *The Population Structure of UPEC is Affected by the Infection Cycle*

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



**Figure 1.** Schematic illustrating different levels of resolution offered by current genetic typing techniques. 16S can be used to identify species, while multi-locus enzyme electrophoresis (MLEE) and multi-locus sequence typing (MLST) offer subspecies resolution into clonal complexes. However, analysis of single-nucleotide polymorphisms (SNPs) allows for increased resolution to the strain level. Adapted from Medini et al. 2008.

progression is complicated and consists of a number of invasion events that restrict the population diversity, thus affecting the underlying population structure as proposed in a standard hypothetical model (Figure 2)<sup>54</sup>.



**Figure 2.** Hypothetical model for UPEC population bottlenecks in acute and change in population structure. This model depicts a hypothetical situation where two virulent and clonal UPEC populations, the Solids and the Pastels, are introduced into the urinary bladder. Passage through this acute bottleneck is linked to clonal expansion via the IBC cycle (a–e) during acute cystitis<sup>54</sup>. If these clones are equally fit to proceed through the IBC cycle, changes in population structure occur through stochastic mechanisms, affecting both the Solids and Pastels equally. QIR formation (f) during acute infection results in another round of stochastic loss in diversity and in the latent QIRs that persist in the urothelium after acute infection resolves, the Solids and Pastels persist equally well<sup>35,54,55</sup>. In this example, the Pastels are more fit than the Solids in competing for limited resources and eventually take over the bladder. Adapted from Hannan *et al.* 2012.

UPEC that invade the bladder are thought to originate in the gastrointestinal tract<sup>56-58</sup>, although direct evidence for this phenomenon has not been provided<sup>10</sup>. 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<sup>59</sup>. Following adherence, UPEC subsequently invade the epithelial cell and establishes a clonal community called an intracellular bacterial community (IBC) in a *fimH* dependent manner<sup>60,61</sup>. 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<sup>54</sup>. Alternatively, quiescent intracellular reservoirs (QIRs) may develop if the UPEC gain entry into the underlying epithelium below the superficial facet cells lining the bladder<sup>35</sup>. In such cases, UPEC may exist in a dormant state and emerge at a later time to cause a recurrent UTI<sup>35</sup>. 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<sup>54,62</sup>. 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*

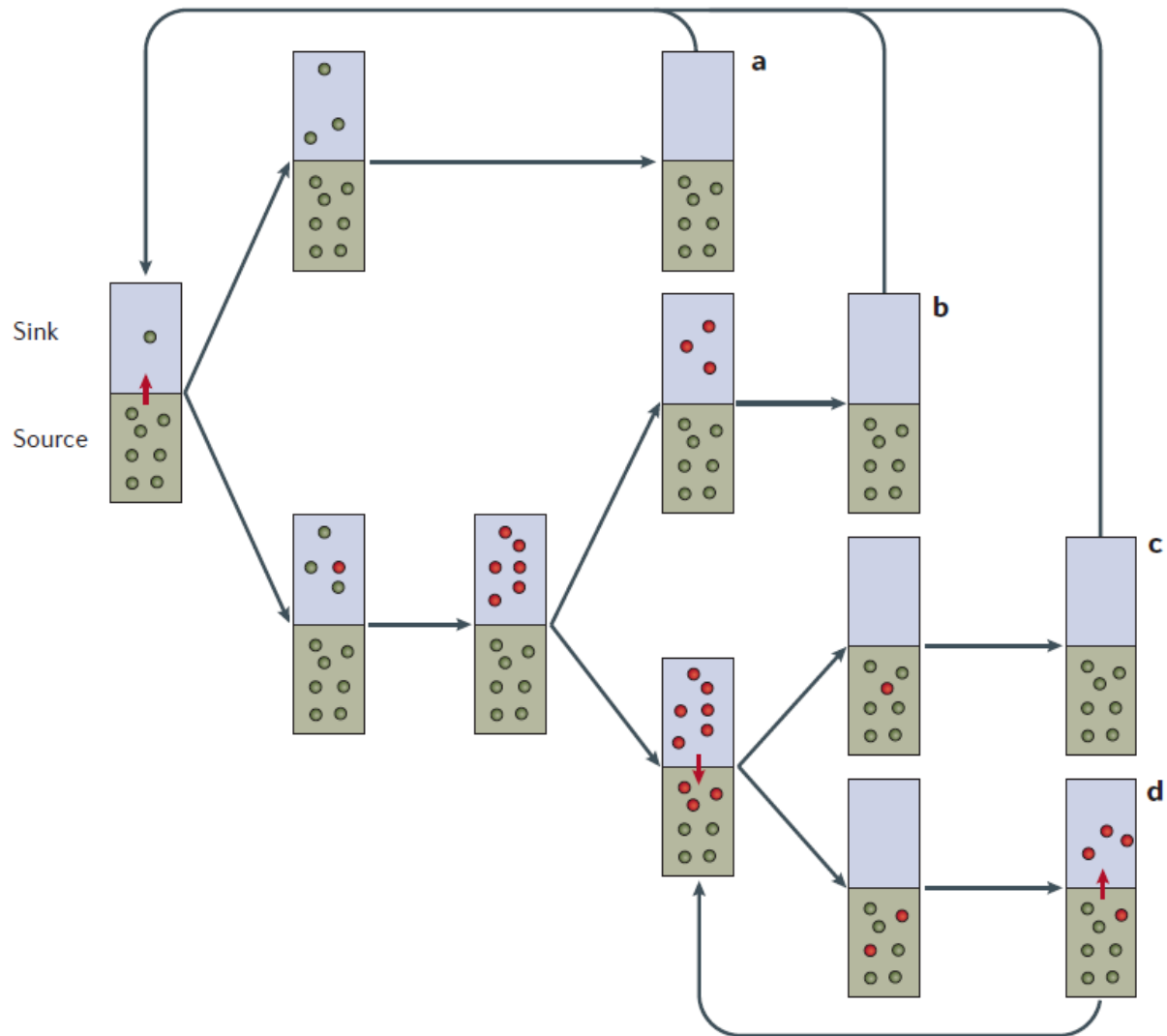
The source-sink model of population dynamics is an explanation of observed population dynamics and migration patterns between niches in which certain "source" habitats support the population of connected "sink" habitats<sup>63</sup>. In this model, populations that exist in source habitats experience 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<sup>22,64–66</sup>. 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<sup>22,23</sup>. 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 pathoadaptation 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<sup>23</sup>.

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<sup>67</sup>.

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<sup>23</sup>. 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<sup>4</sup>, 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<sup>28</sup>.





**Figure 3.** a | A scenario where there is no adaptive evolution during the course of infection; all new infections are caused by repeated invasion of the reservoir-adapted organisms and therefore there is no genetic signature. b | A scenario in which adaptive evolution occurs during the course of infection, but pathoadaptive mutants do not migrate back to the reservoir, or are eliminated very fast on migration, before reaching detectable numbers; virtually all new infections are caused by repeated invasion of the reservoir adapted organisms. The genetic signature associated with this scenario is that the evolutionarily short-lived pathoadaptive mutations are found exclusively in the infecting organisms. c | A scenario in which pathoadaptive mutants migrate back to the reservoir and circulate there in detectable numbers, but are eventually eliminated from the reservoir and replenished with new migrants; new infections can be caused by re-invasion of the pathoadaptive organisms from the reservoir, but their frequency and migration back to the reservoir is not significant enough to continuously sustain a particular pathoadaptive population there. The genetic signature that is associated with this scenario is that the evolutionarily short-lived pathoadaptive mutations are found commonly but not exclusively in the infecting organisms. d | Shows the same scenario as (c) but the pathoadaptive organisms re-invade the virulence niche and migrate back to the reservoir frequently enough to continuously sustain the pathoadaptive population in the reservoir habitat. The genetic signature associated with this scenario is that the pathoadaptive mutations are recently emerged but show signs of evolutionary stability and are found in both infecting and reservoir organisms, with a predominance in infecting organisms. Green box, reservoir, source habitat; blue box, virulence sink habitat; green dots, reservoir-adapted bacteria; red dots, patho-(sink)-adapted mutants; red arrows, inter-habitat migration. Adapted from Sokurenko *et al.* 2006

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<sup>68</sup> and invasion of uroepithelial cells by binding to mono-mannosylated uroplakin receptor UPA1A<sup>59</sup>, 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<sup>69</sup>. Subsequent *in vitro* functional investigations found that the polymorphisms in *fimH* resulted in altered binding affinities for different ligands expressed in differential tissues<sup>22,70</sup>. 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<sup>70</sup>. 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<sup>22</sup>. 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 source-sink model<sup>23,28,71</sup>. This model also predicts that, for these pathoadaptive 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*<sup>22,23,25,28</sup>. 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<sup>64</sup>. 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<sup>65</sup>. 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 Source-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<sup>72</sup>, 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 in redundant genes that encode proteins with similar function, such as Type 1 pili and Type S pili which bind different ligands, but are both involved in adherence and colonization<sup>10</sup>. 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<sup>73</sup>. These compensatory mutations are thought to occur very quickly<sup>74</sup> and can occur even in the absence of antibiotics<sup>73</sup>, 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<sup>75,76</sup>, 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<sup>74</sup>, 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 the FimH 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*<sup>70,75</sup>, knockout of the entire Type 1 pili apparatus, of which *fimH* is an integral part<sup>77</sup>, does not affect gut colonization by *E. coli in vivo*<sup>78,79</sup>. 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<sup>23,28,64,65</sup>. 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<sup>28</sup>. 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<sup>32,50</sup>. 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<sup>72</sup>. 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.

## **Experimental Design and Methods**

### *Rationale for Experiment:*

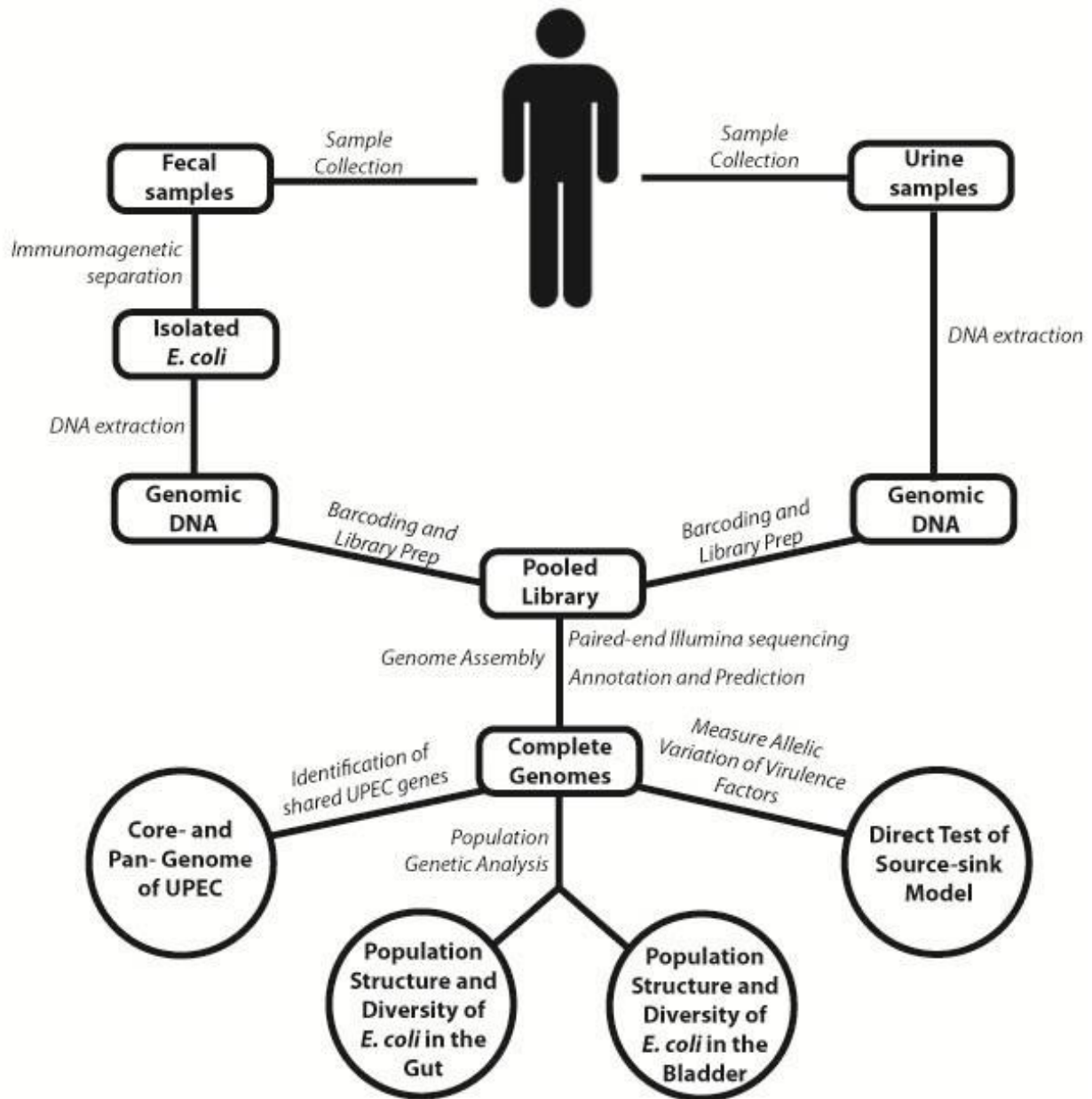
While the population structure and phylogeny of UPEC between patients has been studied extensively, the variability and distribution of UPEC within a single patient yet to fully articulated. Currently, the dominant model of population dynamics of UPEC during UTI is the source-sink model<sup>23,25</sup>. In this model, the *E. coli* population located in the gut is considered a source habitat with robust population growth that bolsters the transient sink population located in the bladder through repeated migrations<sup>23,25,28</sup>. Support for this claim has rested mainly on the presence of pathogen-adaptive (pathoadaptive) changes found within virulence genes, particular the adhesin gene *fimH*, which encodes a critical component of Type 1 pili that are involved in uropathogenicity<sup>22,28,64,65,69,70,76</sup>. In the model of source-sink dynamics, *E. coli* found

in the bladder acquire pathoadaptive mutations that increase their fitness in the bladder environment, however, these mutations are expected also to reduce fitness in the gut environment. As a result, *E. coli* carrying these mutations are expected to occur in a high abundance in the bladder and also occur at low abundance in the gut<sup>22,28,65</sup>. Additionally, because the bladder is considered an evolutionary dead-end, repeated invasions by distinct clones are expected to be responsible for recurrent UTI, rather than a single clonal subpopulation seeding multiple invasions<sup>22,28,65</sup>. Recent evidence suggests the source-sink model may not be accurate (paper in review).

Given the recent evidence, I hypothesize that *UPEC pathoadaptation that results in increased fitness in the bladder does not result in decreased fitness in the gut, meaning that the current source-sink model of UPEC UTI is not biologically accurate*. This hypothesis results in two predictions that are testable using pathogenomic analysis. **If the source-sink model of pathoadaptation and population dynamics is accurate, then (i) strains causing recurrent UTIs should be genotypically distinct from previous strains and (ii) the alleles of virulence factors that confer fitness in the bladder will exist at low frequency in the gut.**

### *Experimental Plan*

In order to fully investigate the effects of pathoadaptation on the population dynamics of UPEC during UTI, two questions must be answered: (i) What is the allelic frequency of known UPEC virulence factors in the gut and bladder habitats? and (ii) Do different clones of UPEC cause recurrent UTI? These questions can be answered through the comparison of whole-genome sequences of UPEC isolated from fecal and urine samples taken from patients suffering from recurrent UTIs (Figure 4).



**Figure 4.** Schematic detailing overview of experimental plan. Fecal and urine samples will be collected by immunomagnetic separation and selective plating, respectively. Genomic DNA will be extracted and prepped for paired-end sequencing on the Illumina HiSeq 2000 platform. Complete genomes will be analyzed to identify the core and accessory genomes of UPEC species, and the structure of the *E. coli* populations will be described and compared to ascertain clonality. Allele frequencies of virulence genes will be determined in order to identify patterns of clustering in either the gut or bladder populations.



**A. Cohort selection and disease definition:**

A cohort of 12-20 otherwise healthy, female patients between the ages of 18 and 44 suffering recurrent UTIs will be enrolled in a study and tracked for several months by a licensed urology clinic. In this study, UTI is defined using a combination of urinary tract symptoms and the presence of culturable bacteria in the urine<sup>4</sup>. The presence of more than  $10^4$  colony forming units (cfu) per mL of urine in samples collected using clean-catch method will be necessary for a diagnosis of UTI. Recurrent UTI is defined as a second UTI within 3 months of a previous episode. These definitions are necessary to maintain consistency between diagnoses and are standard in the field.

**B. Sample collection:**

Fecal and urine samples will be collected on a monthly interval. Additional samples will be collected during clinical visits for treatment of acute or recurrent UTI. This time frame will enable collection of samples before, during, and after recurrent UTI episodes.

**C. Strain isolation from urine, ribotyping, and DNA extraction:**

Bacteria found in urine samples will be selectively plated and subcultured to isolate single colonies for use in colony PCR to identify the strain causing the UTI. The colony PCR will target the variable region of the 16S rRNA gene, allowing for accurate identification of bacterial species. DNA will be then extracted from confirmed *E. coli* cultures.

**D. Strain isolation from feces, ribotyping, and DNA extraction:**

*E. coli* strains will be isolated from the feces using immunomagnetic separation (IMS) following standard protocols<sup>80</sup>. The antibodies used in IMS will target antigens located on the cell surface and are present in all subsets of the *E. coli* species, including UPEC. Isolation using IMS will result in less bias than selective plating techniques, and offer a more accurate

representation of gut microflora. Isolated bacteria will have their DNA extracted and amplified using multiple displacement amplification to obtain sufficient quantities for further analysis<sup>81,82</sup>. Isolation of *E. coli* will be confirmed using 16S ribotyping.

#### **D. Library preparation, and re-sequencing:**

Isolated genomic DNA will be prepared for paired-end sequencing on the Illumina HiSeq 2000 platform using manufacturer's protocols. Paired-end sequencing will be used as it will allow for the identification of inversion and insertion events in the UPEC genome and allow for the greatest resolution between strains within a bacterial species<sup>51</sup>. These data are necessary, given the size of the *E. coli* mobilome and the fact that several important virulence genes in UPEC are regulated through phase variation of promoters<sup>31</sup>.

#### **E. Genome assembly:**

Raw data from the Illumina sequencing process will be used to assemble, map, and annotate whole-genome sequences using standard programs. Bioinformatic analyses will be used to identify shared genetic components common to all of the strains collected from urine samples (a UPEC core genome) as well as the components unique to one or some strains isolated from the bladder (UPEC accessory genome)<sup>32</sup>.

#### **F. Bioinformatic analysis and hypothesis testing:**

Further analysis identify SNPs in the core genome that exist between all strains, which can then be used to identify the level of similarity between strains from the same patient during recurrent UTI, **which is a direct test of the repeated *de novo* mutation and invasion prediction of the source-sink model**. Whole-genome analysis of SNPs will result in a very high-resolution delineation between UPEC strains that is not possible using multi-locus sequence typing or 16S ribotyping. Whole-genome sequence data will then be used to analyze

the read-depth and sequence variation in known urovirulence genes in a semi-quantitative analysis to the relative abundance of the urovirulent alleles in both bladder and habitat, **which is a direct test of the differential allelic frequency prediction of the source-sink model.**

Several studies have identified urovirulent alleles for a number of virulence factors in UPEC<sup>12,24,69</sup>, and the fitness effects of SNP differences in virulence genes are currently being analyzed using *in vivo* competition experiments in a murine model in the Hultgren lab.

### *Expected Results*

These experiments will result in: (i) a phylogenetic analyses that can be used to identify clonality between strains and (ii) a semi-quantitative measure of urovirulent allele frequency in *E. coli* populations found in the gut and bladder environments.

***If strains isolated from the same patient during multiple occurrences of UTI are clonal, then the prediction of the source-sink model that de novo mutations and multiple invasions by genotypically distinct strains results in recurrent UTIs is not biologically accurate in all cases.*** Statistical significance will be measured by the number patients that display this clonal re-invasion phenomenon in their recurrent UTI. It is possible that repeated *de novo* mutations can result in the same pathoadaptive mutations, thus confounding this analysis, but, given the range of pathoadaptive alleles for urovirulence<sup>21</sup>, this phenomenon is unlikely to occur by chance.

***If the semi-quantitative measure of urovirulence alleles shows that alleles in known virulence genes isolated from UPEC in the bladder exist at a high frequency in the gut environment at the same time point, then it is likely that the urovirulent alleles do not reduce fitness in the gut given the context of the UPEC genome and the host environment.*** Given the fact that these urovirulence alleles may exist in high frequency in the gut due to linkage disequilibrium or compensatory mutations elsewhere in the genome, the fitness effects of these alleles will require further investigation. This investigation will occur

using a two-way competition experiment in a murine model of the gut between isogenic UPEC strains with or without the urovirulent allele. This analysis will show if the pathoadaptive allele has a fitness cost in other environments, and how UPEC mediate this fitness cost. These data will inform the biological relevancy of the source-sink model as it pertains to the whole genome of UPEC, rather than single genes.

Taken together these interpretations will aid our understanding of the population dynamics that occur in UPEC during UTI. Additionally, they will also inform the source-sink model of UPEC UTI, which, as recent evidence indicates, may need revision. Ultimately, this information will be used to better understand and predict the evolution of opportunistic pathogens, such as UPEC, to better design strategies of prevention and therapeutic intervention.

### Literature Cited:

1. Hooton, T. M. Recurrent urinary tract infection in women. *International Journal of Antimicrobial Agents* **17**, 259–268 (2001).
2. Scholes, D. *et al.* Risk Factors for Recurrent Urinary Tract Infection in Young Women. *The Journal of infectious diseases* **182**, 1177–1182 (2000).
3. Foxman, B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *The American journal of medicine* **113 Suppl**, 5S–13S (2002).
4. Foxman, B. The epidemiology of urinary tract infection. *Nature Publishing Group* **7**, 653–660 (2010).
5. Nicolas-Chanoine, M. H. *et al.* Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *Journal of Antimicrobial Chemotherapy* **61**, 273–281 (2007).
6. Rogers, B. A., Sidjabat, H. E. & Paterson, D. L. Escherichia coli O25b-ST131: a pandemic, multiresistant, community-associated strain. *Journal of Antimicrobial Chemotherapy* **66**, 1–14 (2010).
7. Dobrindt, U. (Patho-)Genomics of Escherichia coli. *International Journal of Medical Microbiology* **295**, 357–371 (2005).
8. Dobrindt, U., Chowdary, M. G., Krumbholz, G. & Hacker, J. Genome dynamics and its impact on evolution of Escherichia coli. *Medical microbiology and immunology* **199**, 145–54 (2010).
9. Touchon, M. *et al.* Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS genetics* **5**, e1000344 (2009).
10. Hannan, T. J. *et al.* Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic Escherichia coli bladder infection. *FEMS microbiology reviews* **36**, 616–48 (2012).
11. Welch, R. a *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 17020–4 (2002).
12. Chen, S. L. *et al.* Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5977–82 (2006).
13. Luo, C., Hu, G.-Q. & Zhu, H. Genome reannotation of Escherichia coli CFT073 with new insights into virulence. *BMC genomics* **10**, 552 (2009).

14. Avasthi, T. S. *et al.* Genome of multidrug-resistant uropathogenic *Escherichia coli* strain NA114 from India. *Journal of Bacteriology* **193**, 4272–4273 (2011).
15. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449**, 835–42 (2007).
16. Gogarten, J. P. & Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* **3**, 679–687 (2005).
17. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
18. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic acids research* **36**, 6688–719 (2008).
19. Canchaya, C., Fournous, G. & Brüssow, H. The impact of prophages on bacterial chromosomes. *Molecular microbiology* **53**, 9–18 (2004).
20. Schneider, G. *et al.* Mobilisation and remobilisation of a large archetypal pathogenicity island of uropathogenic *Escherichia coli* in vitro support the role of conjugation for horizontal transfer of genomic islands. *BMC microbiology* **11**, 210 (2011).
21. Chen, S. L. *et al.* Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 22439–44 (2009).
22. Sokurenko, E. V. *et al.* Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proceedings of the ...* (1998).at <<http://www.pnas.org/content/95/15/8922.short>>
23. Sokurenko, E. V., Gomulkiewicz, R. & Dykhuizen, D. E. Source–sink dynamics of virulence evolution. *Nature Reviews Microbiology* **4**, 548–555 (2006).
24. Iebba, V. *et al.* Microevolution in fimH Gene of Mucosa-Associated *Escherichia coli* Strains Isolated from Pediatric Patients with Inflammatory Bowel Disease. *Infection and ...* (2012).at <<http://iai.asm.org/content/80/4/1408.short>>
25. Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends in microbiology* **7**, 191–195 (1999).
26. Rakin, A., O’Connell, K. P., Skowronski, E. W., Sulakvelidze, A. & Bakanidze, L. Pathoadaptation of Especially Dangerous Pathogens. *Emerging and Endemic Pathogens* **00**, 109–118 (2010).
27. Maurelli, A. T. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS microbiology letters* **267**, 1–8 (2007).
28. Weissman, S. J. *et al.* Differential stability and trade-off effects of pathoadaptive mutations in the *Escherichia coli* FimH adhesin. *Infection and Immunity* **75**, 3548–3555 (2007).

29. Salvador, E. *et al.* Comparison of asymptomatic bacteriuria *Escherichia coli* isolates from healthy individuals versus those from hospital patients shows that long-term bladder colonization selects for attenuated virulence phenotypes. *Infection and immunity* **80**, 668–78 (2012).
30. Grozdanov, L. & Raasch, C. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *Journal of ...* (2004).doi:10.1128/JB.186.16.5432
31. Hannan, T. J. *et al.* LeuX tRNA-dependent and -independent mechanisms of *Escherichia coli* pathogenesis in acute cystitis. *Molecular Microbiology* **67**, 116–128 (2008).
32. Medini, D. *et al.* Microbiology in the post-genomic era. *Nature reviews. Microbiology* **6**, 419–30 (2008).
33. Rasko, D. A. *et al.* The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. *Journal of Bacteriology* **190**, 6881–6893 (2008).
34. Chaudhuri, R. R. & Henderson, I. R. The evolution of the *Escherichia coli* phylogeny. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 214–26 (2012).
35. Mysorekar, I. U. & Hultgren, S. J. Mechanisms of uropathogenic *Escherichia coli* persistence and eradication from the urinary tract. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14170–5 (2006).
36. Reigstad, C. S., Hultgren, S. J. & Gordon, J. I. Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled. *The Journal of biological chemistry* **282**, 21259–67 (2007).
37. Lloyd, A. L., Henderson, T. a, Vigil, P. D. & Mobley, H. L. T. Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. *Journal of bacteriology* **191**, 3469–81 (2009).
38. Bauchart, P. *et al.* Pathogenomic comparison of human extraintestinal and avian pathogenic *Escherichia coli*--search for factors involved in host specificity or zoonotic potential. *Microbial Pathogenesis* **49**, 105–115 (2010).
39. Nielubowicz, G. R. & Mobley, H. L. T. Host–pathogen interactions in urinary tract infection. *Nature Publishing Group* **7**, 430–441 (2010).
40. Norinder, B. S., Köves, B., Yadav, M., Brauner, A. & Svanborg, C. Do *Escherichia coli* strains causing acute cystitis have a distinct virulence repertoire? *Microbial pathogenesis* **52**, 10–6 (2012).
41. Johnson, T. J. *et al.* Prevalence of avian-pathogenic *Escherichia coli* strain O1 genomic islands among extraintestinal and commensal *E. coli* isolates. *Journal of Bacteriology* **194**, 2846–2853 (2012).

42. Nataro, J. P. & Kaper, J. B. Diarrheagenic *Escherichia coli*. *Clinical microbiology reviews* (1998).at <<http://cmr.asm.org/content/11/1/142.short>>
43. Boudeau, J., Glasser, A. L., Masseret, E., Joly, B. & Darfeuille-Michaud, A. Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infection and Immunity* **67**, 4499–4509 (1999).
44. Darfeuille-Michaud, A. Adherent-invasive *Escherichia coli*: a putative new *E. coli* pathotype associated with Crohn's disease. *International Journal of Medical Microbiology* **292**, 185–193 (2002).
45. Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology* **2**, 123–140 (2004).
46. Russo, T. a & Johnson, J. R. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *The Journal of infectious diseases* **181**, 1753–4 (2000).
47. Rijavec, M. *et al.* High Prevalence of Multidrug Resistance and Random Distribution of Mobile Genetic Elements Among Uropathogenic *Escherichia coli* (UPEC) of the Four Major Phylogenetic Groups. *Current Microbiology* **53**, 158–162 (2006).
48. Poey, M. E., Albin, M., Saona, G. & Laviña, M. Virulence profiles in uropathogenic *Escherichia coli* isolated from pregnant women and children with urinary tract abnormalities. *Microbial Pathogenesis* **52**, 292–301 (2012).
49. Gibreel, T. M. *et al.* Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *Journal of Antimicrobial Chemotherapy* **67**, 346–356 (2012).
50. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology* (2008).doi:10.1038/nrmicro1872
51. Loman, N. J. *et al.* High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology* **10**, 599–606 (2012).
52. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Molecular biology and evolution* **19**, 2226–2238 (2002).
53. Doolittle, W. F. & Papke, R. T. Genomics and the bacterial species problem. *Genome Biology* (2006).at <<http://www.biomedcentral.com/content/pdf/gb-2006-7-9-116.pdf>>
54. Schwartz, D. J., Chen, S. L., Hultgren, S. J. & Seed, P. C. Population dynamics and niche distribution of uropathogenic *Escherichia coli* during acute and chronic urinary tract infection. *Infection and immunity* **79**, 4250–9 (2011).
55. Hannan, T. J., Mysorekar, I. U., Hung, C. S., Isaacson-Schmid, M. L. & Hultgren, S. J. Early severe inflammatory responses to uropathogenic *E. coli* predispose to chronic and recurrent urinary tract infection. *PLoS pathogens* **6**, e1001042 (2010).



56. Moreno, E. *et al.* Relationship between *Escherichia coli* strains causing urinary tract infection in women and the dominant faecal flora of the same hosts. *Epidemiology and infection* **134**, 1015–1023 (2006).
57. Moreno, E. *et al.* Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E. coli* population of the host. *Journal of Clinical Microbiology* **46**, 2529–2534 (2008).
58. Yamamoto, S. *et al.* Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. *The Journal of urology* **157**, 1127–1129 (1997).
59. Zhou, G. *et al.* Uroplakin Ia is the urothelial receptor for uropathogenic *Escherichia coli*: evidence from in vitro FimH binding. *Journal of Cell Science* **114**, 4095–4103 (2001).
60. Mulvey, M. A. *et al.* Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* **282**, 1494–1497 (1998).
61. Anderson, G. G. *et al.* Intracellular bacterial biofilm-like pods in urinary tract infections. *Science (New York, N. Y.)* **301**, 105–7 (2003).
62. Walters, M. S. *et al.* Kinetics of uropathogenic *Escherichia coli* metapopulation movement during urinary tract infection. *mBio* **3**, (2012).
63. Pulliam, H. R. Sources, sinks, and population regulation. *American Naturalist* 652–661 (1988).at <<http://www.jstor.org/stable/10.2307/2461927>>
64. Sokurenko, E. V. *et al.* Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli*. *Molecular biology and evolution* **21**, 1373–1383 (2004).
65. Chattopadhyay, S. *et al.* Haplotype Diversity in “Source-Sink” Dynamics of *Escherichia coli* Urovirulence. *Journal of Molecular Evolution* **64**, 204–214 (2006).
66. Chattopadhyay, S., Paranjpye, R. N., Dykhuizen, D. E., Sokurenko, E. V. & Strom, M. S. Comparative evolutionary analysis of the major structural subunit of *Vibrio vulnificus* type IV pili. *Molecular biology and evolution* **26**, 2185–2196 (2009).
67. Alteri, C. J. & Mobley, H. L. T. *Escherichia coli* physiology and metabolism dictates adaptation to diverse host microenvironments. *Current Opinion in Microbiology* **15**, 3–9 (2012).
68. Hung, C.-S. *et al.* Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Molecular Microbiology* **44**, 903–915 (2002).
69. Sokurenko, E. V., Courtney, H. S., Ohman, D. E., Klemm, P. & Hasty, D. L. FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. *Journal of Bacteriology* **176**, 748–755 (1994).

70. Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. & Hasty, D. L. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in fimH genes. *Journal of ...* (1995).at  
<<http://jb.asm.org/content/177/13/3680.short>>
71. Chattopadhyay, S. *et al.* High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12412–12417 (2009).
72. Shoal, O. *et al.* Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157–1160 (2012).
73. Levin, B. R., Perrot, V. & Walker, N. Compensatory Mutations, Antibiotic Resistance and the Population Genetics of Adaptive Evolution in Bacteria. *Genetics* (2000).at  
<<http://www.genetics.org/content/154/3/985.short>>
74. Sousa, A., Magalhães, S. & Gordo, I. Cost of antibiotic resistance and the geometry of adaptation. *Molecular biology and evolution* **29**, 1417–1428 (2012).
75. Schembri, M. A., Sokurenko, E. V. & Klemm, P. Functional flexibility of the FimH adhesin: insights from a random mutant library. *Infection and Immunity* **68**, 2638–2646 (2000).
76. Weissman, S. J. *et al.* Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Molecular Microbiology* **59**, 975–988 (2006).
77. Hultgren, S. J., Porter, T. N., Schaeffer, A. J. & Duncan, J. L. Role of type 1 pili and effects of phase variation on lower urinary tract infections produced by *Escherichia coli*. *Infection and Immunity* **50**, 370–377 (1985).
78. Bloch, C. A., Stocker, B. A. D. & Orndorff, P. E. A key role for type 1 pili in enterobacterial communicability. *Molecular Microbiology* **6**, 697–701 (1992).
79. McCormick, B. A., Franklin, D. P., Laux, D. C. & Cohen, P. S. Type 1 pili are not necessary for colonization of the streptomycin-treated mouse large intestine by type 1-piliated *Escherichia coli* F-18 and *E. coli* K-12. *Infection and ...* (1989).at  
<<http://iai.asm.org/content/57/10/3022.short>>
80. Enroth, H. & Engstrand, L. Immunomagnetic separation and PCR for detection of *Helicobacter pylori* in water and stool specimens. *Journal of Clinical Microbiology* **33**, 2162–2165 (1995).
81. Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* **11**, 1095–1099 (2001).
82. Binga, E. K., Lasken, R. S. & Neufeld, J. D. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *The ISME journal* **2**, 233–241 (2008).

