Winter 12-15-2017

Understanding the paradox of genetic diversity in uropathogenic E. coli: the uncommon evolution of a common pathogen

Henry Louis Schreiber
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
Scott Hultgren, Chair
Gautam Dantas
Ashlee Earl
Jeffrey Henderson
Phillip Tarr

Understanding the Paradox of Genetic Diversity in Uropathogenic *E. coli*: The Uncommon Evolution of a Common Pathogen

by
Henry L. Schreiber IV

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy of Philosophy

December 2017
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank Dr. Scott Hultgren for being a peerless mentor, an unrelenting champion, and a dedicated coach during these past years. The sincerity and passion you bring to science is inspiring. Without your constant support and unflagging enthusiasm, I would not have been able to risk so much, and thus be so rewarded. Any success that I have moving forward is a reflection of the foundation you have given me. Thank you.

Dr. Ashlee Earl, thank you for your limitless optimism, your zeal in tackling the big ideas, and your patience in collaboration. I have learned how to be a good collaborator from our work. I aspire to match your ability to work with intelligence and innovation in so many fields.

I would like to thank my thesis committee for their guidance and support. I have pursued difficult topics and I would not have succeeded without your advice and benevolent critique. I would like to thank Dr. Phil Tarr, who has always grounded me and taught me how to be skeptical without being uncivil. Thank you to Dr. Jeff Henderson whose guidance early in a critical conversation early in my career gave me impetus to pursue the identity of UPEC. I would also like to thank Dr. Gautam Dantas, who has given me thoughtful and sincere advice on every aspect of graduate school, including how have an outside life.

All of the members of the Hultgren lab, past and present, have been critical for my success. The daily affirmations, wisdom, and congeniality made my life better and always made me happier to be around you. In particular, I would like to thank the support and daily mentorship from Dr. Karen Dodson, Dr. Mike Hibbing, and Dr. Matt Conover, who were instrumental in my training. I would also like to thank Dr. Drew Schwartz and Dr. Sarah Greene, who welcomed me into the lab as a rotation student and who helped me build my ideas. I also want to thank Dr. Valerie O'Brien and Dr. Caitlin Spaulding, who

joined the lab at the same time as me and who have been wonderful companions and collaborators through all of graduate school.

I knew I wanted to be a scientist at an early age, and it was my family who helped me through difficult times so that I could be here. My father taught me about science, my mother taught me how to be a human being, and my sisters taught me how to annoy everyone with constant questions. Thank you for all the love you have given me throughout my entire life. There isn't the space to give you all the credit you deserve.

To my amazing wife, Amanda, thank you. Every day, you make me want to be a better man. Life with you is full and brilliant, even when life seems like it is closing in around me. Your smiles and laughter lighten any room and I am in awe of your gentle kindness. Thank you for being my shelter and my warmth.

My work would not have been possible without the funding support of the Lucille P. Markey Pathway in Pathobiology or the Monsanto Excellence Fund Gradute Student Fellowship. This stipend support allowed me the freedom to pursue a risky project.

Henry L. Schreiber IV

Washington University in St. Louis

December 2017

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Understanding the paradox of genetic diversity in uropathogenic *E. coli*:

the uncommon evolution of a common pathogen by

Henry L. Schreiber IV

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2017

Professor Scott J. Hultgren, Chair

Urinary tract infections (UTIs) are the second most common bacterial infection of people in the U.S.A and are frequently recurrent, as an initial UTI is quickly followed by a second episode in 30-35% of cases despite appropriate antibiotic treatment and clearance of the bacteria from the urine. The vast majority, >80%, of UTIs are caused by uropathogenic *Escherichia coli* (UPEC). UPEC that colonize the bladder are thought to originate in the gut, where they live as commensal organisms. UPEC can be shed in feces to colonize the vagina and/or periurethral area, and then can ascend into the bladder to start a UTI. *E. coli* strains, including UPEC, have been sub-divided into clades (e.g., clades A, B1, B2 and D) based on their genetic relatedness. In the U.S.A, most (50-75%) UPEC fall into clade B2 while the rest (25-50%) are spread through clades A, B1, and D. Many UPEC encode a variety of putative urovirulence factor genes that are thought to enable bladder colonization and whose carriage in has been correlated with both UTI and recurrence in humans. However, in contrast to many other *E. coli* pathotypes and despite decades of research, a clear, genetic definition of UPEC remains elusive. Towards this goal, I pursued a research strategy integrating multiple fields of study, including large-scale bioinformatic analysis, *in vitro* and *in vivo* modeling of pathogenesis, and structural biology, within a holistic view of the UPEC evolutionary history that incorporates their residence in both

the gut and the bladder. Thus, I have shown that clinical UPEC are genetically heterogeneous and that gene carriage alone is not a robust predictor of UPEC's ability to colonize the bladder in mouse models of cystitis. Instead, I have found the transcriptional regulation of core genes shared by all *E. coli* strains can be used to predict the outcome of bladder infections in mice. Further, I have found that evolution has stringently conserved bacterial behaviors that are critical to both bladder and gut colonization by *E. coli*, namely the tension and unwinding of the type 1 pilus rod in response to shear stress. The type 1 pilus is found in the vast majority of *E. coli* strains and nearly every UPEC isolate and has been shown to be critical in bladder colonization in animal models of cystitis, thus underscoring the fact that bacterial features enabling uropathogenicity are common and conserved across many *E. coli* strains. Finally, I have shown that clade B2 UPEC have adopted genetic tools from other gut bacteria that provide them with a selective advantage in gut colonization and persistence, potentially enhancing their ability to cause recurrent UTIs. This may explain why B2 strains are enriched in UPEC overall, especially in those strains causing recurrent UTI, despite the fact that both B2 and non-B2 strains can be robust colonizers of the bladder. Taken together, these findings indicate the bladder pathogenesis may be a "core feature" of most *E. coli* and that the definition of UPEC may be related more to the core bacterial behaviors enabling persistence and survival in multiple body sites than any one specific virulence mechanism or carriage of certain genes. These findings extend beyond UPEC to other bacterial diseases, such as respiratory infections caused by *Klebsiella* or Pneumococcoccus, where bacteria transition from commensal lifestyles in one habitat to pathogenic lifestyles in another body site and further work is needed to understand how conserved bacterial features may be coopted for pathogenicity in the new environment

Chapter 1:Introduction

Excerpted from:

From physiology to pharmacy: developments in the pathogenesis and treatment of recurrent urinary tract infections.

By

Jennifer A. Silverman, Henry L. Schreiber IV, Thomas M. Hooton, Scott J. Hultgren

Current Urology Reports

2013 October. 14(5):448-56. doi: 10.1007/s11934-013-0354-5 PMID 23832844

AND

One size doesn't fit all: unraveling the diversity of factors and interactions that drive *E. coli* urovirulence

By

Henry L. Schreiber IV, Caitlin N. Spaulding, Jonthan. Livny, Scott J. Hultgren

Annals of Translational Medicine

2017 January. 5(2):28 doi: 10.21037/atm.2016.12.73. PMID 28217693

1.1 Overview

Uropathogenic *Escherichia coli* cause the majority of urinary tract infections (UTIs) and are highly recurrent, owing, in part, to their ability to colonize host reservoirs outside of the bladder, such as the gut. In this chapter, I will provide a brief description of the prevailing model of UPEC virulence and the gaps in knowledge within that model. Next I will describe how the UPEC pathotype is unique amongst pathogenic groups of *E. coli* due to the lack of a clear, genetic definition of UPEC strains. Finally, I will conclude with calls for further study of host reservoirs of UPEC outside of the bladder to better understand the totality of the UPEC persistence within hosts. In future chapters, I will describe my work in better understanding the evolution and conservation of features in UPEC that enable their uropathogenesis.

1.2 Understanding uropathogenic *Escherichia coli*

1.2.1 Epidemiology of urinary tract infections

Urinary tract infections (UTIs) are the second most common bacterial infections encountered in the U.S.A with a significant predilection towards women. More than 50% of women experience at least one UTI during their lifetime, and each year approximately 11% of women report having had a UTI (Foxman, 2010). Despite appropriate treatment for the initial infection, UTIs are frequently recurrent, as 25-35% of initial infections are followed by a recurrent UTI (rUTI) (Foxman et al., 2000; Scholes et al., 2000). This high frequency and recurrence leads to significant healthcare costs; in the United States alone, UTIs cost greater than $5 billion dollars per annum including direct and indirect costs (Foxman, 2014). Uncomplicated UTIs (hereafter referred to simply as UTIs) are infections of the lower urinary tract that are not associated with functional or anatomical abnormalities, diabetes, pregnancy, or urinary catheterization, while all other infections are labeled as complicated. The clinical manifestations

and symptomatologies of UTIs are strikingly complex and result from the interactions between diverse uropathogens and host urothelial tissues and include dysuria, foul-smelling or cloudy urine, and fever (Hooton, 2012). UTIs are generally short-lived due to antimicrobial therapy; however, UTIs can last for months in women without antibiotic treatment (Ferry et al., 2004; Mabeck, 1972). Alarmingly, many uropathogens are becoming increasingly resistant to many of the antimicrobials used to treat UTI (Zowawi et al., 2015), thus highlighting the urgent need to develop and deploy new therapeutic options.

1.2.2 Current models of uropathogenic *Escherichia coli* (UPEC) virulence

The most common organism causing UTI is uropathogenic *Escherichia coli* (UPEC), which causes >80% of UTIs (Foxman, 2010). UPEC that colonize the urinary tract are thought to originate from the gut, where they are a minor member of the healthy, adult gut microbiota (Consortium, 2012). UPEC strains, along with all *E. coli*, have been categorized into different clades (e.g., clades A, B1, B2, D, and E) based on their genetic relatedness (Tenaillon et al., 2010). While UPEC strains tend to be genetically heterogeneous, the majority of UPEC strains isolated from women with UTI in the U.S.A. reside in the B2 clade. Notably, clade B2 also contains all of the commonly studied model UPEC strains, including strains UTI89 (Chen et al., 2006), CFT073 (Welch et al., 2002), and 536 (Brzuszkiewicz et al., 2006). Epidemiological studies have shown that host susceptibility is variable and mediated by numerous genetic, environmental, and behavioral risk factors (Scholes et al., 2000). Many of these risk factors have been investigated in experimental mouse infection models. Mice have similar bladder morphology and express many of the same surface proteins as humans and respond to infection with phenotypes that recapitulate clinical findings in humans (e.g., epithelial exfoliation, regulation of cytokine production, pyuria, and expression of antimicrobial peptides, as reviewed

in (Carey et al., 2016) and (Barber et al., 2016)). Using many of these mouse models of cystitis, model, clade B2 UPEC strains have been used to characterize the pathogenic cascade during bladder infection (Hannan et al., 2010; Hung et al., 2009). After ascension into the bladder, UPEC use a specialized adhesive organelle, called type 1 pili, to attach to and subsequently invade superficial umbrella cells lining the bladder lumen (Martinez et al., 2000; Mulvey et al., 1998). Bladder invasion enables UPEC to evade the host immune responses, such as phagocytosis by neutrophils, and UPEC that are unable to invade the epithelial bladder cells are much more likely to be cleared from the bladder (Wright et al., 2005). After invasion, UPEC escape their enclosing vesicle through unknown mechanisms and expand into biofilm-like intracellular bacterial communities (IBCs) (Anderson et al., 2003; Justice et al., 2004) comprised of 10^4-10^5 cells (Wright et al., 2007). IBC development is also dependent on type 1 pili (Wright et al., 2007). In response to attachment, host cells undergo apoptosis and exfoliation (Thumbikat et al., 2009a; 2009b) and, after invasion, UPEC can be exocytosed by the host cells in a TLR-4 dependent process (Song et al., 2007). If not lost to expulsion and host cell exfoliation, bacteria can then flux out of the mature IBCs back into the lumen of the bladder, where they can invade naïve epithelial cells again (Justice et al., 2006). After an infection is cleared, latent bacterial cells, termed quiescent intracellular reservoirs (QIRs), can remain in underlying or superficial bladder epithelial tissue and are capable of causing recurrent UTIs (Mysorekar and Hultgren, 2006). Another potential outcome of these acute events is the establishment of long-lasting, chronic cystitis (Hannan et al., 2010). Chronic cystitis in mice is characterized by persistent, high titer bacteriuria ($>10^4$ colony forming units (CFU)/mL) accompanied by chronic inflammation, and urothelial necrosis and high titer bacterial bladder burdens at sacrifice (Hannan et al., 2010). Importantly, the elucidation of all of these pathogenic mechanisms in mouse models has relied

almost exclusively on model UPEC strains, such as UTI89 and CFT073, which represent only a small section of the overall *E. coli* phylogenetic tree, despite the remarkable diversity of UPEC strains in the clinic. Thus, it is unknown whether the genetically diverse UPEC strains found in the clinic follow the same pathogenic pathways or how diverse UPEC interact with different host susceptibility factors.

1.2.3 The problem of recurrence

The virulence of UPEC is further enhanced by their ability to persist in reservoirs between multiple UTI episodes to cause recurrent bouts of the disease. UTIs recur in 25-35% of young women with cystitis within 6 months after their first episode (Foxman et al., 2000), and the recurrence rate increases with more than 1 prior UTI (Foxman et al., 2000; Scholes et al., 2000). Recurrent UTIs are common in women at all ages, and some women are troubled with frequent recurrences throughout their lives. Although several mechanisms have been postulated to explain recurrent UTI, this syndrome is not well understood. Recurrent UTIs caused by the same UPEC strain can arise from re-colonization of the bladder from UPEC persisting in reservoirs in the gut, vagina (Chen et al., 2013; Hooton, 2012) or possibly in the bladder itself (Mysorekar and Hultgren, 2006). Recurrences with strains different than the first infection only occur through the invasion of the bladder by another strain. When different-strain recurrent UTIs occur, the new strain causing the UTI may originate from the patient's gut (Chen et al., 2013), or be introduced from another environmental source. The rates of same-strain versus different-strain recurrences have been investigated for nearly five decades, but estimates of the percentage of same-strain recurrences vary greatly between studies from 16-82% (Ejrnaes, 2011; Foxman et al., 2000; 1995; Ikaheimo et al., 1996; Kärkkäinen et al., 2000; Kõljalg et al., 2009; Kunin, 1970; Luo et al., 2012). This dramatic range of results may be due to differences in the (1)

demographics of the cohort studied, (2) urine collection and bacterial culture methods, (3) definition of symptoms, (4) treatment regimens, (5) duration of follow-up, and (6)strain typing methods.

1.3 Questioning uropathogenic *Escherichia coli*

1.3.1 UPEC: a common pathogen without a clear definition

Currently, UPEC are defined as any strain that is recovered from the urine of symptomatic UTI patients (Marrs et al., 2005), a classification that fails to account for differences in host susceptibility or the possibility for multiple evolutionary and mechanistic paths to urovirulence. This clinical definition of UPEC stands in contrast to other *E. coli* pathotypes, which are clearly defined by sets of virulence genes, such as the stx gene in Shiga-toxin producing *E. coli* (Croxen et al., 2013; Marrs et al., 2005). The clinical definition of UTI is also plastic as many clinicians define symptomatic criteria differently. For example, while many clinicians require a bacterial burden of $>10^3$ CFU/mL of UPEC in urine to diagnose a UPEC UTI, other clinicians diagnose UTIs with a cutoff of 10^2 to 10^6 CFU/mL (Foxman, 2010). These differing cutoffs are related to the difficulty in collecting urine from female patients without contamination from the vaginal or periurethral flora, which are body sites that *E. coli* can also colonize. Thus, a single *E. coli* strain infecting a single patient's bladder may be considered a UPEC strain in some clinics while being considered as not significant in other clinics. Further, patient susceptibility to UTIs is known to vary due to genetic, environmental, and behavioral differences between women (Foxman, 2010; Hooton, 2012). One of the most significant risk factors for developing a UTI is a previous history of UTIs (Hooton et al., 1996). Recent research has provided insight into this phenomenon by showing in mice that a severe initial UTI results in a remodeling of the bladder epithelium (Hannan et al., 2010; O'Brien et al., 2016). This

remodeling changes the course of subsequent infections and can increase the susceptibility of the bladder to infection by different UPEC strains. Interestingly, previous history results in a colonization resistance upon new exposure. For example, a remodeled bladder epithelium is recalcitrant to IBC formation. However, increased expression of COX-2 breaks this resistance and allows unchecked bacterial replication and recurrent infection in the bladder. This is in stark contrast to the pathogenic cascade in a naïve bladder that requires IBC formation (O'Brien et al., 2016). Thus, a remodeled bladder epithelium of women who have experienced frequent rUTIs may be predisposing to recurrent infections. In addition, a separate set of factors in UPEC may increase bacterial fitness in such a remodeled habitat. These data underscore the fact that the categorization of UPEC into one group, without regard to differences in requirements to colonize distinct host environments or differing clinical definitions, has increased the difficulty in identifying a clear genetic definition of UPEC.

1.3.2 The gut as a reservoir of UPEC

The fecal-perineal-urethral model of ascending UTI posits that UPEC maintain an established reservoir in the gut before transmission to the bladder (Yamamoto et al., 1997). In healthy adults, *E. coli* account for ~0.8% of the total microbiota and are the primary facultative anaerobe in the gut microbial community (Consortium, 2012). While the majority of adults carry *E. coli* asymptomatically in their gut, blooms of *E. coli* are associated with a number of intestinal diseases including diarrheal diseases, inflammatory bowel diseases (IBD) such as Crohn's disease, and colorectal cancer (Arthur et al., 2012; Kotlowski et al., 2007). One prominent host factor that likely influences the *E. coli* population in the gut is inflammation, which can be triggered in numerous ways including infection by a pathogen, an imbalance of the existing microbiota, or exposure to antibiotics during clinical intervention (Stecher et al., 2010; Winter et

al., 2013). Interestingly, biopsy specimens from patients with Crohn's disease and ulcerative colitis, two IBD syndromes, revealed that these patients have a 3-4 log increase in the levels of Enterobacteriaceae in their intestines over healthy controls, with a significant increase in the abundance of *E. coli* from the B2 and D phylogenetic groups (Kotlowski et al., 2007). This has also been observed in conventionally raised mice treated with the antibiotic streptomycin or dextran sulfate sodium (DSS), a chemical used to superficially stimulate Crohn's disease in murine animals (Lupp et al., 2007). Of interest, several clinical studies have found that patients with intestinal inflammatory diseases, like IBD and AIDS, have an increased incidence of rUTI (Evans et al., 1995; Kyle, 1980). Importantly, studies have shown that at the time of UTI, the dominant *E. coli* strain in the urine and feces of an infected woman is the same (Chen et al., 2013; Moreno et al., 2008). Strikingly, in instances of different-strain rUTI (described above), the new strain infecting the bladder is also found to be the dominant strain in the gut as well (Chen et al., 2013). Together, these data suggest that intestinal inflammation may alter conditions in the gut in a way that enhances UPEC colonization, ultimately increasing the likelihood of a downstream UTI, particularly in individuals that are susceptible to chronic/recurrent cystitis.

1.3.3. Putative urovirulence factors (PUFs) in UPEC

While there is no strict genetic definition of UPEC, previous research on clinical UPEC isolates has identified a list of putative urovirulence factors (PUFs) that are enriched in UPEC strains relative to commensal *E. coli* strains through the use of a polymerase chain reaction (PCR) typing technique, which measures the carriage of specific genes of interest (Johnson et al., 2001; 2005). These PUFs include genes for capsule production, iron acquisition systems, toxins, and CUP pili along with other miscellaneous functions (Johnson et al., 2001; 2005). Clinical

UPEC strains with increased numbers of these PUFs were more likely to be associated with same-strain recurrent UTI episodes than strains with fewer virulence factors (Luo et al., 2012) and clade B2 UPEC were more likely than other UPEC to carry many PUF genes (Luo et al., 2012; Picard et al., 1999). Previous research using model B2 UPEC strains has also shown that genetic deletion of some of the PUF genes attenuates colonization of the mouse bladder (Flores-Mireles et al., 2015). Thus, these data show that, within an individual strain model UPEC strain, PUFs can enhance fitness; however, carriage of PUF genes is variable and UPEC strains encoding very few PUFs are capable of causing robust UTIs. Therefore, putative urovirulence factors are not essential for an *E. coli* strain to cause UTI in women. Interestingly, several PUFs have also been implicated in UPEC persistence in the gut (Nowrouzian et al., 2005; 2006). Yet, despite these correlations and widespread acceptance that the gut is a reservoir for UPEC (Chen et al., 2013; Yamamoto et al., 1997), we know very little about the molecular mechanisms that promote the establishment and maintenance of UPEC in this habitat. Taken together, these data suggest that putative urovirulence factors may enhance the bladder colonization of some *E. coli*; however, overall, they are not necessary for bladder colonization in the diverse host bladder habitats by all UPEC strains. Further, PUFs may enhance fitness of UPEC in colonizing other body sites, such as the gut, which may enhance their ability to cause subsequent UTIs in the future.

1.4 Summary

In this chapter, I have described our current understanding of UPEC pathogenesis, which has been elucidated almost exclusively by model B2 UPEC strains in mouse models of cystitis. I have also detailed evidence that implicates the gut as an important, yet understudied, reservoir

that influences UTI recurrence. Finally, I have described the genetic and phylogenetic diversity seen in clinical UPEC strains as well as the failing of current molecular epidemiology studies to fully articulate a clear, genetic definition of UPEC. Thus, towards the goal of filling these gaps in knowledge regarding the genetic basis for uropathogenesis in *E. coli*, I employed a research strategy that integrated large-scale multi-omics and evolutionary analyses of diverse clinical UPEC strains with *in vitro* phenotypic analysis and *in vivo* animal models of both bladder and gut colonization. This strategy has several advantages, including: (1) the results from multi-omics analyses, such as comparative genomics and comparative transcriptomics, allow for a global analysis of bacterial features at a higher resolution (i.e., down to the level of single nucleotide polymorphisms in protein coding genes) than single gene analyses (e.g., PCR typing used to measure PUF carriage that result in binary presence/absence data), (2) the ability to examine the totality of the UPEC lifestyle, including the gut reservoir, which avoids the assumption that all features that enable uropathogenicity in clinical UPEC strains must result from increased ability to colonize the bladder, (3) the integration mouse models of cystitis and *in vitro* measures of virulence with bioinformatic analyses, which enables the direct interrogation of hypotheses generated from large datasets in a setting that better represents the biological reality of patients in the clinic instead of a simplified "one-bug, one-mouse" analysis.

Towards the goal of understanding conserved features that enable UPEC pathogenesis, I will describe three projects that I have completed in my graduate career. In Chapter 2, I will show genetically diverse *E. coli* strains are capable of robust bladder colonization in many different mouse backgrounds and that transcriptional control of core bacterial behaviors is a better predictor of bladder colonization than gene carriage alone. In Chapter 3, I will describe the evolutionary conservation of the structure and function of the type 1 pilus rod as a molecular

spring and describe the fitness effects of mutations in the pilus on the ability of UPEC to colonize the bladder and gut. In Chapter 4, I will characterize my work describing the evolution of CUP pili and their importance in gut colonization by UPEC. Together, my studies have worked towards expanding our understanding of UPEC virulence and evolution. Through this understanding, we are now able to identify and deploy new therapeutic options that have a much greater chance of success in combating a genetically diverse pathogen by targeting the common, evolutionary conserved features that enable UPEC virulence and persistence.

1.6 References

Anderson, G.G., Palermo, J.J., Schilling, J.D., Roth, R., Heuser, J., and Hultgren, S.J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. Science *301*, 105–107.

Arthur, J.C., Perez-Chanona, E., Mühlbauer, M., Tomkovich, S., Uronis, J.M., Fan, T.-J., Campbell, B.J., Abujamel, T., Dogan, B., Rogers, A.B., et al. (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. Science *338*, 120–123.

Barber, A.E., Norton, J.P., Wiles, T.J., and Mulvey, M.A. (2016). Strengths and Limitations of Model Systems for the Study of Urinary Tract Infections and Related Pathologies. Microbiol. Mol. Biol. Rev. *80*, 351–367.

Brzuszkiewicz, E., Brüggemann, H., Liesegang, H., Emmerth, M., Olschläger, T., Nagy, G., Albermann, K., Wagner, C., Buchrieser, C., Emődy, L., et al. (2006). How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc. Natl. Acad. Sci. U.S.a. *103*, 12879–12884.

Carey, A.J., Tan, C.K., Ipe, D.S., Sullivan, M.J., Cripps, A.W., Schembri, M.A., and Ulett, G.C. (2016). Urinary tract infection of mice to model human disease: Practicalities, implications and limitations. Crit. Rev. Microbiol. *42*, 780–799.

Chen, S.L., Hung, C.S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., and Ozersky, P. (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U.S.a. *103*, 5977–5982.

Chen, S.L., Wu, M., Henderson, J.P., Hooton, T.M., Hibbing, M.E., Hultgren, S.J., and Gordon, J.I. (2013). Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Science Translational Medicine *5*, 184ra60.

Consortium, T.H.M.P. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., and Finlay, B.B. (2013). Recent advances in understanding enteric pathogenic *Escherichia coli*. Clinical Microbiology Reviews *26*, 822–880.

Ejrnaes, K. (2011). Bacterial characteristics of importance for recurrent urinary tract infections caused by *Escherichia coli*. Dan Med Bull *58*, B4187.

Evans, J.K., McOwan, A., Hillman, R.J., and Forster, G.E. (1995). Incidence of symptomatic urinary tract infections in HIV seropositive patients and the use of cotrimoxazole as prophylaxis against Pneumocystis carinii pneumonia. Genitourin Med *71*, 120–122.

Ferry, S.A., Holm, S.E., Stenlund, H., Lundholm, R., and Monsen, T.J. (2004). The natural course of uncomplicated lower urinary tract infection in women illustrated by a randomized placebo controlled study. Scand. J. Infect. Dis. *36*, 296–301.

Flores-Mireles, A.L., Walker, J.N., Caparon, M., and Hultgren, S.J. (2015). Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nature Reviews Microbiology *13*, 269–284.

Foxman, B., Gillespie, B., Koopman, J., Zhang, L., Palin, K., Tallman, P., Marsh, J.V., Spear, S., Sobel, J.D., Marty, M.J., et al. (2000). Risk factors for second urinary tract infection among college women. American Journal of Epidemiology *151*, 1194–1205.

Foxman, B., Zhang, L., Tallman, P., Palin, K., Rode, C., Bloch, C., Gillespie, B., and Marrs, C.F. (1995). Virulence characteristics of *Escherichia coli* causing first urinary tract infection predict risk of second infection. J. Infect. Dis. *172*, 1536–1541.

Foxman, B. (2010). The epidemiology of urinary tract infection. Nature Publishing Group *7*, 653–660.

Foxman, B. (2014). Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect. Dis. Clin. North Am. *28*, 1–13.

Hannan, T.J., Mysorekar, I.U., Hung, C.S., Isaacson-Schmid, M.L., and Hultgren, S.J. (2010). Early Severe Inflammatory Responses to Uropathogenic *E. coli* Predispose to Chronic and Recurrent Urinary Tract Infection. PLoS Pathog *6*, e1001042.

Hooton, T.M., Scholes, D., Hughes, J.P., Winter, C., Roberts, P.L., Stapleton, A.E., Stergachis, A., and Stamm, W.E. (1996). A prospective study of risk factors for symptomatic urinary tract infection in young women. N. Engl. J. Med. *335*, 468–474.

Hooton, T.M. (2012). Clinical practice. Uncomplicated urinary tract infection. N. Engl. J. Med. *366*, 1028–1037.

Hung, C.-S., Dodson, K.W., and Hultgren, S.J. (2009). A murine model of urinary tract infection. Nat Protoc *4*, 1230–1243.

Ikaheimo, R., Siitonen, A., Heiskanen, T., Kärkkäinen, U., Kuosmanen, P., Lipponen, P., and Mäkelä, P.H. (1996). Recurrence of urinary tract infection in a primary care setting: analysis of a 1-year follow-up of 179 women. Clin. Infect. Dis. *22*, 91–99.

Johnson, J.R., O'Bryan, T.T., Delavari, P., Kuskowski, M., Stapleton, A., Carlino, U., and Russo, T.A. (2001). Clonal Relationships and Extended Virulence Genotypes among *Escherichia coli*Isolates from Women with a First or Recurrent Episode of Cystitis. J. Infect. Dis. *183*, 1508–1517.

Johnson, J.R., Owens, K., Gajewski, A., and Kuskowski, M.A. (2005). Bacterial characteristics in relation to clinical source of *Escherichia coli* isolates from women with acute cystitis or pyelonephritis and uninfected women. Journal of Clinical Microbiology *43*, 6064–6072.

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J., and Hultgren, S.J. (2004). Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. Proc. Natl. Acad. Sci. U.S.a. *101*, 1333–1338.

Justice, S.S., Hunstad, D.A., Seed, P.C., and Hultgren, S.J. (2006). Filamentation by *Escherichia coli* subverts innate defenses during urinary tract infection. Proc. Natl. Acad. Sci. U.S.a. *103*, 19884–19889.

Kärkkäinen, U., Ikaheimo, R., Katila, M.-L., and Siitonen, A. (2000). Recurrence of Urinary Tract Infections in Adult Patients with Community-acquired Pyelonephritis Caused by *E. coli*: A 1-year Follow-up. Scand. J. Infect. Dis. *32*, 495–499.

Kotlowski, R., Bernstein, C.N., Sepehri, S., and Krause, D.O. (2007). High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. Gut *56*, 669–675.

Kõljalg, S., Truusalu, K., Vainumäe, I., Stsepetova, J., Sepp, E., and Mikelsaar, M. (2009). Persistence of *Escherichia coli* Clones and Phenotypic and Genotypic Antibiotic Resistance in Recurrent Urinary Tract Infections in Childhood. Journal of Clinical ….

Kunin, C.M. (1970). A ten-year study of bacteriuria in schoolgirls: final report of bacteriologic, urologic, and epidemiologic findings. J. Infect. Dis. *122*, 382–393.

Kyle, J. (1980). Urinary complications of Crohn's disease. World J. Surg. *4*, 153–159.

Luo, Y., Ma, Y., Zhao, Q., Wang, L., Guo, L., Ye, L., Zhang, Y., and Yang, J. (2012). Similarity and Divergence of Phylogenies, Antimicrobial Susceptibilities, and Virulence Factor Profiles of *Escherichia coli* Isolates Causing Recurrent Urinary Tract Infections That Persist or Result from Reinfection. Journal of Clinical Microbiology *50*, 4002–4007.

Lupp, C., Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L., Gaynor, E.C., and Finlay, B.B. (2007). Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. Cell Host and Microbe *2*, 119–129.

Mabeck, C.E. (1972). Treatment of uncomplicated urinary tract infection in non-pregnant women. Postgrad Med J *48*, 69–75.

Marrs, C.F., Zhang, L., and Foxman, B. (2005). *Escherichia coli* mediated urinary tract infections: are there distinct uropathogenic *E. coli* (UPEC) pathotypes? FEMS Microbiol. Lett. *252*, 183–190.

Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S., and Hultgren, S.J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. Embo J. *19*, 2803–2812.

Moreno, E., Andreu, A., Pigrau, C., Kuskowski, M.A., Johnson, J.R., and Prats, G. (2008). 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.

Mulvey, M.A., Lopez-Boado, Y.S., Wilson, C.L., Roth, R., Parks, W.C., Heuser, J., and Hultgren, S.J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. Science *282*, 1494–1497.

Mysorekar, I.U., and Hultgren, S.J. (2006). Mechanisms of uropathogenic *Escherichia coli* persistence and eradication from the urinary tract. Proc. Natl. Acad. Sci. U.S.a. *103*, 14170–14175.

Nowrouzian, F.L., Wold, A.E., and Adlerberth, I. (2005). *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J. Infect. Dis. *191*, 1078–1083.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006). Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect. *8*, 834–840.

O'Brien, V.P., Hannan, T.J., Yu, L., Livny, J., Roberson, E.D.O., Schwartz, D.J., Souza, S., Mendelsohn, C.L., Colonna, M., Lewis, A.L., et al. (2016). A mucosal imprint left by prior *Escherichia coli* bladder infection sensitizes to recurrent disease. Nat Microbiol *2*, 16196.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infection and Immunity *67*, 546–553.

Scholes, D., Hooton, T.M., Roberts, P.L., Stapleton, A.E., Gupta, K., and Stamm, W.E. (2000). Risk Factors for Recurrent Urinary Tract Infection in Young Women. J. Infect. Dis. *182*, 1177–1182.

Song, J., Bishop, B.L., Li, G., Duncan, M.J., and Abraham, S.N. (2007). TLR4-initiated and cAMP-mediated abrogation of bacterial invasion of the bladder. Cell Host and Microbe *1*, 287–298.

Stecher, B., Chaffron, S., Käppeli, R., Hapfelmeier, S., Freedrich, S., Weber, T.C., Kirundi, J., Suar, M., McCoy, K.D., Mering, von, C., et al. (2010). Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. PLoS Pathog *6*, e1000711.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. Nature Reviews Microbiology *8*, 207–217.

Thumbikat, P., Berry, R.E., Schaeffer, A.J., and Klumpp, D.J. (2009a). Differentiation-induced uroplakin III expression promotes urothelial cell death in response to uropathogenic *E. coli*. Microbes Infect. *11*, 57–65.

Thumbikat, P., Berry, R.E., Zhou, G., Billips, B.K., Yaggie, R.E., Zaichuk, T., Sun, T.-T., Schaeffer, A.J., and Klumpp, D.J. (2009b). Bacteria-induced uroplakin signaling mediates bladder response to infection. PLoS Pathog *5*, e1000415.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.a. *99*, 17020–17024.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., Laughlin, R.C., Gomez, G., Wu, J., Lawhon, S.D., et al. (2013). Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. Science *339*, 708–711.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2005). Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization. Infection and Immunity *73*, 7657–7668.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli*depends on type 1 pili. Cell. Microbiol. *9*, 2230–2241.

Yamamoto, S., Tsukamoto, T., Terai, A., Kurazono, H., Takeda, Y., and Yoshida, O. (1997). Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. J. Urol. *157*, 1127–1129.

Zowawi, H.M., Harris, P.N.A., Roberts, M.J., Tambyah, P.A., Schembri, M.A., Pezzani, M.D., Williamson, D.A., and Paterson, D.L. (2015). The emerging threat of multidrug-resistant Gram-negative bacteria in urology. Nature Publishing Group *12*, 570–584.

Chapter 2:

Interplay between dynamic bacterial virulence phenotypes of *E. coli* and host susceptibility determines UTI risk

By

Henry L. Schreiber IV, Matt S. Conover, Wen-Chi Chou, Michael E. Hibbing,

Abigail L. Manson, Karen W. Dodson, Thomas J. Hannan, Pacita L. Roberts, Anne E. Stapleton,

Thomas M. Hooton, Jonthan. Livny, Ashlee M. Earl, Scott J. Hultgren

Science Translational Medicine

2017 March 22. 9(382): eaaf1283.

doi: 10.1126/scitranslmed.aaf1283. PMID 28330863

Format Adapted for Dissertation

2.1 Abstract

Urinary tract infections (UTIs) are caused by uropathogenic *Escherichia coli* (UPEC) strains. In contrast to many enteric *E. coli* pathogroups, no genetic signature has been identified for UPEC strains. We conducted a high-resolution comparative genomic study using *E. coli* isolates collected from the urine of women suffering from frequent recurrent UTIs. These isolates were genetically diverse and varied in urovirulence, or the ability to infect the bladder of a mouse model of cystitis. Importantly, we found no set of genes, including previously defined putative urovirulence factors (PUFs), that were predictive of urovirulence. In addition, in some patients, the *E. coli* strain causing a recurrent UTI had fewer PUFs than the supplanted strain. In competitive experimental infections in mice, the supplanting strain was more efficient at colonizing the mouse bladder than the supplanted strain. Despite the lack of a clear genomic signature for urovirulence, comparative transcriptomic and phenotypic analyses revealed that the expression of key conserved functions during culture, such as motility and sugar metabolism, could be used to predict subsequent mouse bladder colonization. Taken together, our findings suggest that UTI risk and outcome may be determined by complex interactions between host susceptibility and the urovirulence potential of diverse bacterial strains.

2.2 Introduction

Urinary tract infections (UTIs) are a health and financial burden, particularly affecting sexually active, premenopausal women (Foxman, 2002). Approximately 50% of women will suffer a UTI in their lifetime with 10.5 million cases of UTIs occurring in 2007 in the United States alone (Foxman, 2014). In placebo-controlled studies of UTI outcome, most women maintain bladder infections for weeks after an acute cystitis episode if not treated with antibiotics, whereas other women clear the infection without antibiotic treatment (Ferry et al., 2004; Mabeck,

1972). These two outcomes of infection, either self-limiting or chronic cystitis, have been modeled in mice (Hannan et al., 2010; 2014; O'Brien et al., 2016). Effective treatment of UTIs is becoming more challenging as antibiotic resistance rates are increasing (Gupta et al., 2011). In addition, even though antibiotic therapy effectively resolves the majority of UTIs, approximately 20-30% of women will experience a recurrent UTI (rUTI) within six months of initial infection (Foxman et al., 2000; Gupta et al., 2011). This highlights the need to better understand the pathogenic mechanisms that facilitate acute bacterial colonization of the urinary tract and determine UTI outcome.

The most common cause of community-acquired UTIs is uropathogenic *E. coli* (UPEC) (Foxman, 2014). *E. coli* strains are diverse and have been characterized into pathotypes based on their ability to cause particular diseases (Kaper et al., 2004; Russo and Johnson, 2000). Pathotypes include intestinal pathogens such as Shiga toxin-producing *E. coli* as well as extra-intestinal pathogenic *E. coli*, strains such as UPEC. *E. coli* are also sub-divided into phylogenetic clades (e.g., A, B1, B2, D, E) based on their genetic similarity (Rasko et al., 2008; Touchon et al., 2009). In the U.S. and Europe, the majority of UPEC are from the B2 clade though members of clades A, B1 and D cause 25-50% of cystitis (Ejrnaes et al., 2011; Piatti et al., 2008; Rijavec et al., 2006; Skjøt-Rasmussen et al., 2011; Starcic Erjavec et al., 2007). In East Asia, clade D strains predominate in community-acquired UTI followed by B2 strains (Luo et al., 2012; Wang et al., 2014). All commonly studied UPEC isolates, including UTI89, CFT073, NU14, and 536 (Brzuszkiewicz et al., 2006; Chen et al., 2006; Hultgren et al., 1986; Welch et al., 2002), are B2 members whereas non-B2 strains have not been extensively characterized. Thus, it is unknown whether non-B2 UPEC strains use the same virulence mechanisms as B2 UPEC strains.

Study of prototypical B2 UPEC isolates, such as the model UPEC strain UTI89, originally isolated from a woman with cystitis (Mulvey et al., 2001), has shown that the chaperone usher pathway (CUP) type 1 pilus is critical for bladder colonization in diverse mouse models (Schwartz et al., 2013). The *fim* operon, which encodes the proteins for assembly of type 1 pili, is part of the core *E. coli* genome, found in nearly every *E. coli* strain (Wurpel et al., 2013). Further, in research of the model strain UTI89, it has been shown that type 1 pili expression is tightly co-regulated with numerous factors including flagella, and S and P CUP pili (Greene et al., 2014). Type 1 pili are tipped by the adhesin FimH, which binds specifically to mannosylated uroplakins lining the superficial bladder epithelium (Wu et al., 1996) and mannosylated proteins expressed on underlying layers of epithelial cells (Eto et al., 2007), thus mediating critical steps in bladder colonization and UTI progression. Subsequent to FimH attachment, *E. coli* can invade the superficial umbrella bladder cells (Martinez et al., 2000) and escape into the cytoplasm where they rapidly replicate to form clonal, biofilm-like intracellular bacterial communities (IBCs) of ~10^4 colony forming units (CFU) while protected from the host immune response (Anderson et al., 2003; Justice et al., 2004). Upon IBC maturation, UTI89 transform into long filaments and burst out of the superficial umbrella cells of the bladder into the bladder lumen where they invade other epithelial cells and form new IBCs (Justice et al., 2004). The majority of clinical UPEC isolates that have been tested are able to form IBCs in mice, and each of six different inbred mouse strains tested support IBC formation (Garofalo et al., 2007). Importantly, after the discovery of IBCs in mouse UTI models, IBCs and filamentous bacteria were documented in human urine collected from clinical UTI cases Robi(Robino et al., 2013; Rosen et al., 2007).

Despite decades of research using model UPEC strains, universal bacterial features that enable *E. coli* uropathogenesis remain largely undefined. Carriage of putative urovirulence

factors (PUFs) is thought to enhance *E. coli* uropathogenicity and is used to measure and categorize clinical UPEC strains isolated from different patient populations (Johnson et al., 2001b; 2015; Luo et al., 2012; Starcic Erjavec et al., 2007). PUFs were identified through their enrichment in UTI isolates when compared to non-UTI associated *E. coli* (Johnson and Stell, 2000; Johnson et al., 2001a; 2001b; 2015). While numerous sets of PUFs have been used to assess urovirulence (Johnson and Stell, 2000; Johnson et al., 2001a; 2001b; 2015; Luo et al., 2012; Starcic Erjavec et al., 2007) and strains with more PUFs are considered more uropathogenic than strains with fewer PUFs, a direct role in pathogenicity remains to be defined for the majority of PUFs.

Here, we examined a set of 43 human urine-associated *E. coli* (UAEC) strains isolated at symptomatic and asymptomatic time points from the urine of 14 women with frequent rUTIs. We tested whether UAEC isolates shared common genes that classified strains with increased virulence from those with lesser virulence in defined experimental mouse models. Surprisingly, we observed no correlation between gene content, including PUFs, and the ability to cause cystitis in C3H/HeN mouse models. However, we found that environmentally-responsive phenotypes and specific transcriptional responses to *in vitro* conditions could discriminate whether strains would be more or less efficient at bladder colonization when tested in the C3H/HeN acute cystitis mouse model. We also determined that some UAEC were robust colonizers in multiple mouse models whereas others were only efficient at colonizing the bladder in some models.

2.3 Results

2.3.1 Phylogenetic origin and relatedness of human urine-associated *E. coli* strains

A total of 43 *E. coli* isolates was collected from 14 women with frequent rUTI (**Figure S1, Table S1, Table S2**). Since all of these strains were isolated from the urine of women, but not necessarily collected during diagnosed UTIs, we refer to these strains as urine-associated *E. coli* (UAEC) rather than UPEC. Whole genome sequences of all 43 UAEC were determined, assembled, annotated and combined with genomic data from 46 previously sequenced representative *E. coli* from various clades and pathotypes to evaluate aspects of their relatedness and gene content (**Table S3**). We found that 60-75% of each strain's genome consisted of core genes, defined as genes present in all strains being analyzed, while the variable genome comprised the remaining 25-40% (**Figure S2A**). We constructed a maximum-likelihood phylogenetic tree using the 2,746 single-copy genes present in the core genome of all the *E. coli* isolates (**Figure 1**). Consistent with previous studies of UPEC (Ejrnaes et al., 2011; Piatti et al., 2008; Rijavec et al., 2006; Skjøt-Rasmussen et al., 2011; Starcic Erjavec et al., 2007), more than two-thirds (67.4%, 29 of 43) of our UAEC strains fell within the B2 clade. The remaining UAEC fell within three other *E. coli* clades (4 in clade A, 8 in clade B1 and 2 in clade D).

In some patients, UAEC isolated at consecutive time points were tightly grouped within the phylogenetic tree indicating same-strain rUTI. In contrast, UAEC collected at consecutive time points in other patients were not closely related, which is consistent with different-strain rUTI. As has been done in previous studies (Sahl et al., 2015), cluster analysis of pairwise SNP distances between all UAEC isolates revealed 13 different clonal groups comprised of two to four isolates each (**Figure S2B** and **Table S2**). Eight UAEC did not group with any other isolate, and were each assigned to their own core clonal group. Based on clonal group membership, we identified 12 same-strain recurrences and 6 different-strain recurrences from a total of 18 rUTI events (**Table S2**). Eleven of 12 same-strain rUTIs were caused by B2 strains, which was a

statistically significant enrichment relative to the phylogenetic distribution of the UAEC tested in our cohort (Hypergeometric test, $p<0.05$). For further analyses, we then selected a representative UAEC strain from each of the clonal groups, including core clonal groups with a single member, resulting in a representative collection of 21 UAEC isolates that encompassed the genetic, phylogenetic, and phenotypic diversity of our strain set, including 11 clade B2, 2 clade A, 6 clade B1 and 2 clade D strains (**Table S2**).

2.3.2 Carriage of PUFs correlates with B2 clade membership in human urine and non-urine isolates of *E. coli.*

As done previously (Johnson and Stell, 2000; Johnson et al., 2001b; 2015), we calculated a PUF score for each representative UAEC and non-UAEC control strain using an alignment-based analysis of 31 previously defined PUFs (Johnson and Stell, 2000; Johnson et al., 2001b; 2015; Luo et al., 2012)(**Table S4**). Carriage of PUFs varied considerably among strains, ranging from 2 to 24 PUFs per UAEC genome (median 13; the model B2 UPEC strain, UTI89, encodes 17 PUFs) to 2 to 19 PUFs per non-UAEC genome (median 5) (**Figure 2**). Out of the 31 PUFs, 28 were found in at least one UAEC genome. We found that B2 *E. coli* strains carried more PUFs than did non-B2 *E. coli* strains whether or not they were associated with urinary disease in humans (Mann-Whitney U test, $p<0.001$) (**Figure 2A**). Further, using unsupervised hierarchical clustering of PUF carriage, we found that strains generally clustered based on B2 clade membership, rather than on their association with urinary disease (**Figure 2B**), indicating that PUF carriage was biased by phylogeny and was not specific to uropathogens.

2.3.3 Bladder colonization in C3H/HeN mice by UAEC is variable and not restricted to B2 strains

Although each of our UAEC strains was recovered from the urine of a patient and thus able to colonize a human bladder, differences in patient genetics, behavior, and infection history

can influence susceptibility to UTI (Foxman et al., 2000; Hawn et al., 2009; Scholes et al., 2000; Stapleton et al., 1995). To control for host genetics and environmental factors that could confound comparisons of colonization efficiency among the UAEC strains, we characterized the ability of each of 21 representative strains to colonize the genetically homogenous C3H/HeN mouse model of cystitis. Each UAEC strain and a model UPEC strain, UTI89, was inoculated into the bladder of at least 5 C3H/HeN mice. Bacterial burdens, measured in colony forming units (CFU), were determined in both the bladder and kidneys at 24 hours post-infection (hpi) (**Figure 3A, Figure S3**). We observed a range of bladder bacterial burdens (200 to 10^8 CFU/bladder) at 24 hpi with three colonizer groups emerging among the tested UAEC strains. Twelve of 21 strains were "robust colonizers" where UAEC titers were greater than 10^4 CFU in every infected mouse bladder; 4 of 21 strains were "deficient colonizers" where UAEC titers were less than 10^4 in every infected bladder, and; 5 of 21 were "variable colonizers" where UAEC titers ranged both above and below 10^4 CFU/bladder of infected mice (**Figure 3A**). Kidney colonization generally correlated with bladder colonization (**Figure S3**). Notably, both B2 and non-B2 strains were among the robust colonizers, capable of colonizing the mouse bladder at rates similar to UTI89, indicating that efficient bladder colonization of C3H/HeN mice was achievable by strains outside of the B2 clade. Importantly, whereas C3H/HeN bladder colonization varied among UAEC at 24 hpi, the titers of these isolates were markedly higher than those of gut-associated *E. coli* strains collected from adult humans that were not associated with urinary tract disease. These strains included model *E. coli* strains and strains from the feces of healthy adults (Gordon et al., 2005; Snyder et al., 2012) and were comparable to our UAEC collection in the distribution of B2 versus non-B2 strains (**Table S5**). We found that only 1 of 9 gut-associated *E. coli*, a clade A strain, was capable of robustly colonizing the bladders at 24 hpi,

while 5 of 9 strains were deficient and the remaining 4 of 9 were variable (**Figure S4a**). Thus, a significantly higher proportion of UAEC strains robustly colonized the C3H/HeN mouse model than gut-associated *E. coli* at 24 hpi (Fisher's Exact test, $p<0.05$) (**Table S5** and **Figure S4**).

Given that early bladder colonization is enhanced by the formation of IBCs characterized using model B2 UPEC strains (Anderson et al., 2003; Justice et al., 2004; Mulvey et al., 1998; Schwartz et al., 2011), we sought to determine if diverse B2 and non-B2 UAEC strains from our collection followed the same pathogenic cascade as model B2 strains. As for the UPEC strain UTI89 (Schwartz et al., 2011), we found that 5 of 6 phylogenetically diverse UAEC were competent for IBC formation (**Figure S5** and **Table S6**) in C3H/HeN mice at 6 hpi, suggesting that both B2 and non-B2 UAEC strains exhibited pathogenic mechanisms similar to that of the model B2 UPEC strain. Notably, the only deficient colonizer strain tested amongst the group clade A strain 11.1a was also the strain that did not form IBCs at this single time point supporting previous reports that IBC formation is linked to bladder colonization (Schwartz et al., 2011).

2.3.4 Carriage of PUFs is not required for bladder colonization in C3H/HeN mice, but may provide a competitive advantage in chronic cystitis.

Although UAEC strains with both high and low PUF scores caused rUTI among the women in our cohort, we sought to determine whether PUF content could explain the observed differences in bladder colonization in the C3H/HeN mouse model (**Figure 3A**). We observed no significant differences in the carriage of PUFs when comparing UAEC strains that were robust (range 2 to 20, median 16), variable (range 2 to 24, median 7) or deficient (range 4 to 12, median 3.5) colonizers of the C3H/HeN mice ($p>0.05$, Kruskal-Wallis test and Mann-Whitney U tests) (**Figure 3B**). Further, we found no significant correlation between PUF carriage and bladder bacterial burden in C3H/HeN mice (Spearman's rank correlation, $p>0.05$) (**Figure 3C**).

In mice, chronic cystitis is defined as the development of persistent high titer (>10^4 CFU/ml) bacteriuria, high titer bladder bacterial burdens and chronic inflammation with persistent lymphoid aggregates at sacrifice >4 weeks post-infection (wpi). Chronic cystitis and bladder inflammation result in large-scale remodeling of the bladder mucosa, which can last for weeks/months after antibiotic treatment to clear the infection (Ferry et al., 2004; Mabeck, 1972). This remodeling is evident in changes to the bladder itself and in host susceptibility to recurrent infection, exemplified by: (i) an altered transcriptional profile; (ii) an altered urothelial membrane proteome; (iii) defects in superficial cell maturation; (iv) changes to bacterial occupation of different habitats and; (v) differential host responses upon subsequent bacterial exposure (O'Brien et al., 2016). To determine the effects of PUF carriage on the ability of UAEC to cause chronic cystitis, we examined three non-B2 strains: 41.4p (clade B1, PUF score=7), 9.1a (clade D, PUF score=13), and 9.2p (clade B1, PUF score=3) and three B2 UAEC strains: 20.1a (PUF score=16), 35.1a (PUF score=16), and 41.1a (PUF score=20) for their ability to cause persistent, high-titer bacteriuria (>10^4 CFU/mL urine), high-titer bladder colonization (>10^4 CFU/bladder) and chronic inflammation at 28 days post-infection in juvenile C3H/HeN mice. We found that all of the tested UAEC strains were capable of causing chronic cystitis at varying frequencies, including the non-B2 UAEC strains (range 20-90%) (**Figure 4**). These rates were comparable to the rates of chronic cystitis caused by B2 UAEC strains and UTI89 (range 10-60%)(Hannan et al., 2010). Taken together, our results indicate that several non-B2 strains carrying few PUFs were capable of causing both acute and chronic UTI in C3H/HeN mice in addition to causing clinical infection in women. While the carriage of many PUFs was not essential for non-B2 UAEC strains to cause chronic cystitis in mice, B2 UAEC strains that carried many PUF genes were enriched in isolates causing same-strain rUTI (11 of 12 same-

strain UTIs were caused by B2 UAEC). Indeed, we found that specific deletion from UTI89 of a large pathogenicity-associated island, PAI II_{UTI89}, carrying 124 genes including 4 PUFs (*papgIII*, *hek*, *cnf1* and *hlyA*) did not result in a competitive defect against the WT strain during acute time points (1-3 dpi) in urine; however, the mutant strain was severely outcompeted in the urine during chronic time points (≥7 dpi)(**Figure S6a**), and in the bladder and kidney tissue at sacrifice 28 dpi (**Figure S6b**).

2.3.5 Later time point rUTI isolates outcompete those collected at enrollment regardless of PUF carriage

Previous research has shown that different-strain rUTIs can occur when the recurrent strain outcompetes the initial UTI strain in the gastrointestinal tract and bladder (Chen et al., 2013). In two patients in our cohort, patient 9 and patient 41, a UAEC strain with a higher PUF score (strain 9.1a, PUF score=13; strain 41.1a, PUF score=20) was supplanted by a UAEC strain with a lower PUF score (strain 9.2p, PUF score=3; strain 41.4p, PUF score=7). We hypothesized that the supplanting strains would be fitter than the initial UTI strains collected from the same patient in colonizing the murine bladder. Thus, in two separate experiments, we inoculated equal numbers of the initial and supplanting strains from patients 9 and 14 into the bladders of C3H/HeN mice, respectively, and the relative bacterial titers were determined from urine at 24hpi using strain-specific antibiotic resistance markers (**Figure 5**). In both experiments, we found that the supplanting strains outcompeted the initial UTI strains at 24 hpi despite the fact that strains 41.4p and 9.2p carried considerably fewer PUF genes. Taken together, these data suggest that bladder colonization was not limited to B2 UAEC strains or strains with high carriage of PUFs.

2.3.6 Variations in UAEC gene expression under defined conditions predicts the outcome of acute cystitis in mice

It is possible that UAEC strains share some genomic feature that is not currently identified as a PUF. Thus, we also performed an in-depth, unbiased comparative genomic analysis between robust and deficient colonizer strains of UAEC to find any possible cryptic PUF genes or genomic signature that would clearly delineate these two groups. Our analysis revealed that no orthologous genes were exclusive to either robust or deficient colonizers and that there was no significant enrichment of specific functional domains or enzymatic pathways in either colonizer group after correcting for variance between strains. Using a relaxed definition of enrichment (i.e., present in >80% of one group and <20% of the other group), we found that two orthogroups were enriched in the deficient colonizer UAEC relative to robust colonizer strains, namely the *brnTA* genes encoding a type II toxin-antitoxin (Heaton et al., 2012)(**Table S7**). The *brnTA* genes, whose function in pathogenesis and regulation in *E. coli* are unknown, were identified in all four deficient colonizer strains; however, they were not specific to deficient colonizers, as they were also found in 2 of the 12 robust colonizers (12.1a and 26.1a). Further, the number of hits identified here fell below our threshold for Type I errors, which was determined by extensive permutations of the enrichment analysis. Thus, we find that the enrichment of these two genes in deficient colonizers is likely due to random chance.

Whereas differences in gene carriage could not clearly discriminate between these two groups of UAEC strains, we hypothesized that differences in how core bacterial functions are regulated, such as pili production, might be able to explain the variation that we observed in mouse bladder colonization and UTI outcome. We found significant variability in pili function amongst our UAEC strains with mannose-sensitive hemagglutination titers, denoting the expression of type 1 pili, ranged from zero to 2^{10} (**Figure S7a**) despite the presence of an intact *fim* operon in 20 of the 21 UAEC strains, as determined using custom BLAST and alignment-

based searches (see Supplemental Methods). Notably, the closely related B2 strains 20.1a, 21.1a and 35.1a had high mannose-resistant hemagluttination titers ($>2^4$), denoting expression of a non-type 1 pilus, under culture conditions that typically induce mannose-sensitive hemagglutination phenotypes in model B2 strains such as UTI89, including S pili (**Figure S7b**)(Chen et al., 2009; Parkkinen et al., 1986). Importantly, we found that robust colonizer strains had significantly higher hemagglutination titers (median $2^{7.25}$) than deficient colonizers (median $2^{2.25}$) (Mann-Whitney U test, $p< 0.01$) or variable colonizers (median 2^6) (Mann-Whitney U test, $p<0.01$) (**Figure S7c**), and that hemagglutination titer was significantly correlated to bladder bacterial burden in C3H/HeN mice at 24 hpi (Spearman's rank correlation, $p<0.001$)(**Figure S7d**). Taken together, these data suggest that the regulatory networks that drive the expression of CUP pili genes differ significantly between UAEC strains and that hemagglutination titer is correlated with the ability to cause acute cystitis in C3H/HeN mice.

To systematically compare core gene expression between phylogenetically and phenotypically diverse UAEC strains, we generated RNA-Seq data from triplicate cultures of a subset of 11 strains after growth in static LB broth (bladder infection inoculum). These strains included: (i) representative UAEC from four phylogenetic clades; (ii) robust, variable, and deficient colonizers; (iii) strains with disparate hemagglutination titers and; (iv) the model B2 strain, UTI89 (**Table S8**). Pairwise comparisons of transcript abundance for the 10 UAEC strains relative to the UTI89 strain revealed 753 genes (of 3,516 core genes or ~21.4%) that were differentially expressed in at least 1 of the 10 UAEC strains under these culture conditions ($P_{adj} < 0.05$, fold change > 3) (**Table S9**) with B2 strains having fewer differentially expressed genes relative to UTI89 than non-B2 strains (**Figure S8A**). However, the expression of a large number of core genes differed among the clinical UAEC strains, including both the B2 and non-B2

strains, even though they were grown under identical culture conditions. Differentially expressed genes were scattered throughout the genome, and many were clustered in operons encoding flagella, pilus, and chemotaxis machinery (**Figure S8B**). Notably, under this culture condition, the average expression of genes in the *fim* operon encoding type 1 pili was lower in deficient colonizers than robust colonizers and corresponded with the results of our hemagglutination assay. We also found that most PUFs present in each strain were expressed under these conditions, although at different levels when comparing across strains (**Table S10**).

To explore whether transcriptional profiles under inoculum growth conditions were able to predict bladder colonization efficiency, we performed principal component analysis (PCA) on the complete set of core genome transcriptional profiles relative to UTI89 (**Figure 6A**). We found a distinct pattern that separated the profiles of deficient and robust colonizers along the principal component 2 (PC2) axis, which explained 19.2% of the variance between the UAEC strains. Variable colonizers grouped with robust colonizers and away from deficient colonizers in this analysis. Further, PCA showed a separation between UAEC strains along PC1 (25.6% of variance) that correlated with clade membership (i.e. B2 vs. non-B2) suggesting that differences in gene expression associated with phylogeny are distinct from those that separate deficient and robust colonizers. Using a linear regression analysis, we systematically evaluated how expression of each gene (using normalized read counts per gene) correlated with bacterial burden at 24 hpi in the mouse bladder. After correcting for false discovery, we identified a total of 42 core genes whose expression in static LB broth correlated with subsequent bladder colonization efficiency in C3H/HeN mice at 24 hpi (**Figure S9** and **Table S11**). These included genes encoding sugar-transport proteins, such as the LamB maltoporin (**Figure 6B**), and other genes associated with maltose transport (Beg et al., 2007; Szmelcman and Hofnung, 1975), as well as

the Tar protein (**Figure 6C**), a chemotactic sensor of aspartate and maltose (Mowbray and Koshland, 1987; Springer et al., 1977). The majority of the linear regression curves had negative beta coefficients indicating that deficient colonizers had higher expression of these genes than robust colonizers. We grouped these 42 core genes into functional pathways (Huang et al., 2009a; 2009b), which showed a total of six functions that were enriched including motility, chemotaxis and carbohydrate transport (**Table S12**). These results indicate that under culture conditions used for growth of bacteria prior to infecting mice, core bacterial behaviors, such as swimming motility and nutrient transport, were regulated differently between robust and deficient colonizers in UAEC strains. Interestingly, despite differences in how genes mediating motility and chemotaxis were expressed, the motility of the clinical UAEC strains did not correlate with bladder colonization in C3H/HeN mice at 24 hpi (**Figure S10**), suggesting that transcriptional control of motility, rather than the *in vitro* phenotype itself, was a predictor of subsequent colonization efficiency.

2.3.7 UAEC strains vary in their ability to colonize the bladders of mice of different genetic backgrounds

Our data suggest that dynamic features (e.g., transcriptional control of core *E. coli* genes) rather than static characteristics (e.g., carriage of PUFs) could predict the relative efficiency of UAEC strain colonization of mouse bladder. Thus, differential regulation of transcriptional networks during *in vitro* growth in the colonization inoculum may have served to prime UAEC bladder colonization. To determine if this "priming" transcriptional response was universal in predicting colonization among different mouse strains, we inoculated a subset of our UAEC strains into C57BL/6 mice, which was a different host environment with a different immune response to infection. As with the C3H/HeN mouse model, the UAEC strains showed a variable ability to colonize the bladders of C57BL/6 mice, with bladder bacterial burdens ranging from

undetectable to $>10^7$ CFU/bladder at 24 hpi (**Figure 7**). Overall, we observed less variation in bladder colonization for each UAEC strain in C57BL/6 mice than seen in C3H/HeN mice e.g., the highly variable strains in C3H/HeN colonization experiments, 2.2r and 17.1a, showed a more narrow distribution of bladder burden at 24hpi in the C57BL/6 mice. Interestingly, whereas most UAEC strains colonized the C57BL/6 mouse bladder as well as the C3H/HeN bladder at 24 hpi, two strains, 31.1a (clade B1, PUF score=4) and 41.1a (clade B2, PUF score=20), showed significantly lower bladder bacterial burdens in C57BL/6 mice (Mann-Whitney U test, $p<0.01$) (**Figure 7**). Notably, the reduction in colonization ability did not appear to be based on genetic background or phylogeny of the bacterial strains, as other UAEC strains carrying similar numbers of PUFs and strains from the same clades as 31.1a and 41.1a colonized C3H/HeN and C57BL/6 mouse bladders at similar levels.

Quiescent intracellular reservoirs are small collections of dormant UPEC cells that remain in LAMP-1-positive vesicles in the underlying uroepithelium and represent another mode of bladder colonization within a different host environment. Quiescent intracellular reservoirs persist in the bladder even after apparent clearance of bacteria from the kidneys and urine and may be able to seed same-strain rUTI (Mysorekar and Hultgren, 2006). In addition to acute cystitis, C57BL/6 mice have been used to model the formation of quiescent intracellular reservoirs. In general, quiescent intracellular reservoirs can be detected through plating bladder homogenates 14dpi after the clearance of bacteriuria and kidney infection, as has been done previously (Mysorekar and Hultgren, 2006). We tested 4 robust colonizer UAEC strains, including 9.1a (clade D, PUF score=13), 9.2p (clade B1, PUF score =3), 35.1a (clade B2, PUF score=16) and 41.1a (clade B2, PUF score=20), and the model B2 UTI89 (PUF score=17) and found that all 5 strains, resulted in detectable bladder CFUs at 14dpi (**Figure S11**), despite

having sterile urines and kidneys, indicating the existence of quiescent intracellular reservoirs. However, the non-B2 strain, 9.1a, produced significantly lower bladder burdens at this time point than other tested strains, including the non-B2 strain, 9.2p, which was isolated from the same patient. Strain 9.1a also produced fewer IBCs than other strains tested (**Table S6**). Thus, diverse UAEC strains could colonize multiple host backgrounds. However, the relative efficiency of mouse bladder colonization varied between different UAEC strains in the same host background, and individual UAEC strains had different relative abilities to infect the bladders of mice with different genetic backgrounds.

2.3.8 UAEC strains that robustly colonized mice were correlated with markers of increased UTI severity in patients.

An overly exuberant inflammatory response to an initial UTI is associated with severity of subsequent infections in mice (Hannan et al., 2010) and susceptibility to recurrent infection in women (Hannan et al., 2014). This inflammatory response is cyclooxygenase-2 (cox-2) dependent and is associated with increased neutrophil infiltration across the urothelium during infection, which results in increased levels of white blood cells in the urine (Hannan et al., 2010; 2014). In addition to neutrophils, the host protects itself from UTI through secretion of lipocalin-2, which is induced upon infection of the bladder and helps to sequester iron away from the infecting organism (Steigedal et al., 2014). Importantly, levels of lipocalin 2 in the urine correlate to increasing levels of bacteria in the urine, or bacteriuria, in women (Steigedal et al., 2014). To determine if UAEC strains having increased virulence in the mouse model caused more severe UTI in women, we examined levels of white blood cells and lipocalin 2 in our cohort of patients infected with robust, variable, and deficient colonizer UAEC. We found that strains defined as robust and variable colonizer in the C3H/HeN mouse model were associated with markedly higher levels of white blood cells in the urine than deficient colonizers during

UTI episodes (**Table S13**). Further, we found that robust colonizers were associated with higher levels of lipocalin 2 in the urine than either variable or deficient colonizers (**Table S13**).

2.4 Discussion

2.4.1 Summary of findings

Although virulence in other *E. coli* pathotypes has been tightly linked to carriage of specific genes, UPEC are distinct in their lack of a signature set of genes that universally distinguishes them from non-UPEC strains. This likely reflects the broad definition of UPEC – any *E. coli* strain that is recovered from the urine of a symptomatic UTI patient – a classification that fails to account for differences in host susceptibility, or the possibility that multiple evolutionary and mechanistic paths can lead to urovirulence. To better define the determinants of urovirulence, we focused our study on a collection of 43 diverse UAEC strains isolated from a cohort of 14 women with frequent rUTI. Here, we have integrated genomic and transcriptomic approaches with *in vitro* and *in vivo* characterization of isolate phenotypes, including their colonization and pathogenic potential in defined experimental mouse models of cystitis. Our findings expand our understanding of the genomic diversity of UPEC, show that differential transcriptional regulation of core genes contributes to pathogenic potential and that the relative virulence of different *E. coli* strains varies depending on host background. We conclude that the range of clinical UTI outcomes reflects a range of complex host-pathogen interactions driven, in part, by variation in how uropathogens regulate core functions, impacting their pathogenic potential in hosts with different genetics and health status.

2.4.2 Implications of UAEC genetic heterogeneity

Phylogenomic analysis of this collection of UAEC strains revealed patterns of phylogenetic diversity previously reported in other clinical studies (Ejrnaes et al., 2011; Luo et

al., 2012; Piatti et al., 2008; Rijavec et al., 2006; Skjøt-Rasmussen et al., 2011; Starcic Erjavec et al., 2007), including an enrichment of clade B2 strains (67% of total) and a mixture of same-strain and different-strain rUTIs, with B2 strains dominating same-strain rUTIs (11 of 12 same-strain rUTI isolates fell within the B2 clade) (Luo et al., 2012). We observed variation in the genomic content of these isolates, up to 40% in pairwise comparisons between strains, similar to the level of genomic diversity observed in *E. coli* overall (Rasko et al., 2008; Touchon et al., 2009). Genomic diversity among *E. coli* strains is thought to be key to the species' ability to thrive in a multitude of environments (Touchon et al., 2009). Thus, it is possible that the genetic diversity observed in our cohort of UAEC strains, where two strains may share only 60% of the genes in their genome, reflects the sum of selection pressures across multiple habitats including but not limited to reservoirs in the gut, intracellular and extracellular locales in the bladder and diverse host backgrounds. In addition to other genes, we found that UAEC were variable in their carriage of PUF genes, which have previously been identified as being enriched in UTI isolates relative to other *E. coli* strains (Johnson and Stell, 2000; Johnson et al., 2001b; 2015; Luo et al., 2012). In contrast to this paradigm, we found that, if phylogeny was taken into account, there was no enrichment of PUFs in UAEC strains relative to non-UAEC strains. Thus, the enrichment of PUFs in urinary isolates of *E. coli* reported previously may be due to a combination of the enrichment of clade B2 in urinary isolates and the phylogenetic bias in PUF carriage in clade B2 strains.

2.4.3 Limitations of current study

The molecular mechanisms of pathogenicity for many PUFs remain poorly defined and may have functions outside of bladder pathogenicity that still enhance the overall virulence of UPEC strains, such as increasing successful migration from the gut reservoir to the bladder or

enabling persistence in host niches outside of the bladder, such as the vagina. One limitation of our study is that it is solely focused on bladder virulence and thus, functions outside of the bladder were not tested. Further, given the fact that many of these factors are co-localized on pathogenicity-associated islands common to the B2 clade and absent from other *E. coli* (Dobrindt et al., 2010), it is possible that the enrichment of PUFs results from a type of "genetic hitchhiking" due to some selection for B2 strains, such as their transcriptional profiles or persistence in gut reservoirs, rather than selection for the PUF genes themselves. Strikingly, we found that PUF carriage did not correlate with urovirulence in C3H/HeN mice. Indeed, non-B2 UAEC that carried few PUFs were as proficient as B2 UAEC carrying many PUFs in causing cystitis in mice. We also observed that in two patients, a UAEC strain with a higher number of PUFs was supplanted by a UAEC strain with a lower PUF score. In addition, we found in a mouse model, that the supplanting strains outcompeted the initial UTI strains in co-infection experiments. However, we found that same-strain rUTI isolates were much more likely to be from clade B2 than from other clades, which suggests that B2-specific genomic features, such as carriage of PUFs, other genes and/or transcriptional regulatory machinery, may provide an advantage for persisting within hosts between symptomatic episodes of UTI. Further, while PUF carriage was neither necessary nor sufficient for bladder colonization at acute time points, we did find that deletion of a large pathogenicity island from UTI89, which carried 124 genes including 4 PUFs, resulted in a loss of fitness during competition in chronic cystitis. While this result does not directly implicate these four PUFs in bladder colonization, it does show that carriage of at least some of the genes in this island may provide competitive advantages during different aspects of disease. All of our studied isolates came from the urine of women either during or between episodes of rUTI and thus, all of them, by definition, were uropathogens able to colonize at least

one human bladder. However, since host genetics, environment and behavior are important for UTI susceptibility, it is likely that many UAEC strains isolated from a single patient may be unable to cause infection in all or even most human bladders. To control for these inter-host variations, we assessed the colonization potential of UAEC and non-UAEC strains in genetically defined animal models under identical environmental conditions. While mouse bladder colonization by UAEC strains was markedly higher than by gut-associated *E. coli*, we observed that bladder colonization varied among UAEC isolates under these conditions. However, even after comparing the gene content of robust and deficient colonizer UAEC isolates, we found no significant differences in specific gene content or overall functional potential that could account for the variability of these strains in colonizing mouse bladders. Another important limitation of our study is our focus on gene carriage among strains. Since our comparative analysis focused solely on gene content, as determined by orthology, we would not have been able to identify more subtle genomic differences that could influence uropathogenicity including variation in promoter sequences or in coding sequences that might impact gene expression or protein structure and function.

2.4.4 The role of gene regulation in preparing UAEC for bladder colonization

While we found no clear genomic signatures that could be identified to discriminate between robust and deficient colonizers, we found that differences in the expression of core genes shared among all UAEC strains were predictive of urovirulence in C3H/HeN mouse bladders. These included genes involved in maltose transport, chemotaxis and flagellar assembly that mediate core bacterial behaviors such as nutrient utilization and motility. The variance in expression of core functions contrasts with studies showing that shared genes are expressed similarly in different *E. coli* strains (Bielecki et al., 2014), possibly reflecting different

definitions of the "core genome/transcriptome" between the studies. Our definition was based on comparisons of genomic sequences while previous reports have identified genes that were "commonly transcribed" to identify a core transcriptome. Importantly, this variance in transcriptional regulation of core bacterial behaviors has a direct impact on the ability of UAEC strains to colonize host bladders. This is exemplified by the differential regulation of the conserved *fim* operon, which encodes type 1 pili. While all but one of our UAEC strains carried an intact *fim* operon, the strains varied considerably in their hemagglutination titers, a measure of piliation. Further, hemagglutination titers correlated well with colonization of the C3H/HeN mouse bladder. We posit that that the regulatory networks that control type 1 pili expression within the different UAEC strains respond differently to environmental cues. For example, a set of highly related strains (from patients 20, 21, and 35), all carrying the *fim* gene cluster, expressed S pili instead of type 1 pili, when grown under standard type 1 pilus-inducing conditions (Hultgren et al., 1986). These strains were robust colonizers of C3H/HeN mice, indicating that their increased expression of S pili under inoculum conditions did not prevent their urovirulence. In the UTI89 strain, expression of S and type 1 pili is inversely controlled so that blocking expression of one pilus type induces the expression of the other (Wurpel et al., 2013). Further, differences in type 1 pili regulation are seen in the ways that UAEC regulate their motility and pili expression relative to model B2 UPEC strains. For example, increased expression of type 1 pili is inversely correlated with expression of flagellar genes in UTI89 (Lane et al., 2007). Here, we found that the non-B2 strains exhibiting high expression of type 1 pili also had increased motility, indicating that the networks mediating coordinated regulation of flagella and type 1 pilus expression may differ between B2 and non-B2 UAEC strains. Taken together, our findings implicate the divergence of transcriptional and regulatory networks as a

key driver of UPEC pathogenesis and underscore the importance of examining transcriptional regulation of genes in addition to their patterns of carriage to more fully understand the factors underlying bacterial virulence and host-pathogen interactions. Moreover, our findings suggest that transcriptional responses to the host environment likely also diverge among UAEC strains, and that to gain richer insights into the determinants of UTI risk and outcome, we need a better understanding of how gene expression differences during infection influence bacterial physiology and interactions with the host.

2.4.5 Translation of findings to human disease

The progression and outcome of a UTI is determined not only by the virulence potential of the infecting bacteria, but by myriad factors in the host environment. Mouse models of UTI recapitulate the histological markers of human UTIs and multiple mouse models have been developed that each reflect a portion of the diversity of UTI pathology seen in the clinic [reviewed in (Barber et al., 2016; Carey et al., 2016)]. Consistent with their ability to cause UTIs in humans, many of our UAEC strains were able to elicit key hallmarks of human pathogenesis in mouse models of UTI, including acute cystitis, the formation of IBCs, the development of chronic cystitis, and persistence in quiescent intracellular reservoirs. Importantly, we found that the ability of certain UAEC strains to cause acute cystitis in the C3H/HeN mice correlated with markers of increased UTI severity in patients. Specifically, UAEC strains that exhibited an increased ability to cause acute cystitis in the C3H/HeN mouse model were associated with an increase in leukocytes and lipocalin 2 in the urine of the patients from which the strains were derived. Lipocalin 2 is known to be associated with increased activation of the immune response in human UTIs (Steigedal et al., 2014). Additionally, transmigration of neutrophils across the

bladder epithelium, which results in an increase in leukocytes in the urine, has been associated with UTI severity and recurrence (Hannan et al., 2014).

2.4.6 Diversity in host-pathogen interactions during UTI

Our study suggests that each infection by a single UAEC strain in a single mouse background captures only a small part of the complex landscape of factors that govern UTI progression and outcome. Specifically, we found that, while most UAEC strains had similar success colonizing the bladder of C57BL/6 and C3H/HeN mice, two phylogenetically and genetically diverse strains were able to colonize the C3H/HeN bladder much better than the C57BL/6 bladder, indicating that some feature of the C57BL/6 bladder environment presents a specific barrier to colonization that could be overcome by some strains but not others. These findings suggest that the barriers to infection vary among mouse backgrounds as do the capacities of different UAEC strains to overcome these barriers. Thus, the outcome of each UTI is determined by the compatibility of bacterial virulence and host susceptibility factors involved in that infection. Gaining a more comprehensive understanding of the factors that determine UTI risk and outcome will require studies of urovirulence using more combinations of UAEC isolates and mouse backgrounds that better represent the diversity of bacterial urovirulence potentials and host susceptibilities.

2.5 Materials and Methods

2.5.1 Study design

This study was conducted to identify conserved bacterial features that enabled *E. coli* bladder colonization and virulence and determine if these features enabled virulence in all host backgrounds. In a previous study that was approved by the Human Subjects Review Committee at the University of Washington, 104 women aged 18-49 years with a self-reported history of

UTI and a current diagnosis of acute cystitis were enrolled in the analysis cohort and self-collected mid-stream urine samples daily for 90 days (d), as described previously (Czaja et al., 2009). Women were recruited from University of Washington Health Centers in Seattle, Washington. Exclusion criteria for this cohort included known anatomic of functional abnormalities of the urinary tract, chronic illness, pregnancy, and development of acute pyelonephritis (Czaja et al., 2009). From this cohort, a total of 29 women experienced a rUTI within the 90d study window and provided urine samples containing urine-associated *E. coli* (UAEC). We examined strains isolated from the first 14 of these 29 women, resulting in a total of 43 UAEC isolates collected throughout the study (**Figure S1** and **Table S2**). These isolates 14 isolates collected at enrollment, 18 isolates collected during rUTI, and 11 isolates from daily urine samples collected at home by the patients in the days preceding the diagnosis of UTI (**Table 1**). A single *E. coli* isolate was collected from each urine sample using selective plating techniques (Czaja et al., 2009), and all isolates were stored at -80ºC. When possible, levels of white blood cells and lipocalin 2 were measured in the urine samples, as described previously (Czaja et al., 2009). *E. coli* rUTI was defined based on: i) the presence of *E. coli* in the urine at levels $\geq 10^2$ CFU/mL and ii) medical diagnosis of a UTI based on symptoms of acute cystitis, such as foul-smelling urine, increased frequency of urination, and/or urgency to urinate.

2.5.2 Additional strains used in this project

The model strains MG1655, Sakai and Nissle 1917 were derived from pure cultures and stored in glycerol at -80ºC. As described in a previous study, gut-associated *E. coli* were isolated from human feces collected from healthy adults (Gordon et al., 2005; Snyder et al., 2012). Briefly, single colonies were identified as *E. coli* with isolation plating on differential media and glycerol stocks were stored at -80ºC.

2.5.3 DNA extraction

UAEC isolates were grown shaking overnight at 37ºC in LB and then subcultured 1:100 into fresh LB for exponential growth (OD_{600} of ~0.4). Bacterial cells were then pelleted, and their DNA was extracted using the Wizard DNA Purification Kit (Promega Corp) following manufacturer's instructions. DNA quality and quantity was measured using the NanoDrop1000 (Thermo Scientific) spectrophotometer and gel electrophoresis.

2.5.4 Genome dataset construction

A multiplex library of paired-end 101bp sequencing reads was produced using the Illumina HiSeq 2000 platform with a 300bp insert size. This library was demultiplexed and unordered contigs were assembled *de novo* using Velvet v1.2 (Zerbino and Birney, 2008)(**Table S14**). A preliminary phylogeny was constructed using *in silico* multi-locus sequence typing (Enright and Spratt, 1999) to identify closely-related reference *E. coli* genomes, which were then used to order each strain's contigs using Mauve v2.3.1 (Rissman et al., 2009). In order to contextualize the UAEC genomes, we constructed a database containing the 43 draft UAEC genomes as well as 46 closed *E. coli* genomes from the National Center for Biotechnological Information (NCBI) (**Table S3**). All genomes were re-annotated using the Broad Institute's prokaryote annotation pipeline (Lebreton et al., 2013).

2.5.5 Orthogroup clustering

Orthogroup clustering was done using the reciprocal best BLAST hits method (Salichos and Rokas, 2011). All-vs-all BLAST searches were performed between the genes in all 89 genomes. In cross-genome BLAST searches, only the top 5 best hits were captured. In self-searches, up to 1000 top matches were captured. The resulting BLAST hits were clustered using single-linkage clustering (transitive closure). Paralogs were identified by searching the self-genome BLAST results, identifying gene pairs (A, B) where gene A is in an ortholog cluster, and

gene B matches A with a BLAST score that exceeds that of gene A with any of its cross-species orthologs.

2.5.6 Phylogenetic analysis

A single-copy core genome was concatenated from 2746 orthogroups shared between all 89 *E. coli* genomes and aligned using MUSCLE (Edgar, 2004). A phylogenetic tree was then constructed with RAxML and the GTRCAT model (Stamatakis et al., 2005). Bootstrapping was performed using RAxML's rapid bootstrapping algorithm (Stamatakis, 2006). To identify clusters of highly related strains, we calculated the number of SNPs differing between each pair of strains between the core genes in our concatenated single-copy core orthogroup alignment using custom scripts. These SNP counts were then used in unsupervised hierarchical clustering to identify groups of highly related strains. Clustering was performed with the Euclidean distance metric and the Complete clustering algorithm and visualized with the package pheatmap in R. Using this clustergram, we defined "core clonal groups" as collections of UAEC isolates with $<1.5x10^{-5}$ SNPs/bp in pairwise alignments that were isolated from the same patients. Within our dataset, we identified same-strain rUTI as events caused by strains that: i) were from the same core clonal group (as described above); ii) were isolated at time of UTI (either enrollment or recurrence), and; iii) were isolated from the same patient.

2.5.7 Global gene content analysis

Gene content analysis was performed using orthogroup clusters from representative UAEC strains that were robust or deficient colonizers of C3H/HeN mouse bladders. Using stringent criteria, we searched for orthogroups present in all members of one category (i.e. either robust or deficient colonizers), and absent in all members of the second category. For a more relaxed comparison, we searched for orthogroups present in >80% of one group and <20% of the

other group. Further, we also identified the gene categories (Pfam (Finn et al., 2008) and KEGG (Ogata et al., 1999)) most expanded or reduced when comparing groups of strains by using Fisher's Exact test with multiple hypothesis test correction. In order to select for gene categories uniformly enriched in one set of strains, we additionally required that the standard deviation of copy number in each set of strains be less than the difference between the mean of the copy number in the two sets of strains. Further, a large percentage of these variable annotations are associated with transposon or phage genes, which are difficult to assemble using *de novo* methods and prone to errors in their copy number within draft genomes. Therefore, we removed Pfam domain predictions associated bacteriophages or transposons.

2.5.8 Modeling of Type I error rate in gene content analysis

To estimate the number of genes that would be enriched in robust and deficient UAEC due to random chance, we performed a repeated permutation experiment (Sefik et al., 2015) where we randomly assigned 6 robust and deficient UAEC strains and UTI89 (for a total of 17 strains) into sets of 13 and 4 strains (to match the distribution of strains in the robust and deficient groups) and compared the gene carriage of these random groupings using the relaxed definition of enrichment. From 1000 permutations, the mean number of hits between the randomly assigned groups was 6.6 genes. Notably, a limited sample size in comparative genomics is likely to result in an overestimation of genetic enrichment (i.e. low N is usually associated with a higher rate of Type I errors).

2.5.9 Selection of representative UAEC used in phenotypic and colonization assays

Representative UAEC strains were selected to represent the genomic and patient diversity in our collection. Thus, we selected only one isolate from each clonal strain group in each patient, which resulted in a total of 21 unique UAEC strains represented in these analyses. There were no

significant differences in guinea pig red blood cell hemagglutination, swimming motility, or curli and biofilm production phenotypes between strains in the same clonal group from the same patient (data not shown).

2.5.10 Targeted interrogation of virulence factor carriage, type 1 pilus and flagella genes

A total of 31 putative urovirulence factor (PUF) gene sequences were selected for this analysis based on: i) their enrichment in UPEC isolates collected from women with UTIs relative to other *E. coli*, ii) their impact on urovirulence in mouse models of cystitis, and iii) their use as genetic markers in studies of UPEC virulence in previous research (**Table S4**). The distribution of putative PUFs in our clinical *E. coli* isolates was determined using custom BLAST searches using the 31 PUF gene sequences as queries against a database containing the draft genome sequences of our clinical *E. coli* isolates and the reference genome of UTI89. A "hit" was considered as any genome sequence that matched the entire length of the query sequence with >75% identity. As a control to prevent false negatives in the BLAST search, DNA sequencing reads from each representative UAEC isolate were mapped against a reference sequence constructed by concatenating all the PUF gene sequences with 100 N's added to the gene sequences to separate the concatenated genes using Geneious v6.1.7 (Kearse et al., 2012). Any PUF gene that was completely covered by the DNA sequencing reads to a depth of 4X was considered a "hit". This same protocol was used to investigate the type 1 pili genes in the *fim* operon and the genes mediating flagella assembly and function. A binary hit matrix was constructed using the results from this analysis, where each hit for each PUF within each strain was counted as a 1 and all other data points were set to 0. Two-dimensional cluster analysis on this matrix was performed and visualized in the R software package using complete linkage clustering on the Jaccard Distance metric and ggplot2 for visualization.

2.5.11 Construction of antibiotic resistance marked strains

A chloramphenicol resistance marker was integrated into the HK-site of strains 41.4p and 9.1a, while a kanamycin resistance marker was integrated into the HK-site of 9.2p using the λ red recombinase system (Datsenko and Wanner, 2000). For strain 41.1a, a spectinomycin resistance marker was integrated into the *attB* site by first transforming in the integrase expressing vector pINT-TS followed by integration of pPSSH10 containing a spectinomycin resistance cassette, as described previously (Wright et al., 2005).

2.5.12 Mouse infections

Single acute infections

~$1x10^8$ bacteria suspended in PBS were inoculated into lightly anesthetized 7-8 week old female C3H/HeN or C57BL/6 mice transurethrally in a 50µl injection (Conover et al., 2015; Hung et al., 2009). At 24 hpi bladders and kidneys were harvested from these mice and homogenized in sterile PBS. Homogenates were serially diluted and plated for bacterial enumeration.

Competitive acute infections

For competitive infections, strains were inoculated with equal CFU of the two strains transurethrally into 7-8 week old female C3H/HeN mice. Urines were collected at the indicated days post-infection and plated on duplicate LB plates containing spectinomycin, kanamycin or chloramphenicol to enumerate CFU of each strain. For competitive infections, strains were differentially marked with antibiotic resistance cassettes. Competitive indexes were calculated based on the ratio of the supplanting strain (either 9.2p or 41.4p) over the initial strain (either 9.1a or 41.1a) for each respective competition.

Single chronic infections

7-8 week old female C3H/HeN mice were transurethrally inoculated with $1x10^8$ UAEC and urines were collected at the denoted time points for 28 days for determination of bacteriuria. After 28 days the bladders and kidneys were harvested, homogenized and plated for bacterial enumeration.

Competitive chronic infections

Strain TJH2 (UTI89ΔPAI II_{UTI89}) was constructed as previously described (Hannan et al., 2008). 7-8 week old female C3H mice were transurethrally inoculated with an equal numbers of TJH2 and UTI89 WT together totaling ~$1x10^8$ CFU. Urines were collected at indicated time points for 28 days for determination of bacteriuria and to track chronic competition. After 28 days, bladders and kidneys were harvested, homogenized and replica plated on LB plates containing the appropriate antibiotics for strain enumeration. Competitive indexes were calculated based on the ration of TJH2 over UTI89 WT.

Enumeration of intracellular bacterial community formation

7-8 week old female C3H/HeN mice were inoculated with ~$1x10^7$ CFUs of the indicated strain. Mice were sacrificed 6 hpi and bladders were bisected and splayed for fixation in 4% paraformeldahyde for 1.5 hours and staining with rabbit α-*E. coli* antibody detected with Alexa Fluor 488 labeled goat α-rabbit secondary antibody. Bladder cells were counter-stained with Alexa Fluor 594 labeled wheat germ agglutinin.

Enumeration of persistent bladder reservoirs

To assess reservoir populations, 6 week old C57BL/6 mice purchased from were transurethrally inoculated with 50 µl of PBS containing ~$1x10^7$ CFU of the indicated UAEC strains. Urine titers were monitored for 2 weeks by dilution plating of collected urines on MacConkey agar plates, at which point the mice were sacrificed and bladder and kidney tissues were homogenized and bacterial burdens determined by dilution plating. Mice that developed

chronic infection, as indicated by persistent bacteriuria (>10^4 CFU/ml urine), or that had kidney abscesses at sacrifice (determined visually) were omitted from the analysis as the bladder bacterial burdens of these mice are not conclusively due to reservoir populations (Mysorekar and Hultgren, 2006).

2.5.13 Phenotypic analyses

Guinea pig hemagglutination assays

Clinical UAEC were grown under type 1 pili-inducing conditions as described. Bacteria were harvested and resuspended in PBS and normalized to an optical density at 600nm (OD_{600}) of 10 in 100 uL (~10^8 CFU) and then serially 2-fold diluted in either PBS or PBS with 4% methyl-α-D-mannopyranoside and incubated overnight at 4ºC with equal amounts of guinea pig erythrocytes (resuspended to an OD_{640} of 2 in PBS)(Hultgren et al., 1986). In HA assays, the titers indicate the maximum dilution still capable of agglutinating guinea pig erythrocytes (Greene et al., 2014). Methyl-α-D-mannopyranoside acts as a competitive inhibitor of type 1 pilus-mediated adhesion, abolishing agglutination by type 1 pili. Comparison of the HA titers of the same bacterial strain incubated with and without mannose analogs is used to measure the MSHA of bacteria, corresponding to type 1 pili binding. The results represent a mean average of 2 technical replicates in each of 2-4 biological replicates.

Swimming motility assays

Swimming motility was assessed after culture in in our inoculum conditions and normalization to an OD_{600} of 1 in 1mL. A sterile inoculating loop was dipped into the sample and then stabbed into the center of 12mL of 0.25% agar LB plugs in six-well plates, as described previously (Greene et al., 2014). Inoculated plates were incubated for 6h at 37ºC and visually examined the plates for evidence of non-swimming motility, such as swarming. Swimming

motility was determined as the mean of the diameter of the spread of swimming bacteria in 3 technical replicates.

2.5.14 RNA-Seq experiments

RNA-Seq data generation

In three separate experiments, RNA was extracted from select UAEC strains after growth in conditions known to induce type 1 pili expression and prepare *E. coli* strains for inoculation into mice (see Mouse Infections). Bacteria were then pelleted and resuspended in TRIzol (Invitrogen), preheated to 65ºC, and flash frozen in an ethanol-dry ice bath. Cells were lysed using 0.1mm Silica-Zirconium beads (BioSpec Products), and the supernatant was transferred to new tubes containing 100% ethanol. RNA was then extracted from this mixture using the Direct-ZOL RNA MiniPrep Kit (Zymogen) following manufacturer's instructions. DNA was removed using the TURBO DNase Kit (Ambion), and the RNA was concentrated using the RNA Clean and Concentrator Kit (Zymo Research), both following manufacturer's instructions. RNA quantity and quality was measured using the BioAnalyzer chip (Ambion). Illumina cDNA libraries were generated using the RNAtag-Seq protocol as described (Shishkin et al., 2015). Briefly, 1ug of total RNA was fragmented, depleted of genomic DNA, and dephosphorylated prior to its ligation to DNA adapter carrying 5'-AN8-3' barcodes with a 5' phosphate and a 3' blocking group. Barcoded RNAs were pooled and depleted of rRNA using the RiboZero rRNA depletion kit (Epicentre). These pools of barcoded RNAs were converted to Illumina cDNA libraries in 3 main steps: (i) reverse transcription of the RNA using a primer designed to the constant region of the barcoded adaptor; (ii) degradation of the RNA and ligation of a second adaptor to the single-stranded cDNA; (iii) PCR amplification using primers that target the constant regions of the 3' and 5' ligated adaptors and contain the full sequence of the Illumina sequencing adaptors. cDNA libraries were sequenced on Illumina HiSeq 2000.

RNA-Seq data analysis

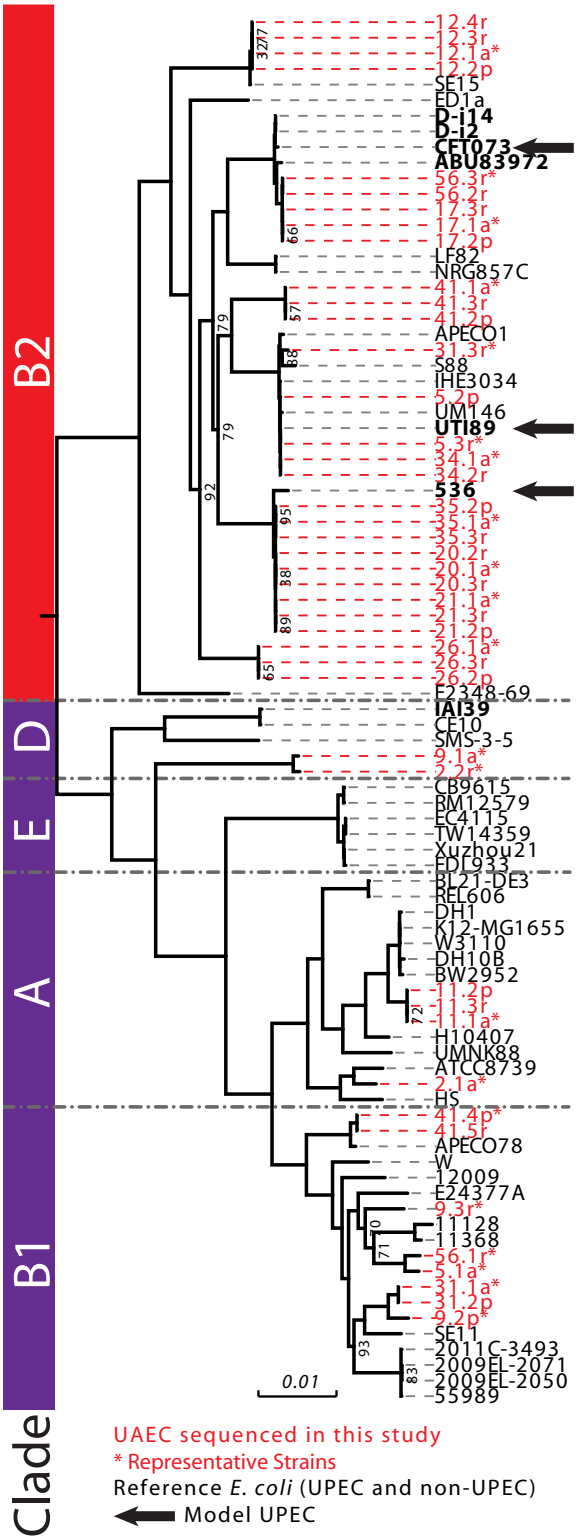
For the analysis of RNAtag-Seq data, reads from each sample in the pool were identified based on their associated barcode using custom scripts, and up to 1 mismatch in the barcode was allowed with the caveat that it did not enable assignment to more than one barcode. Barcode sequences were removed from reads, and the reads from each sample were aligned to genes in their cognate strain using BWA (Li and Durbin, 2010). To enable comparison of gene expression across different strains, we examined 3,516 "core transcriptome genes" defined as genes that were present in UTI89 and 9/10 of the clinical UAEC as determined by clustering of orthologous gene sequences. These genes were assigned the gene ID of their UTI89 homologues and all other groups of homologous genes were assigned a unique arbitrary ID. The number of reads aligning to genes corresponding to each ID was then calculated for each strain, with gene IDs not represented in a particular strain assigned the minimum value of the rest of the strains. Differential expression analysis was conducted with raw reads counts per gene using DESeq (Anders and Huber, 2010). We normalized the mapped read count by reads per million (RPM) to get normalized transcript abundance and normalized transcript abundance by transcripts per million (TPM). The expression level for each gene was estimated by averaging normalized read counts across triplicate samples. We excluded genes that had average expression levels that fell below the first quartile across 11 strains, and the remaining 2,637 genes were used in principle component analysis. Linear regression was performed on $\log_2$ transformed colonization measurements (i.e. CFU/bladder) and normalized individual gene expression levels. P-values of the linear regression were adjusted by Benjamini-Hochberg method to correct for multiple testing.

2.5.15 Statistical analysis

Colonization groups were defined based on the median value of mouse bladder colonization by UAEC strains at 24 hpi. Repeated non-parametric Mann-Whitney U tests were used to identify statistically significant differences in PUF gene carriage, hemagglutination assay titers, and swimming motility between colonization groups. Further, repeated Mann-Whitney U tests were used to identify differences in PUF gene carriage between clades, colonization efficiency of different inbred mouse strains by UAEC isolates, and persistence of quiescent intracellular reservoirs by UAEC in C57BL/6 mice. Correlations between median bladder burden in C3H/HeN mice and PUF gene carriage, hemagglutination assay titer, or motility in UAEC strains was determined using the Spearman rank correlation test in separate analyses. UAEC were clustered into related groups using unsupervised hierarchical clustering of SNP distances defined through pairwise alignments of the core genome. Further, unsupervised clustering was used to identify UAEC that shared similar carriage of PUF genes. A Hypergeometric Distribution test was used to compare enrichment of clade B2 strains causing same-strain rUTI relative to the distribution of B2 strains with our collection of clinical isolates. Differences in expression of core genes were measured using the DESeq algorithm. Principle component analysis (PCA) was used to visualize similarities in core gene expression and linear regression analysis was used to correlate these differences in gene expression to median bladder burden at 24 hpi in C3H/HeN mice. Type I error modeling was established with repeated permutation analysis with 1000 replicates and counts of gene presence/absence and is described in detail in the Supplementary Materials. Where p-values could be assigned, a threshold of significance was established at $P<0.05$ after multiple hypothesis testing correction ($P_{adj}<0.05$), except for the linear regression analysis of gene expression to bladder burden, which used a significance threshold of $P_{adj}<0.1$.

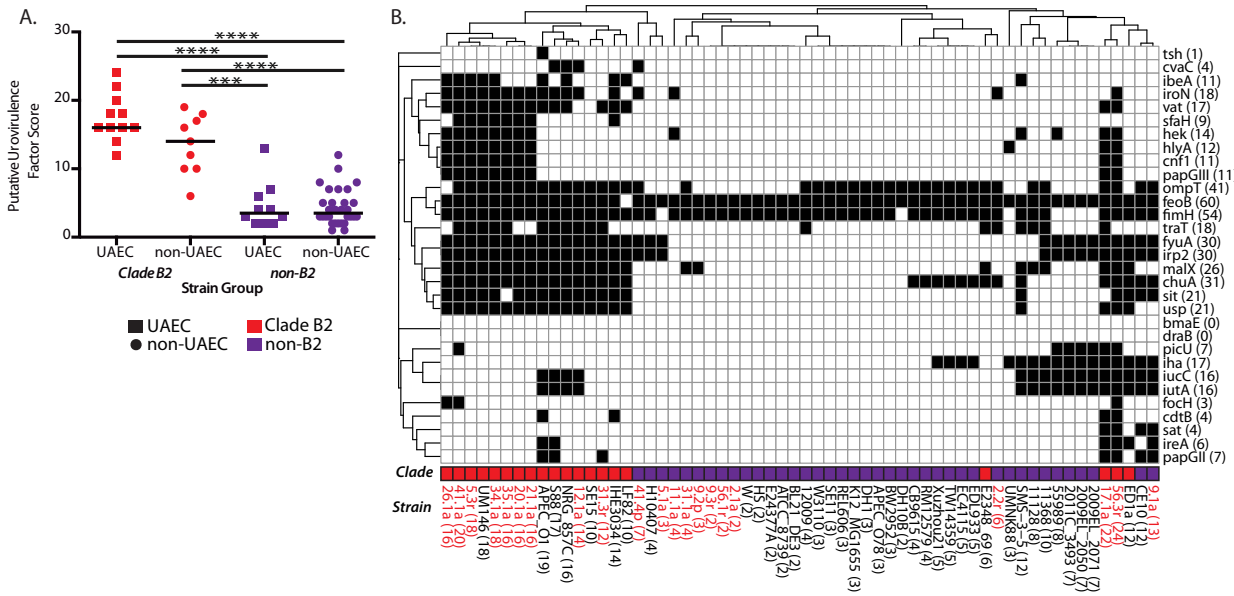
2.6 Figures

2.6.1 Figure 1. Phylogenetic distribution of UAEC strains from rUTI patients.



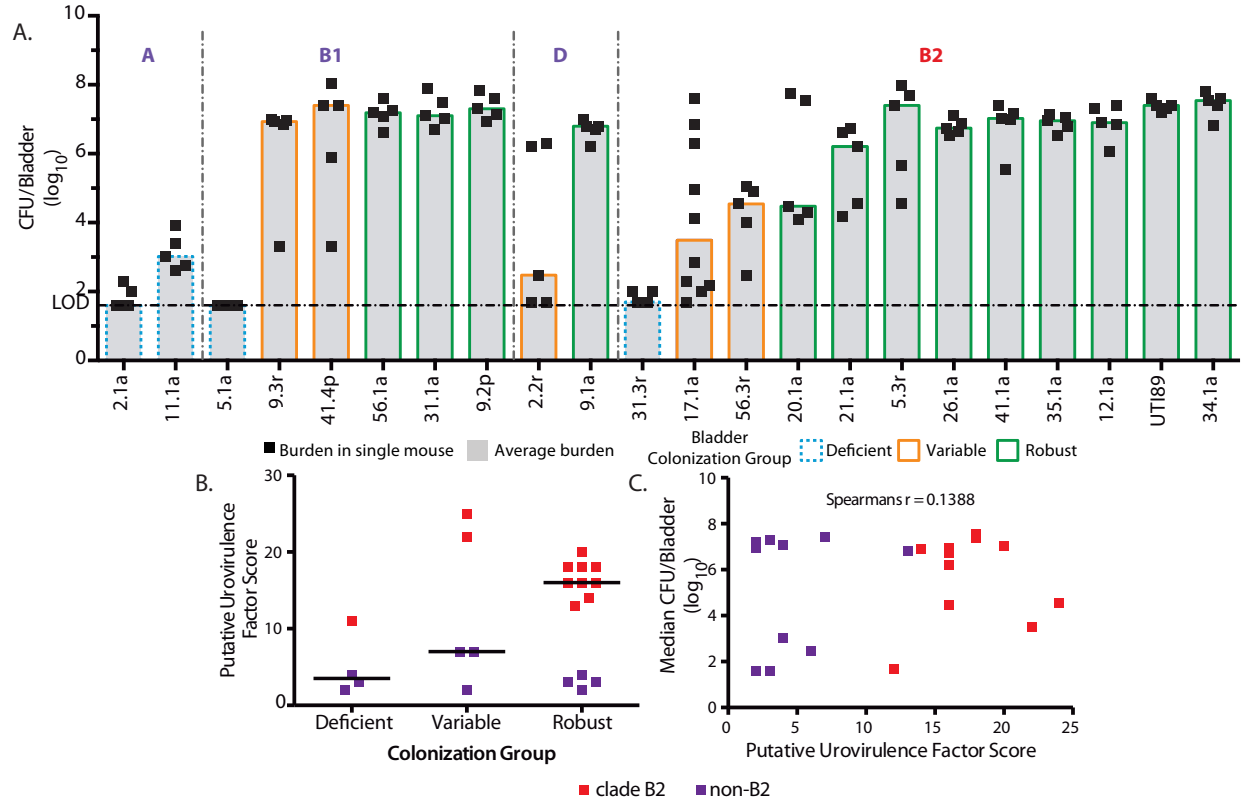
The phylogenetic relatedness of the urine-associated *E. coli* (UAEC) strains (n=43, taxon labels in red) was contextualized within the broader phylogeny of reference *E. coli* strains (n=46, taxon labels in black) by comparing the single-copy core genes of the strains using the RAxML algorithm. Reference *E. coli* strains that were associated with urinary disease (e.g. cystitis, pyelonephritis, or asymptomatic bacteriuria) are in bold. Asterisks indicate UAEC strains chosen as representative isolates for their clonal clusters (Fig. S2B). Bootstrap supports are indicated at internal nodes, and bootstrap values >95 have been removed. UAEC strains were found in four out of five *E. coli* clades (indicated by red and purple bars on the left). Black arrows indicate model UPEC strains commonly used in UTI research.

2.6.2 Figure 2. Carriage of putative urovirulence factors (PUFs) is enriched in both UAEC and non-UAEC strains from the B2 clade.



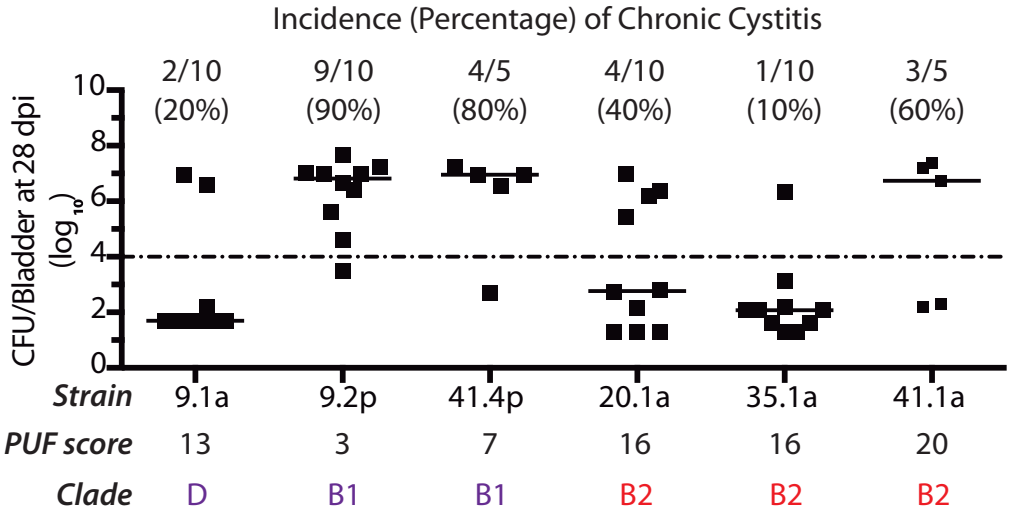
(A) Comparisons between indicated groups were performed using the Mann-Whitney U test. Significant differences were calculated for each group and significant results are indicated as: ***, *P<0.001;* ****, *P<0.0001*. (B) Clinical UAEC strains and reference *E. coli* strains that were not associated with urinary disease (non-UAEC) (x-axis) were examined for the presence of 31 PUF sequences (y-axis) using BLAST and alignment-based searches. The binary presence of PUFs (indicated by black squares) was tallied for each strain and for each PUF (indicated in parentheses on both axes). Two-dimensional hierarchical clustering identified clusters of PUFs that tended to co-occur in UAEC strains (dendrogram along the y-axis) and showed that PUF carriage was associated with phylogeny (dendrogram along the x-axis, phylogeny indicated in column labeled 'Clade').

2.6.3 Figure 3. Increased PUF carriage does not correlate with increased colonization efficiency.



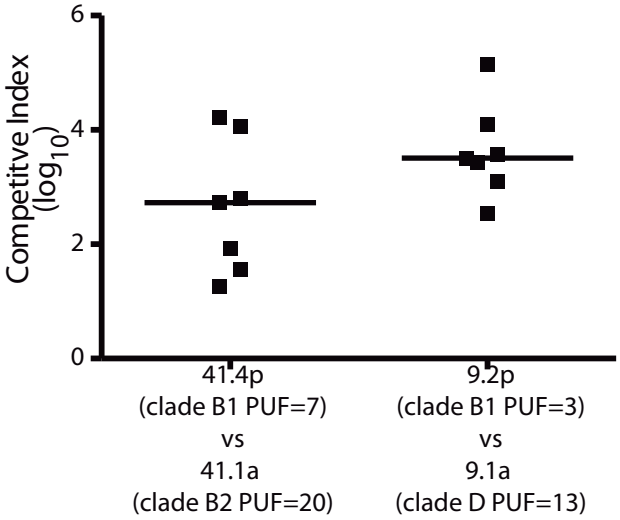
(A) The colonization efficiencies of 21 representative UAEC strains and the model strain UTI89 from indicated clades were tested in C3H/HeN mice. Bacteria were enumerated from individual harvested bladders at 24 hours post-infection (hpi) (black boxes). Each UAEC strain was categorized as: "deficient" (n=4, blue stippled outline) at $<10^4$ CFU/bladder, "variable" (n=5, orange outline) above and below 10^4 CFU/bladder, or "robust" (n=12, green outline) at $>10^4$ CFU/bladder. The black broken horizontal line represents the limit of detection of bacteria. Data presented represent the median (gray bar) of mouse inoculations for each strain. (B) No enrichment of PUF carriage, as measured by PUF scores, was found in comparisons of robust, variable or deficient colonizer strains. Comparisons were performed with Mann-Whitney U test; horizontal solid bars indicate median values. (C) Using Spearman's rank correlation (ρ statistic indicated at the top), there was no significant correlation between carriage of PUFs and bladder burden with UAEC strains at 24 hpi. Squares indicate the median bladder colonization of B2 (red) or non-B2 (purple) UAEC strains.

2.6.4 Figure 4. Both B2 and non-B2 UAEC strains cause chronic cystitis in mice.



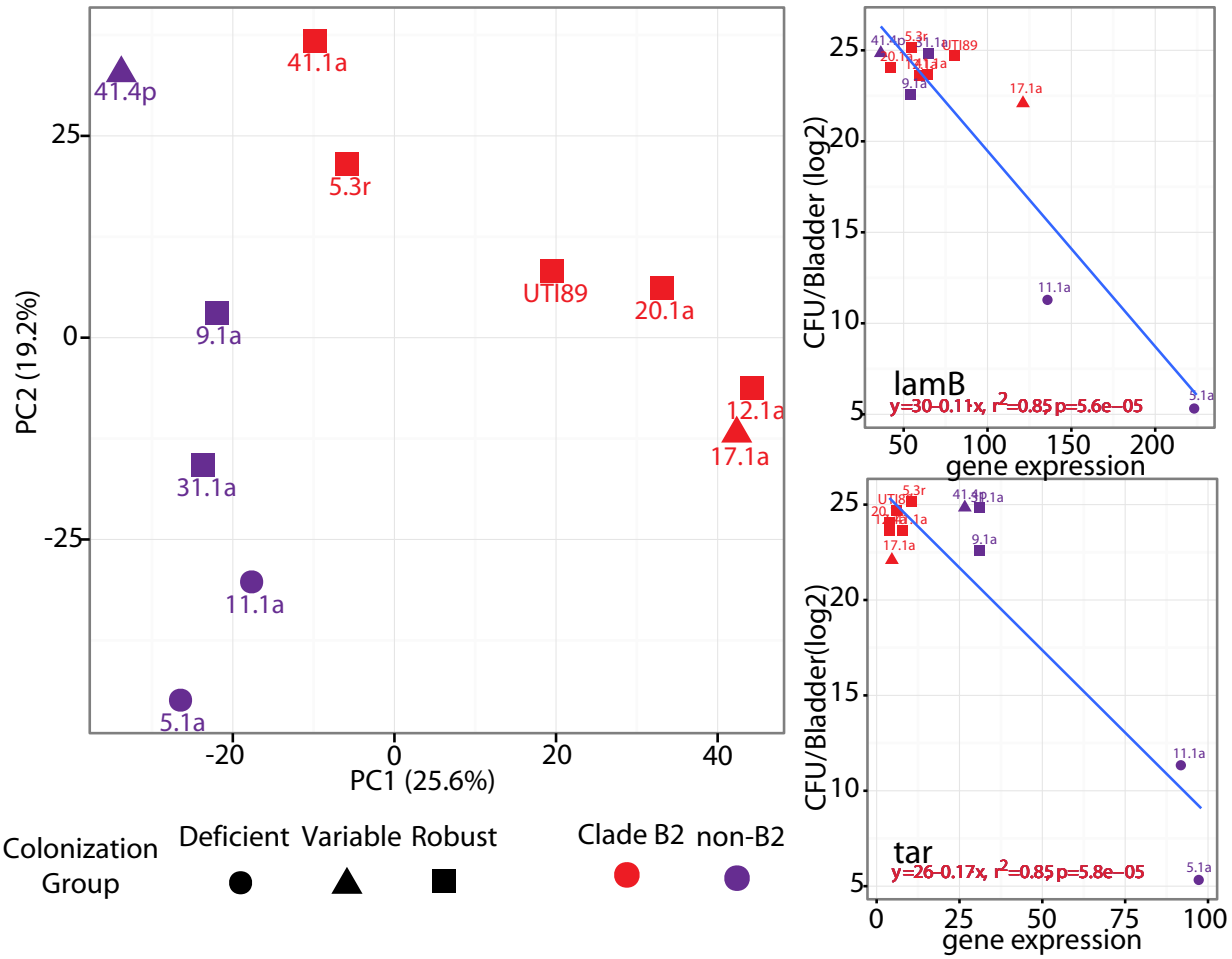
Incidence of chronic cystitis, defined as persistent high-titer bacteriuria (>10^4 CFU/mL urine), high-titer bladder colonization (>10^4 CFU/bladder) and chronic inflammation at 28 days post-infection (dpi), was measured for a subset of B2 and non-B2 UAEC strains infecting C3H/HeN mice. Both B2 and non-B2 UAEC strains could cause chronic cystitis in mice. PUF scores and incidence of chronic cystitis are indicated for each strain. Horizontal bars indicate median values.

2.6.5 Figure 5. PUF carriage does not increase competitive advantage during co-infection.



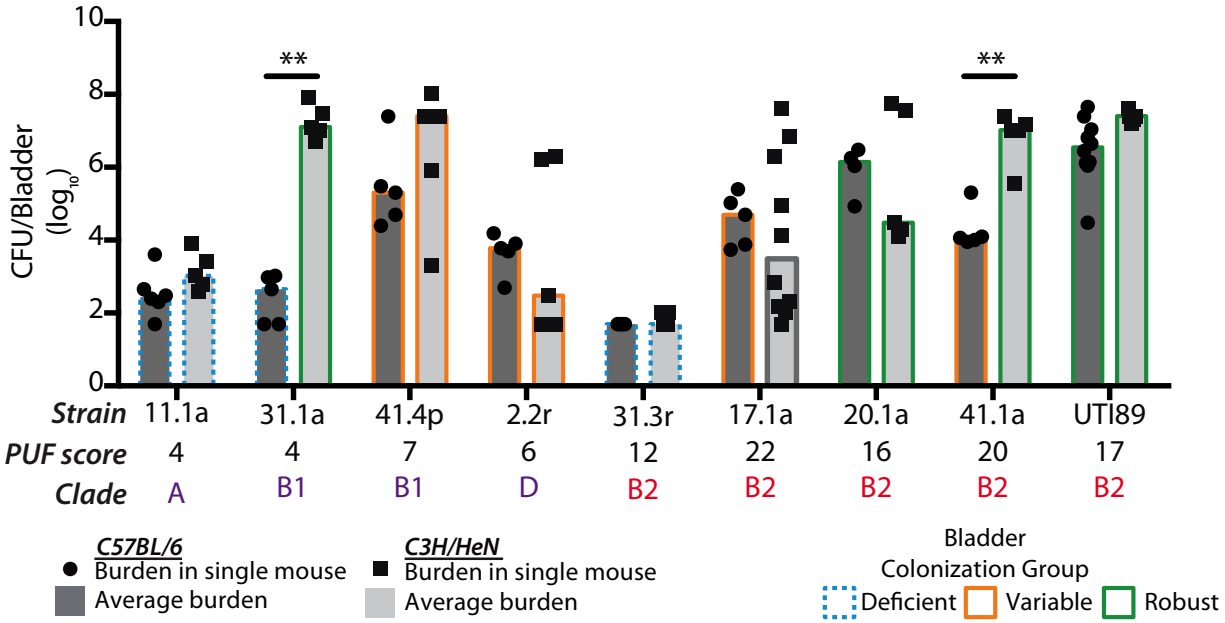
UAEC strains isolated from two patients (9 and 41) with different strain-induced rUTIs were differentially marked with antibiotic resistance markers. C3H/HeN mice were then coinfected with the supplanting strain (9.2p or 41.4p) at an equal dose with their enrollment strain from the same patient (9.1a or 41.1a, respectively). A competitive index (CI) was calculated from the ratio of the supplanting strains over their enrollment strains in the urine from each mouse at 24 hpi. Clades and PUF scores are indicated for each strain. Horizontal bars indicate median values.

2.6.6 Figure 6. Differential expression of core genes by UAEC strains distinguishes robust from deficient colonizers.



(A) Principle component analysis (PCA) was used to cluster UAEC strains based on their expression of 3,340 core genes under type 1 pili-inducing culture conditions used to culture bacteria before inoculation into mouse bladders. B2 strains separated from non-B2 strains along PC1, whereas robust colonizers separated from deficient colonizers along PC2. Linear regression identified the expression of 42 core genes in type 1-inducing conditions as correlated to bladder bacterial burden at 24hpi in C3H/HeN mice after correction for false discovery (*P-adjusted<0.1*). These 42 genes included pathways genes mediating nutrient uptake, such as *lamB* (B) and chemotaxis, such as and *tar* (C), whose correlations to bladder burden are visualized here as representative examples. Data presented represent averages derived from three independent experiments that were corrected for variance in read depth between samples and gene length.

2.6.7 Figure 7. Colonization efficiency by UAEC strains varies between mouse models.



Select UAEC strains were inoculated into C3H/HeN mice (black squares and light gray bars) and C57BL/6 mice (black circles and dark gray bars). Bacteria were enumerated from individual harvested bladders at 24 hpi (black squares and circles). Median values for each infection are given (gray bars). Each UAEC infection of the two mouse strains was categorized separately as "deficient" (blue stippled outline), "variable" (orange outline), or "robust" (green outline). Bladder colonization between C3H/HeN and C57BL/6 mice was significantly different for strains 31.1a and 41.1a by Mann-Whitney U test with **, $P<0.01$.

2.7 Tables

2.7.1 Table 1. Summary of 43 UAEC isolates from women with recurrent UTIs.

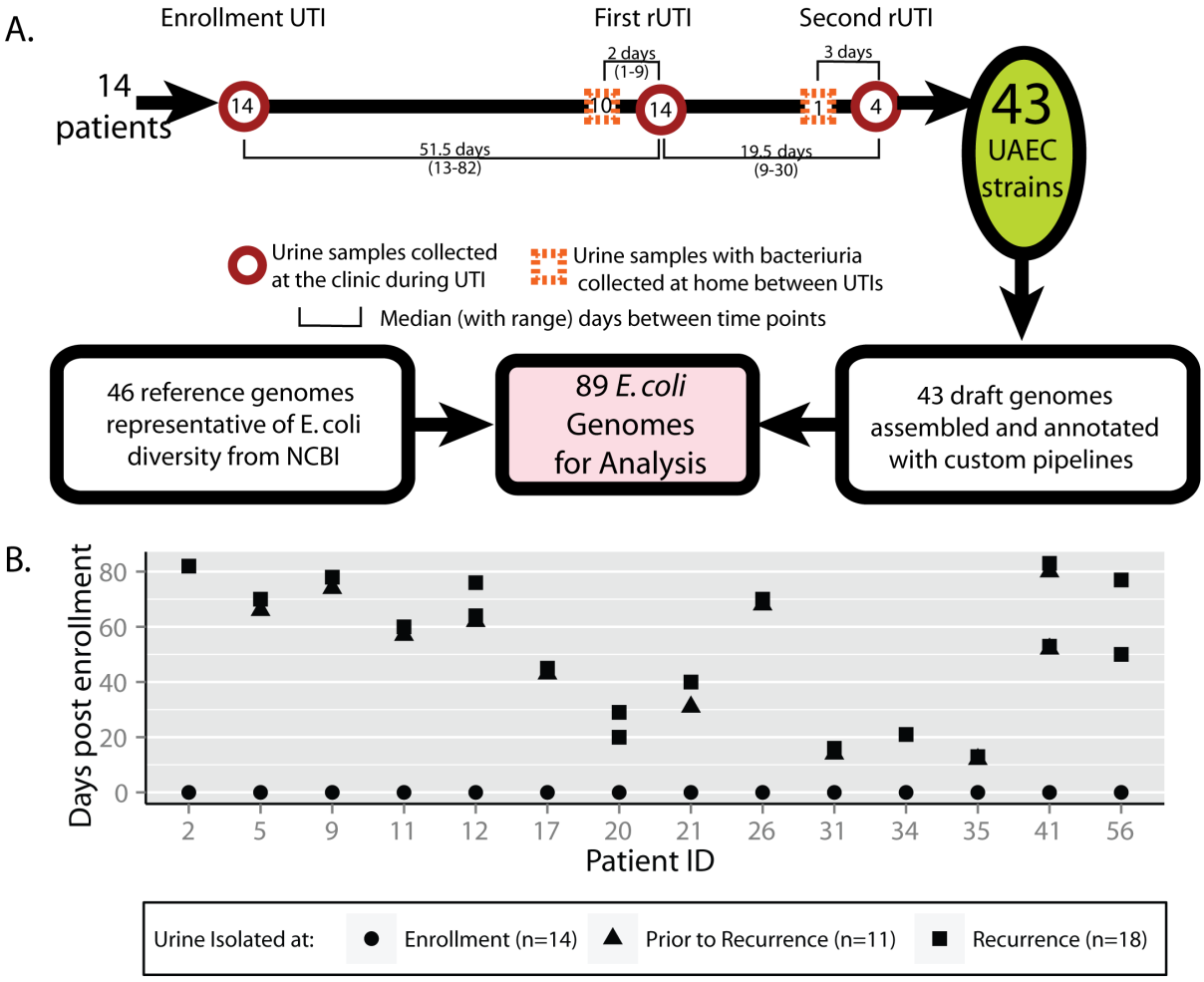
Isolation Time Point	Number of Isolates
Enrollment*	14
rUTI*	18
Prior to recurrence**	11
Total	43

* Diagnosed UTI

** No UTI diagnosis

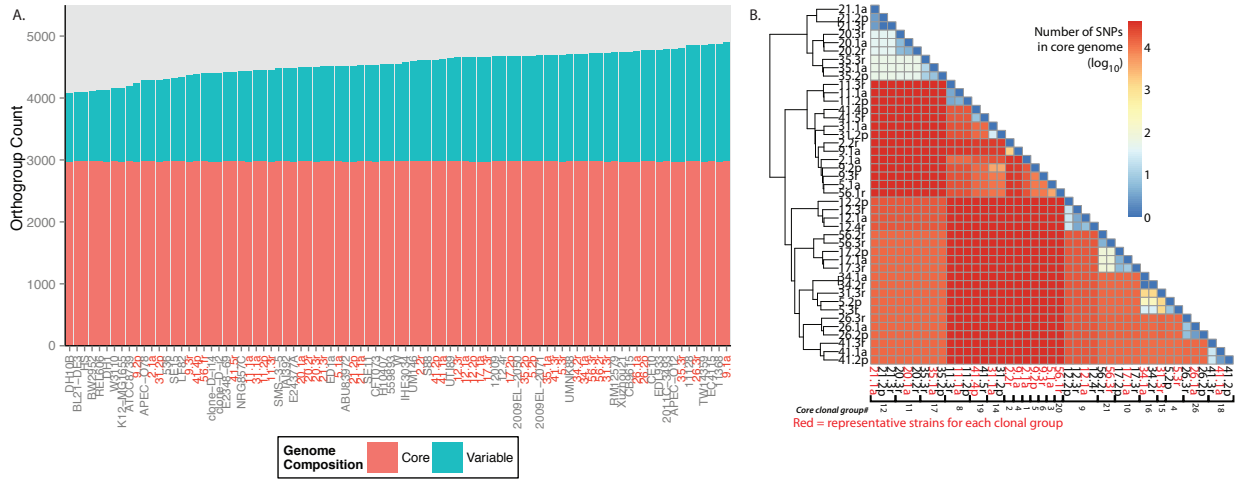
2.8 Supplemental Figures

2.8.1 Figure S1. Sample collection and patient time line.



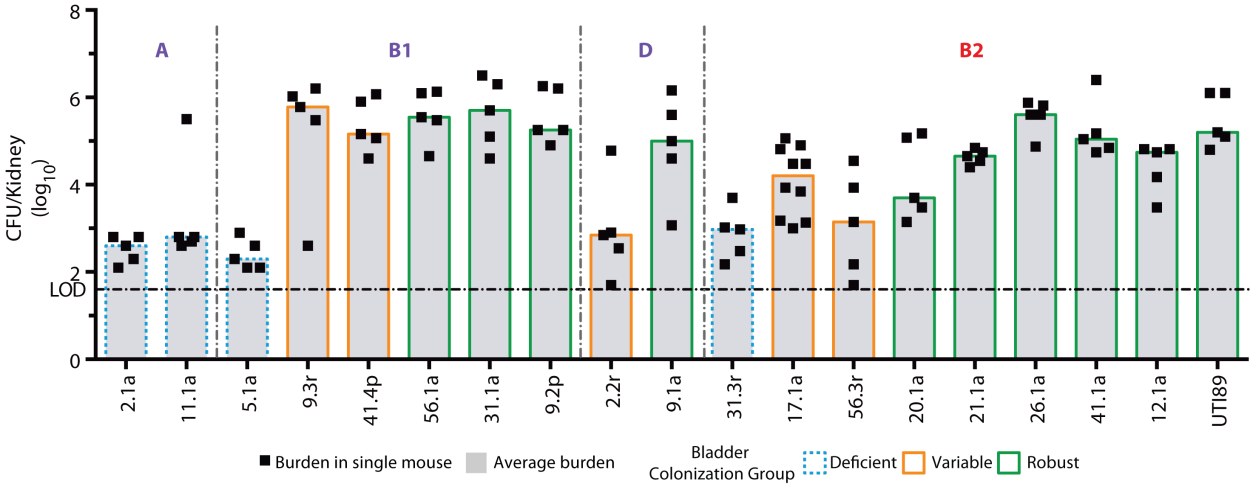
(A) In a previous study, urine samples were collected from a cohort of 104 women with frequent rUTI (defined as ≥3 UTIs in the previous year) and *E. coli* was isolated . A subset of 29 patients had rUTI events during the study. We examined the first 14 of these patients, resulting in 43 isolates collected during enrollment (a), rUTI episodes (r) and prior to recurrence (p). These isolates were subjected to whole genome sequencing and added to a collection of 46 reference genomes that were representative of the rest of the *E. coli* phylogeny for subsequent analysis. Red boxes and orange circles represent samples collected at the clinic or at home, respectively, and the numbers inside the shapes indicate the number of samples collected during those different time points. Brackets under the time line indicate the average days between sample collections. (B) Sample collection dates with positive *E. coli* cultures are plotted from days since the enrollment UTI (circles) to recurrences with diagnosed UTIs (squares). Positive *E. coli* samples collected prior to recurrence were collected at home in the days prior to a UTI diagnosis (triangles). The timing of rUTI relative to the date of enrollment varied for each patient. After treatment for an initial infection, all patients had seven days with no bacteriuria.

2.8.2 Figure S2. Measurement of UAEC gene carriage and nucleotide diversity in core genome.



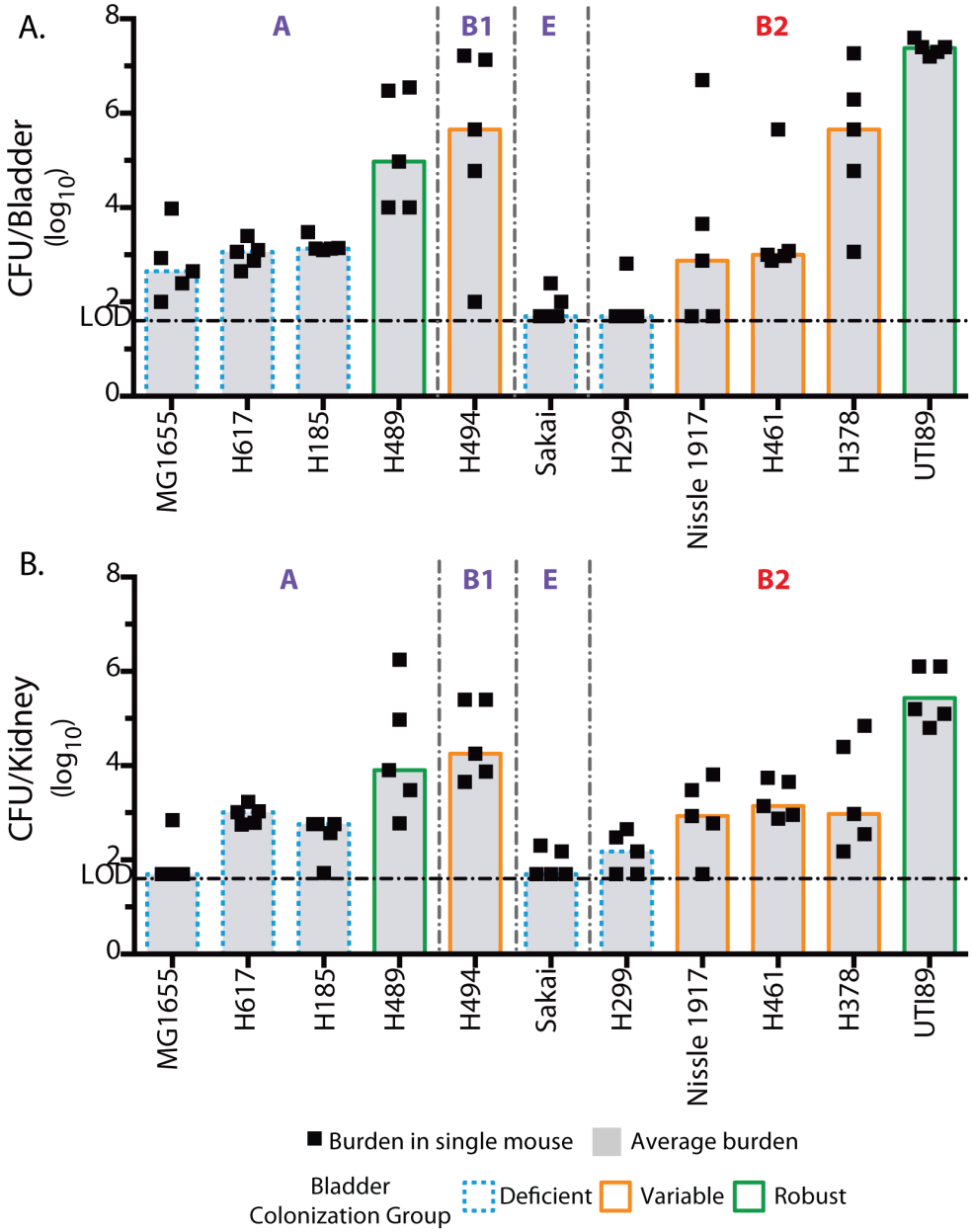
(A) Orthogroup clustering (see Methods) was used to identify the distribution of orthologous genes in reference *E. coli* strains (gray text) and UAEC sequenced as part of this study (red text). Core genes (salmon-pink bars) were defined as orthologous groups that were present in every strain in the data set and comprised 50-75% of each strain's genome. Variable genes (turquoise bars) were defined as orthologous groups that were present in one or more strains, but not in all strains. (B) Pairwise alignment of all 43 UAEC isolates was performed using MUSCLE and the number of SNPs from each alignment was used to hierarchically cluster (see the dendrogram on the y-axis) the strains into 21 "core clonal groups" (black numbers below the x-axis) of 1-4 isolates each. Representatives (in red text) were selected from each core clonal group by the quality of their draft genome.

2.8.3 Figure S3. Kidney colonization by B2 and non-B2 UAEC.



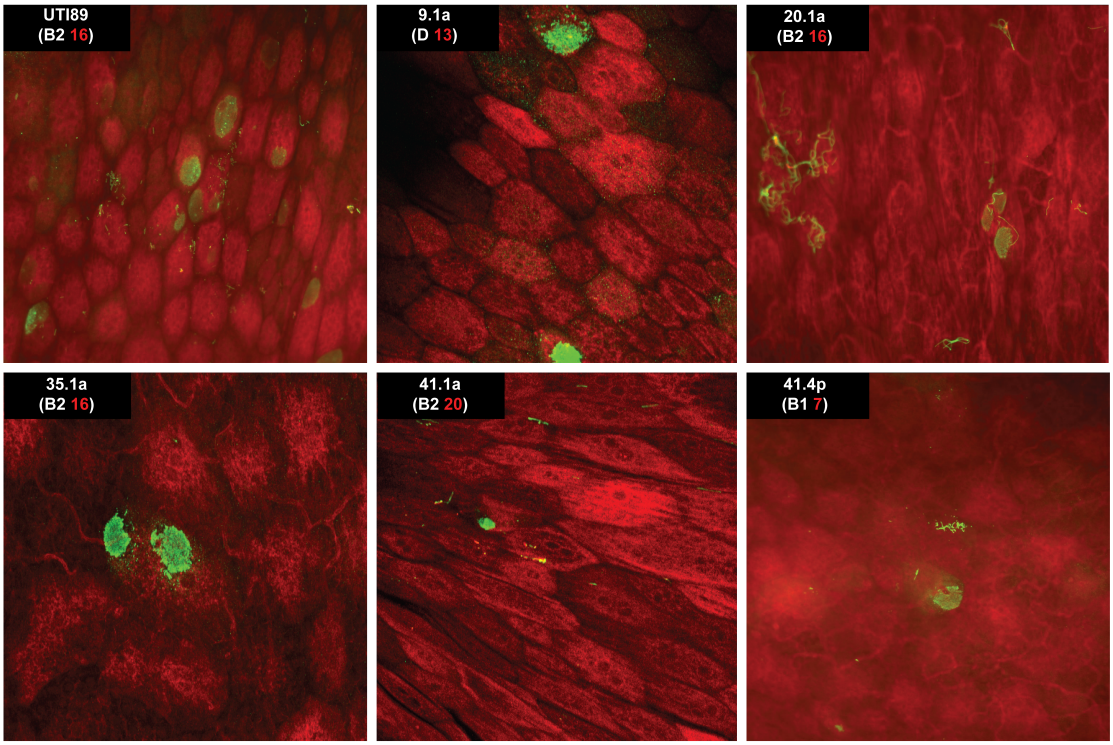
The colonization efficiencies of representative UAEC and UTI89 were tested in C3H/HeN mice. Bacteria were enumerated from harvested kidneys at 24 hpi (black boxes). Kidney colonization largely mirrored bladder burden at 24 hpi in UAEC strains. Strains were separated by clade membership (colored text at the top), ordered by ascending mean bladder bacterial burden at 24 hpi (Fig 3A) and defined as Deficient (dotted, blue outline), Variable (solid, orange outline) and Robust (solid, green outline) colonizers based on their bladder colonization phenotypes. The broken line represents the limit of detection of bacteria. Data from individual mouse inoculations are shown in squares and the gray bar indicates the median values.

2.8.4 Figure S4. Gut-associated *E. coli* are poor colonizers in the C3H/HeN mouse models of UTI.



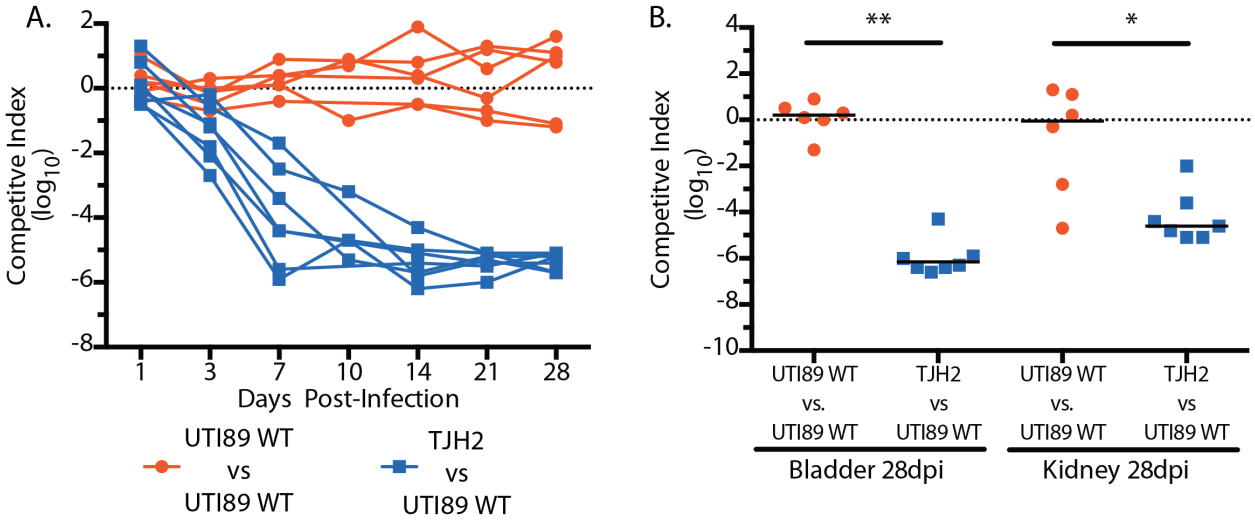
A total of 10 gut-associated *E. coli* that are not associated with urinary disease were tested for their colonization efficiency in C3H/HeN mice. UTI89 was included to represent robust colonizer *E. coli* strains. Bacteria were enumerated from individual harvested (A) bladders and (B) kidneys at 24 hpi (black boxes). Strains were separated by clade membership (colored text at the top), and ordered by their median colonization of the bladder. Each strain was categorized as "deficient", "variable", or "robust" based on the strain median colonization of the bladder using the same parameters as for the UAEC strains (Fig. 3a). The broken line represents the limit of detection of bacteria. Data from individual mouse inoculations are shown in squares and the gray bar indicates the median values.

2.8.5 Figure S5. IBC formation by B2 and non-B2 UAEC.



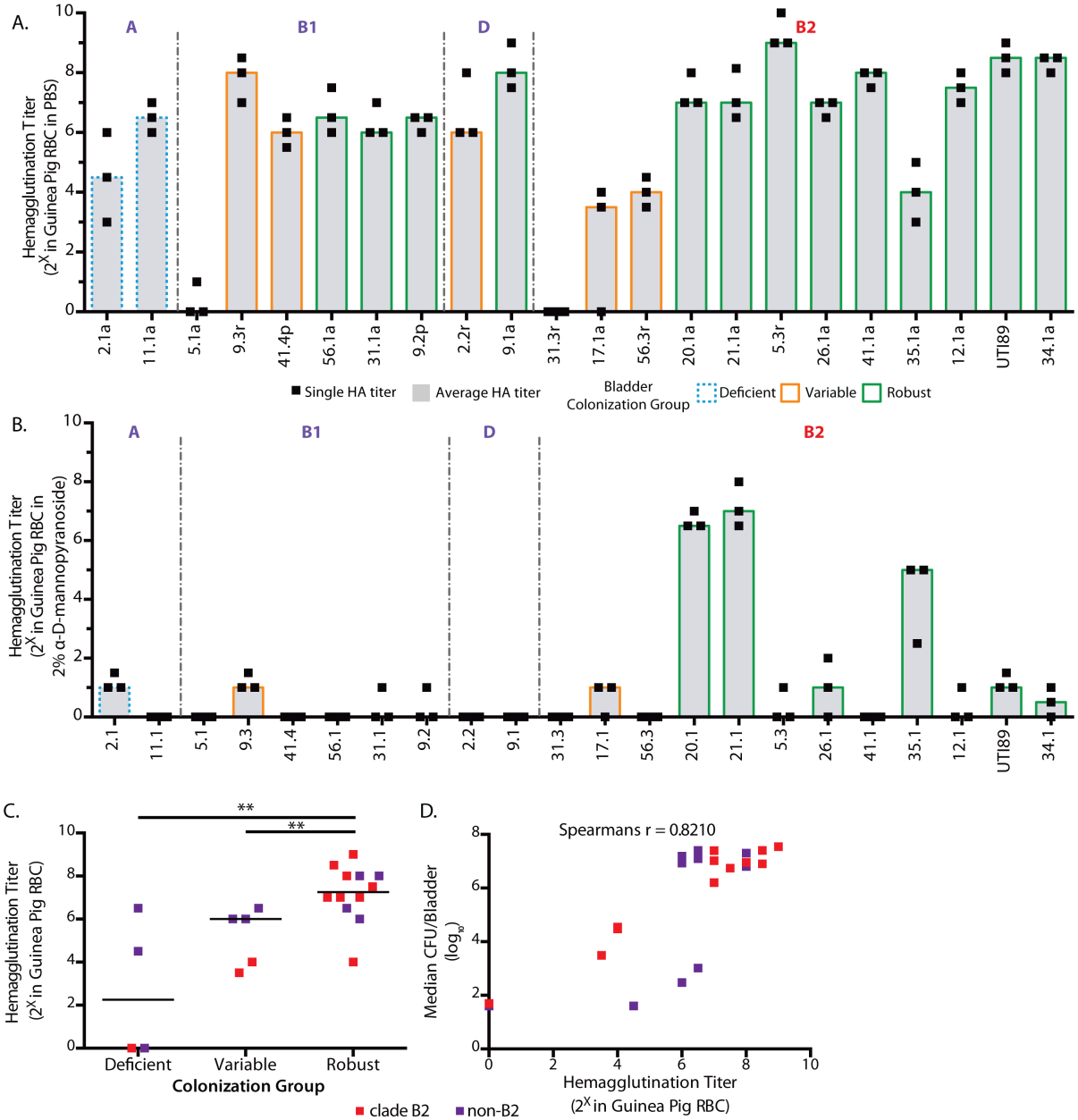
A total of 7 UAEC strains were inoculated into 3-5 C3H/HeN mice each. At 6 hpi bladders were harvested, fixed, stained with an α-*E. coli* antibody and counter-stained with wheat-germ agglutinin. Of the 7 strains tested, only strain 11.1a failed to produce IBCs in any mouse at this time point while the other non-B2 and B2 UAEC were competent for IBC formations.

2.8.6 Figure S6. Carriage of PAI II_{UTI89} enhances competitive fitness in chronic UTI in C3H/HeN mice.



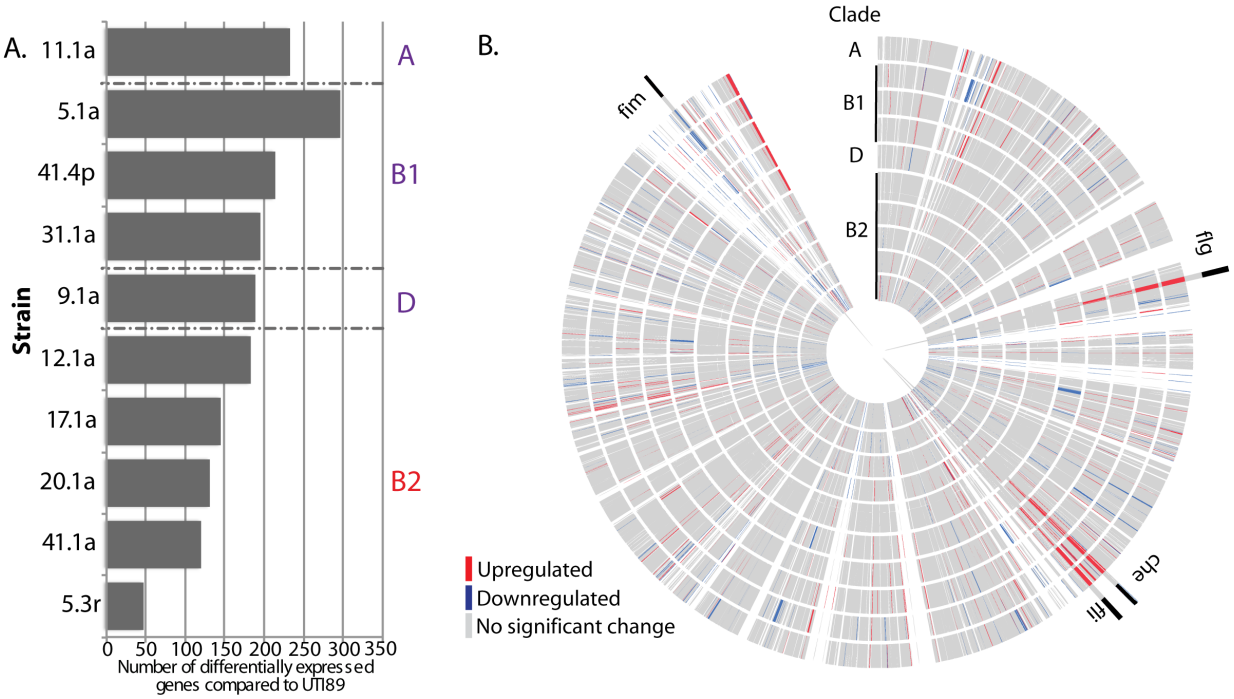
(A) TJH2 (UTI89ΔPAI II_{UTI89}), which lacks the PUF genes *papGIII, hek, cnf1,* and *hlyA*, was co-infected with UTI89 and bacterial titers were enumerated from the urine at indicated days post-infection (dpi). Competitive indexes (CI) were calculated as the ratio of UTI89 (red) or TJH2 (blue) over UTI89, normalized to the ratio of the actual inoculum titers. UTI89 WT was able to outcompete TJH2 at chronic time points (≥7 dpi) but did not display competitive advantages during acute time points (≤3 dpi) (B) After 28 dpi, bladders and kidneys were harvested and bacteria were enumerated to calculate CIs as above. UTI89 WT was able to dramatically outcompete TJH2 as measured by Mann-Whitney U test with **, $P<0.01$.

2.8.7 Figure S7. Hemagglutination is correlated to colonization efficiency in both B2 and non-B2 UAEC.



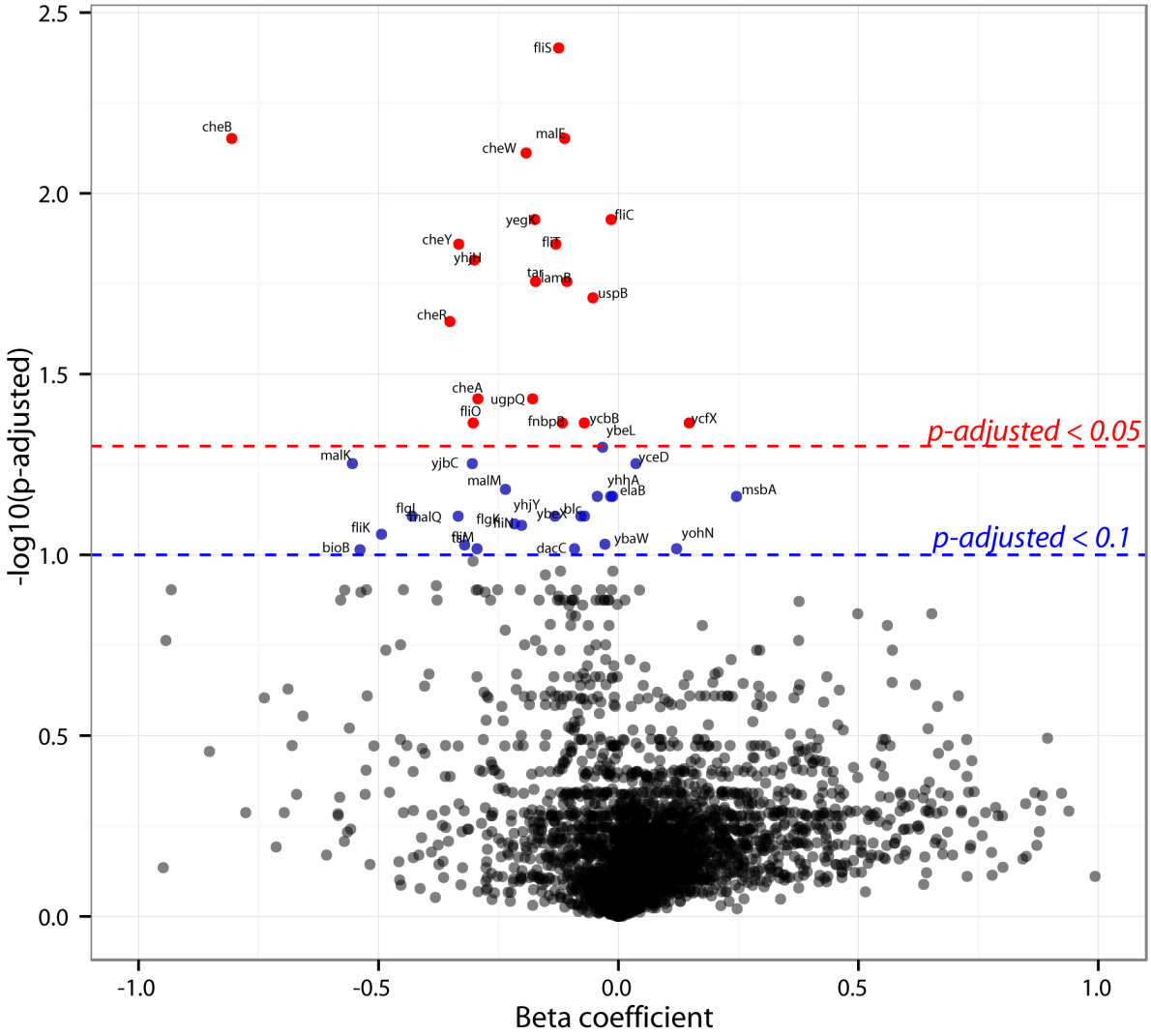
Hemagglutination (HA) titers, which are a measure of UPEC piliation, including type 1 and S pili, were calculated for each representative UAEC strain in the indicated clade using guinea pig erythrocytes in PBS (A) or 2% methyl α-D-mannopyranoside (B). Data presented represent median values (gray bars) derived from three independent experiments (black squares). Strains were categorized by clade (colored text) and ordered by mean bladder burden at 24 hpi (Fig 3a) into deficient (blue, dotted outline), variable (orange, solid outline), and robust (green, solid outline) colonizers. (C) Robust colonizers have significantly higher median HA titers than either variable or deficient colonizers, as measured by Mann-Whitney U test with **, $P<0.01$. Horizontal bars indicate median values for each group and colors indicate either B2 (red) or non-B2 strains (purple) (D) Median HA titer is significantly correlated to bladder burden at 24 hpi in C3H/HeN mice, as measured by Spearman's rank correlation (ρ statistic indicated at the top). Squares indicate individual B2 (red) or non-B2 (purple) UAEC.

2.8.8 Figure S8. Differentially expressed genes in UAEC relative to UTI89.



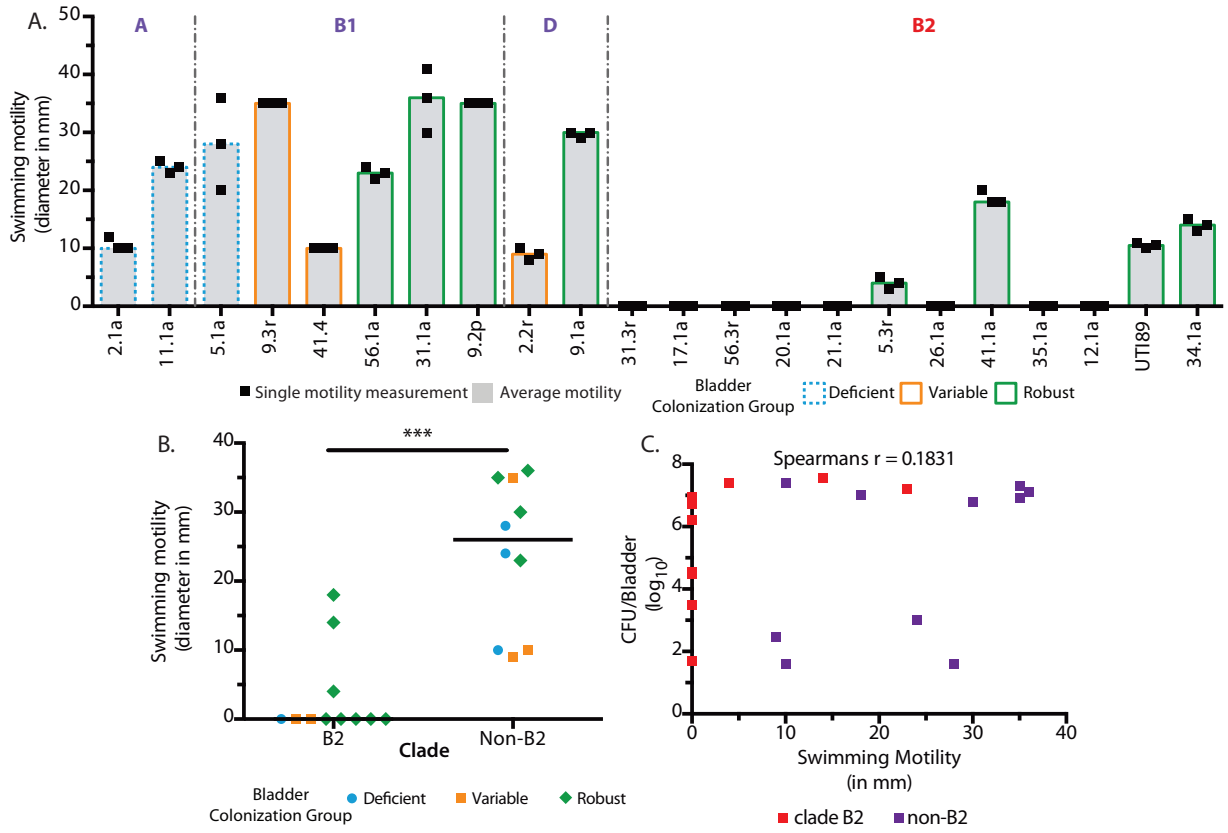
RNA-Seq was performed on a subset of UAEC grown under type 1 pilus-inducing culture conditions used for mouse inoculation. RNA-Seq data were then compared to UTI89 to identify changes in the expression of the 3,516 core genes shared between UAEC and UTI89. (A) Total number of differentially expressed core genes relative to UTI89. Strains are separated by clade (colored text) and ordered by increasing number of differentially expressed genes relative to UTI89. (B) Each tick mark represents a core gene in the order of the UTI89 genome. Each concentric circle represents an UAEC strain and is ordered the same as panel A (strain 5.3r is the centermost and 11.1a is the outermost). Genes are colored based on their expression level relative to their ortholog in UTI89, with gray indicating no significant difference, red indicating significant upregulation and blue indicating significant downregulation. Genes in the *fim, fli, che* and *flg* operons, among others, were differentially expressed and are indicated. Data presented represent averages derived from three independent experiments that were corrected for variance in read depth between samples and false discovery (P-adjusted <0.05, fold-change >3).

2.8.9 Figure S9. Expression of core genes in type 1-inducing culture conditions correlates to bladder colonization in UAEC.



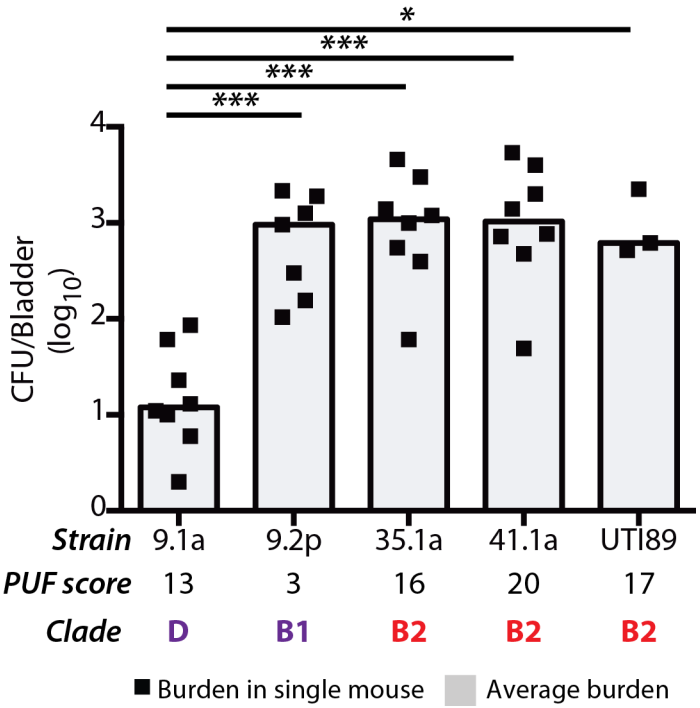
The linear regression curve for each gene (dot) is plotted according to its beta coefficient (x-axis) and adjusted p-value (y-axis). A total of 42 genes were identified whose expression was significantly correlated (blue dots, *p-adjusted*<0.1; red dots, *p-adjusted*<0.05) with bladder colonization at 24 hpi including genes associated with swimming motility and maltose utilization. The majority of significantly correlated genes had negative beta coefficients, indicating a negative correlation between bladder colonization and gene expression.

2.8.10 Figure S10. Motility is not correlated to colonization efficiency in either B2 or non-B2 UAEC.



(A) Swimming motility in soft agar was measured for each representative UAEC strain after growth in type 1-inducing conditions. Strains were categorized by clade (colored text) and ordered within clades by mean bladder burden at 24 hpi (Fig. 3a) showing deficient (blue, dotted outline), variable (orange, solid outline), and robust (green, solid outline) colonizers. Data represents median values (gray bars) derived from three independent experiments (black squares) (B) Clade B2 UAEC exhibited significantly less motility after growth in type 1-inducing conditions relative to non-B2 strains, as measured by Mann-Whitney U test with ***, *P<0.001* (C) Swimming motility is not significantly correlated to bladder burden at 24 hpi, as measured by Spearman's rank correlation (ρ statistic indicated at the top). Squares indicate individual B2 (red) or non-B2 (purple) UAEC. Data presented represent means derived from three independent experiments.

2.8.11 Figure S11. Both B2 and non-B2 UAEC are capable of forming persistent bladder reservoirs in C57BL/6 mice.



UTI89 and a subset of 4 UAEC strains with indicated PUF scores and clade membership were tested for their ability to form persistent bladder reservoirs in the absence of bacteriuria and kidney colonization. Bacteria were enumerated at 14 days post-infection (dpi). Individual mouse burdens (black squares) were used to calculate a median value of bladder burden (gray bar). Strain 9.1a was less able to persist in the bladder than other UAEC strains as determined by Mann-Whitney U test with *, *P<0.05; ***, P<0.001.*

2.9 Supplementary Tables

2.9.1 Table S1. Clinical information of enrolled patients.

Patient ID	Patient Age	Age of First UTI	Mother, UTI History†	Sister, UTI History†	Race	Hispanic Ethnicity	Number of Treated UTIs (Previous Year)	BCM Used in Previous Month††
2	25 or Older	17	No	Yes	White	No	2	OCP
5	18-24	20	Yes	N/A	Asian	No	2	OCP, Condoms
9	18-24	22	Unknown	Unknown	Other	No	1	OCP
11	25 or Older	23	No	Unknown	White	No	3	OCP
12	18-24	13	Unkown	Unknown	White	No	2	OCP, Condoms
17	18-24	19	No	No	Asian	No	1	Spermicide-coated Condoms
20	18-24	6	Yes	No	White	No	3	Condoms
21	18-24	18	Yes	N/A	Asian	No	4	OCP
26	18-24	13	Unknown	No	White	No	1	Spermicide-coated Condoms
31	18-24	14	No	No	White	No	5	OCP
34	18-24	17	Unknown	No	Other	No	1	OCP
35	18-24	19	No	Yes	White	No	1	Spermicide-coated Condoms
41	18-24	16	No	No	White	No	3	OCP, Condoms
56	25 or Older	18	Unknown	Yes	White	No	3	OCP
† Family history of UTI was reported by the patient, if known. N/A indicates patients did not have any sisters.								
†† Birth control methods including oral contraceptive pills								

2.9.2 Table S2. Characteristics of UAEC Isolates.

Strain ID†	Patient ID	Date Collected	Sample Time Point††	Clade	Core Clonal Group	Representative Strain
2.1a	2	2/5/2007	Enrollment UTI	A	1	*
2.2r	2	4/28/2007	Different-strain rUTI	D	2	*
5.1a	5	3/4/2007	Enrollment UTI	B1	3	*
5.2p	5	5/9/2007	Prior to Recurrence	B2	4	
5.3r	5	5/13/2007	Different-strain rUTI	B2	4	*
9.1a	9	3/19/2007	Enrollment UTI	D	5	*
9.2p	9	6/1/2007	Prior to Recurrence	B1	6	*
9.3r	9	6/5/2007	Different-strain rUTI	B1	7	*
11.1a	11	4/1/2007	Enrollment UTI	A	8	*
11.2p	11	5/28/2007	Prior to Recurrence	A	8	
11.3r	11	5/31/2007	Same-strain rUTI	A	8	
12.1a	12	4/16/2007	Enrollment UTI	B2	9	*
12.2p	12	6/17/2007	Prior to Recurrence	B2	9	
12.3r	12	6/19/2007	Same-strain rUTI	B2	9	
12.4r	12	7/1/2007	Same-strain rUTI	B2	9	
17.1a	17	4/29/2007	Enrollment UTI	B2	10	*
17.2p	17	6/11/2007	Prior to Recurrence	B2	10	
17.3r	17	6/13/2007	Same-strain rUTI	B2	10	
20.1a	20	5/1/2007	Enrollment UTI	B2	11	*
20.2r	20	5/21/2007	Same-strain rUTI	B2	11	
20.3r	20	5/30/2007	Same-strain rUTI	B2	11	
21.1a	21	5/2/2007	Enrollment UTI	B2	12	*
21.2p	21	6/2/2007	Prior to Recurrence	B2	12	
21.3r	21	6/11/2007	Same-strain rUTI	B2	12	
26.1a	26	6/5/2007	Enrollment UTI	B2	13	*
26.2p	26	8/12/2007	Prior to Recurrence	B2	13	
26.3r	26	8/14/2007	Same-strain rUTI	B2	13	

31.1a	31	7/23/2007	Enrollment UTI	B1	14	*
31.2p	31	8/6/2007	Prior to Recurrence	B1	14	
31.3r	31	8/8/2007	Different-strain rUTI	B2	15	*
34.1a	34	10/21/2007	Enrollment UTI	B2	16	*
34.2r	34	11/11/2007	Same-strain rUTI	B2	16	
35.1a	35	10/22/2007	Enrollment UTI	B2	17	*
35.2p	35	11/3/2007	Prior to Recurrence	B2	17	
35.3r	35	11/4/2007	Same-strain rUTI	B2	17	
41.1a	41	1/21/2008	Enrollment UTI	B2	18	*
41.2p	41	3/13/2008	Prior to Recurrence	B2	18	
41.3r	41	3/14/2008	Same-strain rUTI	B2	18	
41.4p	41	4/10/2008	Prior to Recurrence	B1	19	*
41.5r	41	4/13/2008	Different-strain rUTI	B1	19	
56.1a	56	7/28/2008	Enrollment UTI	B1	20	*
56.2r	56	9/16/2008	Different-strain rUTI	B2	21	
56.3r	56	10/13/2008	Same-strain rUTI	B2	21	*

† The suffix "a" indicates isolate collection at the enrollment UTI, "p" indicates collection prior to recurrence and "r" indicates collection during recurrent UTIs

†† UTI and rUTI timepoints represent symptomatic events while "prior to recurrence" isolates were collected before a UTI diagnosis.

2.9.3 Table S3. Reference *Escherichia coli* Strains.

Strain	Pathotype	Clade	BioProject (NCBI ID)
536	UPEC	B2	PRJNA16235
11128	EHEC	B1	PRJDA32513
11368	EHEC	B1	PRJDA32509
12009	EHEC	B1	PRJDA32511
55989	EAEC	B1	PRJNA33413
2009EL-2050	EHEC	B1	PRJNA81097
2009EL-2071	EHEC	B1	PRJNA81099
2011C-3493	EHEC	B1	PRJNA81095
ABU 83972	ABU	B2	PRJNA38725
APEC O1	APEC	B2	PRJNA16718
APEC O78	APEC	B1	PRJNA184588
ATCC 8739	Commensal	A	PRJNA18083
BL21(DE3)	Lab Strain	A	PRJNA20713
BW2952	Lab Strain	A	PRJNA33775
CB9615	EPEC	E	PRJNA42729
CE10	NMEC	F	PRJNA63597
CFT073	UPEC	B2	PRJNA313
clone D i14	UPEC	B2	PRJNA52023
clone D i2	UPEC	B2	PRJNA52021
DH1	Lab Strain	A	PRJDA52077
DH10B	Lab Strain	A	PRJNA20079
E2348/69	EPEC	B2	PRJEA32571
E24377A	ETEC	B1	PRJNA13960

EC4115	EHEC	E	PRJNA27739
ED1a	Commensal	B2	PRJNA33409
EDL933	EHEC	E	PRJNA259
H10407	ETEC	A	PRJEA42749
HS	Commensal	A	PRJNA13959
IAI39	UPEC	F	PRJNA33411
IHE3034	NMEC	B2	PRJNA43693
K12 MG1655	Commensal	A	PRJNA40075
LF82	AIEC	B2	PRJNA33825
NRG 857C	AIEC	B2	PRJNA41221
REL606	Lab Strain	A	PRJNA18281
RM12579	EPEC	E	PRJNA68245
S88	Commensal	B2	PRJNA33375
SE11	Commensal	B1	PRJNA18057
SE15	Commensal	B2	PRJDA19053
SMS-3-5	Environmental	F	PRJNA19469
TW14359	EHEC	E	PRJNA30045
UM146	AIEC	B2	PRJNA50883
UMNK88	ETEC	A	PRJNA42137
UTI89	UPEC	B2	PRJNA16259
W	Lab Strain	B1	PRJNA48011
W3110	Lab Strain	A	PRJNA16351
Xuzhou21	EHEC	E	PRJNA45823

2.9.4 Table S4. List of putative urovirulence factors

PUF	Category	Source Genome	NCBI ID	Putative Function	Reference
bmaE	Adhesin	2009C-3133	GI:944399742	Blood group M–specific adhesin	Johnson et al. 2001
cdtB	Toxin	APECO1	GI:115512768	Cytolethal distending toxin	Johnson et al. 2001
chuA	Iron-Uptake	UTI89	GI:91213002	Heme binding outer membrane	Johnson et al. 2015
cnf1	Toxin	UTI89	GI:91213869	Cytotoxic necrotizing factor 1	Johnson et al. 2001
cvaC	Misc	pECOS88	GI:218349786	Colicin V	Johnson et al. 2001
draB	Adhesin	EC958	GI:641686625	Dr antigen-specific adhesin	Johnson et al. 2015
feoB	Iron-Uptake	UTI89	GI:91212884	Ferrous iron transport	Luo et al. 2012
fimH	Adhesin	UTI89	GI:91213965	D-mannose–specific adhesin	Johnson et al. 2001
focH	Adhesin	CFT073	GI:26107519	F1C pili adhesin (marker for *foc* operon)	This study
fyuA	Iron-Uptake	UTI89	GI:91211204	Yersinia siderophore receptor	Johnson et al. 2001
hek	Adhesin	UTI89	GI:91213830	Auto-aggregating adhesin and invasin	Bateman et al. 2013
hlyA	Toxin	UTI89	GI:91213874	a-Hemolysin	Johnson et al. 2001
ibeA	Adhesin	UTI89	GI:91213983	Invasion of brain endothelium	Johnson et al. 2001
iha	Adhesin	CFT073	GI:26109853	Nonhemagglutinin adhesin	Johnson et al. 2001
ireA	Iron-Uptake	CFT073	GI:26111414	Iron-regulated element (siderophore receptor)	Johnson et al. 2015
iroN	Iron-Uptake	UTI89	GI:91210147	Salmochelin siderophore receptor	Johnson et al. 2001
irp2	Iron-Uptake	UTI89	GI:91211199	Yersinia biosynthetic protein	Luo et al. 2012
iucC	Iron-Uptake	CFT073	GI:26109868	Aerobactin biosynthetic protein	Luo et al. 2012

iutA	Iron-Uptake	CFT073	GI:26109866	Ferric aerobactin receptor	Johnson et al. 2001
malX	Misc	UTI89	GI:91212209	Marker of pathogenicity-associated island	Johnson et al. 2001
ompT	Protectin	UTI89	GI:91209611	Outer membrane protein T (protease)	Johnson et al. 2001
papGII	Adhesin	CFT073	GI:26109826	P pili adhesin - Pyelonephritis-associated papG variant	Johnson et al. 2001
papGIII	Adhesin	UTI89	GI:91213835	P pili adhesin - Cystitis-associated papG variant	Johnson et al. 2001
picU	Toxin	Nissle-1917	GI:660510468	Protein involved in Intestinal Colonization - serine protease	Johnson et al. 2015
sat	Toxin	CFT073	GI:26109862	Secreted Autotransporter Toxin	Johnson et al. 2015
sfaH	Adhesin	UTI89	GI:91210144	S pili adhesin (marker for *sfa* operon)	This study
sitA	Toxin	UTI89	GI:91210366	Periplasmic iron transport protein	Luo et al. 2012
traT	Protectin	pUTI89	GI:91075831	Temperature-sensitive hemagglutinin - serine protease	Johnson et al. 2015
tsh	Toxin	pVM01	GI:168831012	Surface exclusion	Johnson et al. 2001
usp	Misc	Nissle-1917	GI:660510280	Uropathogenic-Specific Protein (bacteriocin)	Johnson et al. 2015
vat	Toxin	Nissle-1917	GI:660510497	Vacolating autotransporter - serine protease	Johnson et al. 2015

2.9.5 Table S5. Gut associated *E. coli* Strain Characteristics.

Strain	Pathotype	Clade	Source
MG1655	Commensal/Lab Strain	A	Blattner et al. 1997
Nissle 1917	Commensal	B2	Grozdanov et al. 2004
Sakai	EHEC	E	Hayashi et al. 2001
H185	Commensal	A	Gordon et al. 2005
H299	Commensal	B2	Gordon et al. 2005
H378	Commensal	B2	Gordon et al. 2005
H461	Commensal	B2	Gordon et al. 2005
H489	Commensal	A	Gordon et al. 2005
H494	Commensal	B1	Gordon et al. 2005
H617	Commensal	A	Gordon et al. 2005

2.9.6 Table S6. IBC Formation in Select UAEC at 6 hpi in C3H/HeN mice.

Strain	Clade	PUF Score	IBC Formation
11.1a	A	4	-
41.4p	B1	7	++
9.1a	D	13	+
20.1a	B2	16	++
35.1a	B2	16	++
41.1a	B2	20	+
UTI89	B2	17	+++

- = No IBC formed
+ = Low rate of IBC formation
++ = Moderate rate of IBC formation
+++ = High rate of IBC formation

2.9.7 Table S7. Presence of *brnAT* genes in Robust and Deficient colonizer UAEC strains.

Colonizer Type	Strain	pfam14384 brnA	pfam04365 brnT
Deficient	11.1a	Present	Present
Robust	12.1a	Present	Present
Robust	20.1a	Absent	Absent
Robust	21.1a	Absent	Absent
Robust	26.1a	Present	Present
Deficient	2.1a	Present	Present
Robust	31.1a	Absent	Absent
Deficient	31.3r	Present	Present
Robust	34.1a	Absent	Absent
Robust	35.1a	Absent	Absent
Robust	41.1a	Absent	Absent
Robust	56.1r	Absent	Absent
Deficient	5.1a	Present	Present
Robust	5.3r	Absent	Absent
Robust	9.1a	Absent	Absent
Robust	9.2p	Absent	Absent
Robust	UTI89	Absent	Absent
Total Robust		**2/13 strains**	**2/13 strains**
Total Deficient		**4/4 strains**	**4/4 strains**

2.9.8 Table S8. Reads mapped to core genome of UAEC.

Sample	Clade Colonization group	Total Reads	Reads Aligned to Core Genome	Proportion of Reads Aligned to Core Genome
5.1a_A	A	1.37E+07	7.04E+06	51.34%
5.1a_B	Deficient	7.26E+06	3.72E+06	51.25%
5.1a_C		1.05E+07	5.53E+06	52.85%
5.3r_A	B2	2.13E+07	1.00E+07	47.05%
5.3r_B	Robust	1.98E+07	8.75E+06	44.22%
5.3r_C		1.54E+07	8.54E+06	55.55%
9.1a_A	D	1.56E+07	6.55E+06	41.95%
9.1a_B	Robust	2.14E+07	9.36E+06	43.72%
9.1a_C		3.89E+07	1.74E+07	44.86%
11.1a_A	A	3.02E+07	1.73E+07	57.53%
11.1a_B	Deficient	1.73E+07	1.02E+07	58.59%
11.1a_C		2.80E+07	1.53E+07	54.73%
12.1a_A	B2	2.73E+07	1.65E+07	60.54%
12.1a_B	Robust	3.30E+07	2.00E+07	60.74%
12.1a_C		1.75E+07	1.04E+07	59.12%
41.1a_A	B2	2.19E+07	1.08E+07	49.26%
41.1a_B	Robust	1.09E+07	5.34E+06	49.00%
41.1a_C		3.25E+07	1.69E+07	52.08%
17.1a_A	B2	5.35E+07	3.11E+07	58.19%
17.1a_B	Variable	2.06E+07	1.20E+07	58.15%
17.1a_C		1.89E+07	1.09E+07	57.67%
41.4p_A	B1	2.49E+07	1.47E+07	58.97%
41.4p_B	Variable	1.79E+07	1.10E+07	61.19%
41.4p_C		6.53E+07	3.49E+07	53.39%
20.1a_A	B2	2.23E+07	1.38E+07	61.73%
20.1a_B	Robust	3.65E+07	1.99E+07	54.63%
20.1a_C		4.19E+07	2.40E+07	57.22%
31.1a_A	B1	2.08E+07	1.25E+07	60.02%
31.1a_B	Robust	1.91E+07	1.00E+07	52.56%
31.1a_C		1.93E+07	9.31E+06	48.15%
UTI89A	B2	1.30E+07	6.99E+06	53.73%
UTI89B	Robust	5.24E+06	2.79E+06	53.16%
UTI89C		6.80E+06	3.66E+06	53.80%

2.9.9 Table S9. DEseq Analysis of UAEC transcription of core genome.

Available in Appendix A

2.9.10 Table S10. Expression of PUF genes in select UAEC.

PUF Gene	5.1a	5.3r	9.1a	11.1a	12.1a	41.1a	17.1a	41.4p	21.1a	31.1a	UTI89
sfaH	0	68.7182	0	0	0	281.967	0	0	415.875	0	52.437
iha	0	0	233.727	0	0	0	69.9788	0	0	0	0
traT	0	712.752	0	0	1922.51	1296.68	1049.21	0	0	0	792.094
hek	0	0.0731843	0	50.5309	0	2883.04	832.19	0	1195.47	0	765.862
papGII	0	0	15.9602	0	0	0	380.251	0	0	0	0
iucC	0	0	145.253	0	233.068	0	63.2812	0	0	0	0
cvaC	0	0	0	0	157.803	0	0	150.128	0	0	0
cnf1	0	0.393869	0	0	0	442.011	545.493	0	220.622	0	216.935
iutA	0	0	186.706	0	371.568	0	77.1131	0	0	0	0
feoB	242.288	293.207	111.426	225.914	344.52	219.492	197.9	212.194	384.731	242.847	278.868
papGIII	0	0.142911	0	0	0	104.9	226.243	0	111.103	0	40.5398
focH	0	0	0	0	0	0	0	0	0	0	0
tsh	0	521.981	0	31.4803	0	368.807	132.812	0	199.607	0	85.0636
hlyA	0	0.31254	0	0	0	2102.95	832.367	0	731.642	0	805.574
ibeA	0	13.4008	0	0	0	21.2063	0	0	0	0	5.83029
iroN	0	4248.55	0	10.4705	371.397	4225.54	0	1011.09	2481.4	0	2640.86
chuA	0	48.2425	19.6964	0	67.6079	0	24.784	0	73.3455	0	30.8678
ireA	0	0	0	0	0	0	621.306	0	0	0	0
sitA	0	228.317	98.0461	0	107.68	29.756	0	0	46.4247	0	37.7494
fyuA	30.8512	92.3986	21.4078	0	19.8853	39.6031	16.2512	33.0504	34.3733	0	57.7845
ompT	0	963.823	123.215	0	110.548	248.457	93.3238	272.355	137.231	350.373	267.572
fimH	0	343.639	164.765	69.3973	140.877	249.284	71.2325	197.281	182.799	184.001	351.813
pic	0	0	0	0	0	225.387	152.25	0	0	0	0
sat	0	0	434.958	0	0	0	253.857	0	0	0	0
usp	0	87.432	0	0	19.6396	83.5182	21.5099	0	58.2775	0	49.1393
irp2	19.9497	126.473	44.0038	0	27.5027	33.8929	15.0818	29.5727	24.6589	0	48.5425

Values are represented as Reads Per Kilobase of transcript per Million mapped reads (RPKM)

2.9.11 Table S11. Linear correlation of core gene expression to bladder colonization.

Available in Appendix A

2.9.12 Table S12. Functions enriched in genes associated with bladder colonization.

Available in Appendix A

2.9.13 Table S13. UAEC clinical data

	White Blood Cells (per mL Urine)		Lipocalin 2 (ng/mL urine)	
Colonization Group	N*	Median	N*	Median
Deficient	5	38	3	13
Variable	6	225	1	23
Robust	18	194.5	9	260

N = number of observations (i.e. tests done on urine samples collected at UTI visits in the cohort)

* = not all urine samples were tested for the presence of white blood cells or lipocalin 2 levels

2.9.14 Table S14. UAEC sequencing and genome assembly results

Strain ID	Reads	Reads Mapped	n50	Largest Scaffold	Post-Trim Length	Read Coverage	Gene Count	Scaffold Count
2.1a	5.72E+06	5.40E+06	9.51E+04	2.51E+05	4.85E+06	111	4603	215
2.2r	5.92E+06	5.39E+06	2.14E+04	6.34E+04	4.78E+06	113	4901	557
5.1a	7.50E+06	7.06E+06	8.31E+04	2.43E+05	5.05E+06	140	4957	281
5.2p	9.43E+06	8.48E+06	3.20E+04	1.27E+05	4.68E+06	181	4380	116
5.3r	6.27E+06	5.21E+06	9.27E+03	3.85E+04	5.09E+06	102	4939	150
9.1a	5.70E+06	4.92E+06	1.08E+04	5.85E+04	5.15E+06	96	5594	1163
9.2p	8.53E+06	7.90E+06	2.15E+05	4.05E+05	4.86E+06	163	4694	304
9.3r	8.18E+06	7.70E+06	9.41E+04	3.67E+05	5.42E+06	142	4600	187
11.1a	6.52E+06	6.14E+06	1.25E+05	3.52E+05	4.80E+06	128	4648	207
11.2p	9.56E+06	8.25E+06	5.89E+04	2.06E+05	4.99E+06	165	4906	673
11.3r	7.38E+06	6.76E+06	3.96E+04	1.15E+05	4.85E+06	139	4691	361
12.1a	7.04E+06	6.65E+06	1.64E+05	3.66E+05	4.99E+06	133	4816	220
12.2p	8.51E+06	7.76E+06	1.77E+04	6.18E+04	5.14E+06	151	5093	504
12.3r	4.77E+06	4.50E+06	1.86E+05	3.66E+05	4.99E+06	90	4825	232
12.4r	5.12E+06	4.83E+06	1.25E+05	2.33E+05	5.00E+06	97	4842	267
17.1a	2.30E+07	2.11E+07	1.50E+05	5.95E+05	5.17E+06	408	4958	227
17.2p	8.36E+06	7.80E+06	1.55E+05	6.23E+05	5.16E+06	151	4957	224
17.3r	4.29E+06	4.08E+06	1.55E+05	4.85E+05	5.16E+06	79	4927	197
20.1a	1.27E+07	1.16E+07	2.12E+05	7.61E+05	4.99E+06	233	4677	174
20.2r	9.04E+06	8.43E+06	2.62E+05	7.61E+05	4.98E+06	169	4674	168
20.3r	8.84E+06	8.35E+06	9.64E+04	2.52E+05	4.98E+06	168	4704	246
21.1a	4.39E+06	4.14E+06	1.44E+05	4.12E+05	4.99E+06	83	4738	196
21.2p	6.03E+06	5.67E+06	1.44E+05	3.26E+05	4.99E+06	114	4729	187
21.3r	2.44E+07	2.24E+07	2.17E+05	7.78E+05	5.00E+06	448	4711	162
26.1a	6.24E+06	5.89E+06	1.53E+05	4.79E+05	5.21E+06	113	5037	193

26.2p	1.22E+07	1.15E+07	1.45E+05	3.66E+05	5.21E+06	222	5044	210
26.3r	4.84E+06	4.53E+06	1.41E+04	6.86E+04	5.21E+06	87	5259	892
31.1a	2.46E+06	2.22E+06	3.90E+03	2.13E+04	4.57E+06	49	5263	2127
31.2p	8.50E+06	7.98E+06	1.89E+05	5.01E+05	4.68E+06	170	4445	124
31.3r	4.07E+06	3.71E+06	6.90E+03	3.22E+04	5.00E+06	74	5302	1486
34.1a	7.07E+06	6.65E+06	1.98E+05	6.58E+05	5.15E+06	129	4956	146
34.2r	6.28E+06	5.89E+06	2.39E+05	6.58E+05	5.15E+06	114	4949	152
35.1a	2.93E+06	2.79E+06	8.92E+04	3.05E+05	5.14E+06	54	4917	215
35.2p	5.50E+06	5.23E+06	1.89E+05	4.34E+05	5.14E+06	102	4886	156
35.3r	3.95E+06	3.63E+06	1.19E+04	4.80E+04	5.13E+06	71	5145	950
41.1a	4.37E+06	4.17E+06	1.44E+05	3.96E+05	5.05E+06	83	4807	177
41.2p	1.09E+07	9.97E+06	2.89E+05	5.14E+05	5.06E+06	197	4827	166
41.3r	5.74E+06	5.28E+06	1.53E+04	5.20E+04	5.05E+06	104	5004	767
41.4p	6.88E+06	6.34E+06	1.37E+05	2.90E+05	4.78E+06	133	4572	164
41.5r	1.90E+06	1.83E+06	3.92E+04	1.38E+05	4.79E+06	38	4630	333
56.1a	3.86E+06	3.66E+06	1.73E+05	2.89E+05	4.84E+06	75	4621	159
56.2r	3.62E+06	3.39E+06	1.69E+05	4.63E+05	5.25E+06	65	5048	257
56.3r	6.74E+06	6.30E+06	1.45E+05	4.62E+05	5.26E+06	120	5055	248

2.10 Acknowledgements

We thank members of the S.J.H. laboratory for their helpful suggestions and Daniel Stoebel for the generous gift of the gut-associated *E. coli* strains.

2.10.1 Funding

H.L.S.IV. was supported in part by a 2013 Monsanto Excellence Fund Graduate Fellowship and a 2012 Lucille P. Markey Pathway award. This project was funded by the National Institutes of Health (NIH), specifically National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant number RO1 DK051406 and Office of Research on Women's Health (ORWH) Specialized Center of Research (SCOR) grant number P50 DK064540 to S.J.H. and National Institute of Allergy and Infectious Disease (NIAID) grant number U19AI110818 to the Broad Institute. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

2.10.2 Competing interests

T.J.H. is a part-time employee at Fimbrion Therapeutics (St. Louis, MO), a company co-founded by T.M.H. and S.J.H. that is developing small molecule therapeutics to treat UTIs caused by UPEC. T.M.H. is an advisory board member with GlaxoSmithKlein (Brentford, United Kingdom), Paratek Pharmaceuticals (Boston, MA), Achaogen Pharmaceuticals (San Francisco, CA), and Ocean Spray (Middleborough, MA). S.J.H. is an advisory board member of Genentech (San Francisco, CA) and Roche Pharmaceuticals (Basel, Switzerland), a consultant for Regeneron (Tarrytown, NY) and Wellspect Healthcare (El Segundo, CA) and a founding member of QureTech Bio (Umea, Sweden).

2.11 References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anderson, G.G., Palermo, J.J., Schilling, J.D., Roth, R., Heuser, J., and Hultgren, S.J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. Science *301*, 105–107.

Barber, A.E., Norton, J.P., Wiles, T.J., and Mulvey, M.A. (2016). Strengths and Limitations of Model Systems for the Study of Urinary Tract Infections and Related Pathologies. Microbiol. Mol. Biol. Rev. *80*, 351–367.

Beg, Q.K., Vazquez, A., Ernst, J., de Menezes, M.A., Bar-Joseph, Z., Barabási, A.-L., and Oltvai, Z.N. (2007). Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc. Natl. Acad. Sci. U.S.a. *104*, 12663–12668.

Bielecki, P., Muthukumarasamy, U., Eckweiler, D., Bielecka, A., Pohl, S., Schanz, A., Niemeyer, U., Oumeraci, T., Neuhoff, von, N., Ghigo, J.M., et al. (2014). In vivo mRNA profiling of uropathogenic *Escherichia coli* from diverse phylogroups reveals common and group-specific gene expression profiles. mBio *5*, e01075–14.

Brzuszkiewicz, E., Brüggemann, H., Liesegang, H., Emmerth, M., Olschläger, T., Nagy, G., Albermann, K., Wagner, C., Buchrieser, C., Emődy, L., et al. (2006). How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. Proc. Natl. Acad. Sci. U.S.a. *103*, 12879–12884.

Carey, A.J., Tan, C.K., Ipe, D.S., Sullivan, M.J., Cripps, A.W., Schembri, M.A., and Ulett, G.C. (2016). Urinary tract infection of mice to model human disease: Practicalities, implications and limitations. Crit. Rev. Microbiol. *42*, 780–799.

Chen, S.L., Hung, C.S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., and Ozersky, P. (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U.S.a. *103*, 5977–5982.

Chen, S.L., Hung, C.S., Pinkner, J.S., Walker, J.N., Cusumano, C.K., Li, Z., Bouckaert, J., Gordon, J.I., and Hultgren, S.J. (2009). 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 *106*, 22439–22444.

Chen, S.L., Wu, M., Henderson, J.P., Hooton, T.M., Hibbing, M.E., Hultgren, S.J., and Gordon, J.I. (2013). Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Science Translational Medicine *5*, 184ra60.

Conover, M.S., Flores-Mireles, A.L., Hibbing, M.E., Dodson, K., and Hultgren, S.J. (2015). Establishment and Characterization of UTI and CAUTI in a Mouse Model. J Vis Exp e52892.

Czaja, C.A., Stamm, W.E., Stapleton, A.E., Roberts, P.L., Hawn, T.R., Scholes, D., Samadpour, M., Hultgren, S.J., and Hooton, T.M. (2009). Prospective Cohort Study of Microbial and Inflammatory Events Immediately Preceding *Escherichia coli* Recurrent Urinary Tract Infection in Women. J. Infect. Dis. *200*, 528–536.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U.S.a. *97*, 6640–6645.

Dobrindt, U., Chowdary, M.G., Krumbholz, G., and Hacker, J. (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. Med. Microbiol. Immunol. *199*, 145–154.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Ejrnaes, K., Stegger, M., Reisner, A., Ferry, S., Monsen, T., Holm, S.E., Lundgren, B., and Frimodt-Moller, N. (2011). Characteristics of *Escherichia coli* causing persistence or relapse of urinary tract infections: Phylogenetic groups, virulence factors and biofilm formation. Virulence *2*, 528–537.

Enright, M.C., and Spratt, B.G. (1999). Multilocus sequence typing. Trends Microbiol. *7*, 482–487.

Eto, D.S., Jones, T.A., Sundsbak, J.L., and Mulvey, M.A. (2007). Integrin-mediated host cell invasion by type 1-piliated uropathogenic *Escherichia coli*. PLoS Pathog *3*, e100.

Ferry, S.A., Holm, S.E., Stenlund, H., Lundholm, R., and Monsen, T.J. (2004). The natural course of uncomplicated lower urinary tract infection in women illustrated by a randomized placebo controlled study. Scand. J. Infect. Dis. *36*, 296–301.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.L., et al. (2008). The Pfam protein families database. Nucleic Acids Research *36*, D281–D288.

Foxman, B., Gillespie, B., Koopman, J., Zhang, L., Palin, K., Tallman, P., Marsh, J.V., Spear, S., Sobel, J.D., Marty, M.J., et al. (2000). Risk factors for second urinary tract infection among college women. American Journal of Epidemiology *151*, 1194–1205.

Foxman, B. (2002). Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Am. J. Med. *113 Suppl 1A*, 5S–13S.

Foxman, B. (2014). Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect. Dis. Clin. North Am. *28*, 1–13.

Garofalo, C.K., Hooton, T.M., Martin, S.M., Stamm, W.E., Palermo, J.J., Gordon, J.I., and Hultgren, S.J. (2007). *Escherichia coli* from urine of female patients with urinary tract infections is competent for intracellular bacterial community formation. Infection and Immunity *75*, 52–60.

Gordon, D.M., Stern, S.E., and Collignon, P.J. (2005). Influence of the age and sex of human hosts on the distribution of *Escherichia coli* ECOR groups and virulence traits. Microbiology (Reading, Engl.) *151*, 15–23.

Greene, S.E., Pinkner, J.S., Chorell, E., Dodson, K.W., Shaffer, C.L., Conover, M.S., Livny, J., Hadjifrangiskou, M., Almqvist, F., and Hultgren, S.J. (2014). Pilicide ec240 disrupts virulence circuits in uropathogenic *Escherichia coli*. mBio *5*, e02038.

Gupta, K., Hooton, T.M., Naber, K.G., Wullt, B., Colgan, R., Miller, L.G., Moran, G.J., Nicolle, L.E., Raz, R., Schaeffer, A.J., et al. (2011). International Clinical Practice Guidelines for the Treatment of Acute Uncomplicated Cystitis and Pyelonephritis in Women: A 2010 Update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin. Infect. Dis. *52*, e103–e120.

Hannan, T.J., Mysorekar, I.U., Chen, S.L., Walker, J.N., Jones, J.M., Pinkner, J.S., Hultgren, S.J., and Seed, P.C. (2008). LeuX tRNA-dependent and -independent mechanisms of *Escherichia coli* pathogenesis in acute cystitis. Molecular Microbiology *67*, 116–128.

Hannan, T.J., Mysorekar, I.U., Hung, C.S., Isaacson-Schmid, M.L., and Hultgren, S.J. (2010). Early Severe Inflammatory Responses to Uropathogenic *E. coli* Predispose to Chronic and Recurrent Urinary Tract Infection. PLoS Pathog *6*, e1001042.

Hannan, T.J., Roberts, P.L., Riehl, T.E., van der Post, S., Binkley, J.M., Schwartz, D.J., Miyoshi, H., Mack, M., Schwendener, R.A., Hooton, T.M., et al. (2014). Inhibition of Cyclooxygenase-2 Prevents Chronic and Recurrent Cystitis. EBioMedicine *1*, 46–57.

Hawn, T.R., Scholes, D., Li, S.S., Wang, H., Yang, Y., Roberts, P.L., Stapleton, A.E., Janer, M., Aderem, A., Stamm, W.E., et al. (2009). Toll-like receptor polymorphisms and susceptibility to urinary tract infections in adult women. PLoS ONE *4*, e5990.

Heaton, B.E., Herrou, J., Blackwell, A.E., Wysocki, V.H., and Crosson, S. (2012). Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of Brucella abortus. J. Biol. Chem. *287*, 12098–12110.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research *37*, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Hultgren, S.J., Schwan, W.R., Schaeffer, A.J., and Duncan, J.L. (1986). Regulation of production of type 1 pili among urinary tract isolates of *Escherichia coli*. Infection and Immunity *54*, 613–620.

Hung, C.-S., Dodson, K.W., and Hultgren, S.J. (2009). A murine model of urinary tract infection. Nat Protoc *4*, 1230–1243.

Johnson, J.R., and Stell, A.L. (2000). Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J. Infect. Dis. *181*, 261–272.

Johnson, J.R., Delavari, P., Kuskowski, M., and Stell, A.L. (2001a). Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. J. Infect. Dis. *183*, 78–88.

Johnson, J.R., O'Bryan, T.T., Delavari, P., Kuskowski, M., Stapleton, A., Carlino, U., and Russo, T.A. (2001b). Clonal Relationships and Extended Virulence Genotypes among *Escherichia coli*Isolates from Women with a First or Recurrent Episode of Cystitis. J. Infect. Dis. *183*, 1508–1517.

Johnson, J.R., Porter, S., Johnston, B., Kuskowski, M.A., Spurbeck, R.R., Mobley, H.L.T., and Williamson, D.A. (2015). Host Characteristics and Bacterial Traits Predict Experimental Virulence for *Escherichia coli* Bloodstream Isolates From Patients With Urosepsis. Open Forum Infect Dis *2*, ofv083.

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J., and Hultgren, S.J. (2004). Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. Proc. Natl. Acad. Sci. U.S.a. *101*, 1333–1338.

Kaper, J.B., Nataro, J.P., and Mobley, H.L.T. (2004). Pathogenic *Escherichia coli*. Nature Reviews Microbiology *2*, 123–140.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Lane, M.C., Simms, A.N., and Mobley, H.L.T. (2007). Complex interplay between type 1 fimbrial expression and flagellum-mediated motility of uropathogenic *Escherichia coli*. Journal of Bacteriology *189*, 5523–5533.

Lebreton, F., van Schaik, W., Manson McGuire, A., Godfrey, P., Griggs, A., Mazumdar, V., Corander, J., Cheng, L., Saif, S., Young, S., et al. (2013). Emergence of Epidemic Multidrug-Resistant Enterococcus faecium from Animal and Commensal Strains. mBio *4*, e00534–13–e00534–13.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Luo, Y., Ma, Y., Zhao, Q., Wang, L., Guo, L., Ye, L., Zhang, Y., and Yang, J. (2012). Similarity and Divergence of Phylogenies, Antimicrobial Susceptibilities, and Virulence Factor Profiles of *Escherichia coli* Isolates Causing Recurrent Urinary Tract Infections That Persist or Result from Reinfection. Journal of Clinical Microbiology *50*, 4002–4007.

Mabeck, C.E. (1972). Treatment of uncomplicated urinary tract infection in non-pregnant women. Postgrad Med J *48*, 69–75.

Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S., and Hultgren, S.J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. Embo J. *19*, 2803–2812.

Mowbray, S.L., and Koshland, D.E. (1987). Additive and independent responses in a single receptor: aspartate and maltose stimuli on the tar protein. Cell *50*, 171–180.

Mulvey, M.A., Lopez-Boado, Y.S., Wilson, C.L., Roth, R., Parks, W.C., Heuser, J., and Hultgren, S.J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. Science *282*, 1494–1497.

Mulvey, M.A., Schilling, J.D., and Hultgren, S.J. (2001). Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. Infection and Immunity *69*, 4572–4579.

Mysorekar, I.U., and Hultgren, S.J. (2006). Mechanisms of uropathogenic *Escherichia coli* persistence and eradication from the urinary tract. Proc. Natl. Acad. Sci. U.S.a. *103*, 14170–14175.

O'Brien, V.P., Hannan, T.J., Yu, L., Livny, J., Roberson, E.D.O., Schwartz, D.J., Souza, S., Mendelsohn, C.L., Colonna, M., Lewis, A.L., et al. (2016). A mucosal imprint left by prior *Escherichia coli* bladder infection sensitizes to recurrent disease. Nat Microbiol *2*, 16196.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research *27*, 29–34.

Parkkinen, J., Rogers, G.N., Korhonen, T., Dahr, W., and Finne, J. (1986). Identification of the O-linked sialyloligosaccharides of glycophorin A as the erythrocyte receptors for S-fimbriated *Escherichia coli*. Infection and Immunity *54*, 37–42.

Piatti, G., Mannini, A., Balistreri, M., and Schito, A.M. (2008). Virulence factors in urinary *Escherichia coli* strains: phylogenetic background and quinolone and fluoroquinolone resistance. Journal of Clinical Microbiology *46*, 480–487.

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., et al. (2008). The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. Journal of Bacteriology *190*, 6881–6893.

Rijavec, M., Erjavec, M.S., Avguštin, J.A., Reissbrodt, R., Fruth, A., Križan-Hergouth, V., and Žgur-Bertok, D. (2006). High Prevalence of Multidrug Resistance and Random Distribution of Mobile Genetic Elements Among Uropathogenic *Escherichia coli* (UPEC) of the Four Major Phylogenetic Groups. Curr Microbiol *53*, 158–162.

Rissman, A.I., Mau, B., Biehl, B.S., Darling, A.E., Glasner, J.D., and Perna, N.T. (2009). Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics *25*, 2071–2073.

Robino, L., Scavone, P., Araujo, L., Algorta, G., Zunino, P., and Vignoli, R. (2013). Detection of intracellular bacterial communities in a child with *Escherichia coli* recurrent urinary tract infections. Pathog Dis *68*, 78–81.

Rosen, D.A., Hooton, T.M., Stamm, W.E., Humphrey, P.A., and Hultgren, S.J. (2007). Detection of intracellular bacterial communities in human urinary tract infection. Plos Med *4*, e329.

Russo, T.A., and Johnson, J.R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J. Infect. Dis. *181*, 1753–1754.

Sahl, J.W., Sistrunk, J.R., Fraser, C.M., Hine, E., Baby, N., Begum, Y., Luo, Q., Sheikh, A., Qadri, F., Fleckenstein, J.M., et al. (2015). Examination of the Enterotoxigenic *Escherichia coli* Population Structure during Human Infection. mBio *6*, e00501.

Salichos, L., and Rokas, A. (2011). Evaluating ortholog prediction algorithms in a yeast model clade. PLoS ONE *6*, e18755.

Scholes, D., Hooton, T.M., Roberts, P.L., Stapleton, A.E., Gupta, K., and Stamm, W.E. (2000). Risk Factors for Recurrent Urinary Tract Infection in Young Women. J. Infect. Dis. *182*, 1177–1182.

Schwartz, D.J., Chen, S.L., Hultgren, S.J., and Seed, P.C. (2011). Population Dynamics and Niche Distribution of Uropathogenic *Escherichia coli* during Acute and Chronic Urinary Tract Infection. Infection and Immunity *79*, 4250–4259.

Schwartz, D.J., Kalas, V., Pinkner, J.S., Chen, S.L., Spaulding, C.N., Dodson, K.W., and Hultgren, S.J. (2013). Positively selected FimH residues enhance virulence during urinary tract infection by altering FimH conformation. Proceedings of the National Academy of Sciences *110*, 15530–15537.

Sefik, E., Geva-Zatorsky, N., Oh, S., Konnikova, L., Zemmour, D., McGuire, A.M., Burzyn, D., Ortiz-Lopez, A., Lobera, M., Yang, J., et al. (2015). Individual intestinal symbionts induce a distinct population of RORγ$^+$ regulatory T cells. Science *349*, 993–997.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nature Methods *12*, 323–325.

Skjøt-Rasmussen, L., Hammerum, A.M., Jakobsen, L., Lester, C.H., Larsen, P., and Frimodt-Moller, N. (2011). Persisting clones of *Escherichia coli* isolates from recurrent urinary tract infection in men and women. Journal of Medical Microbiology *60*, 550–554.

Snyder, E., Gordon, D.M., and Stoebel, D.M. (2012). *Escherichia coli* lacking RpoS are rare in natural populations of non-pathogens. G3 (Bethesda) *2*, 1341–1344.

Springer, M.S., Goy, M.F., and Adler, J. (1977). Sensory transduction in *Escherichia coli*: two complementary pathways of information processing that involve methylated proteins. Proc. Natl. Acad. Sci. U.S.a. *74*, 3312–3316.

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics *21*, 456–463.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

Stapleton, A., Hooton, T.M., Fennell, C., Roberts, P.L., and Stamm, W.E. (1995). Effect of secretor status on vaginal and rectal colonization with fimbriated *Escherichia coli* in women with and without recurrent urinary tract infection. J. Infect. Dis. *171*, 717–720.

Starcic Erjavec, M., Rijavec, M., Križan-Hergouth, V., Fruth, A., and Žgur-Bertok, D. (2007). Chloramphenicol- and tetracycline-resistant uropathogenic *Escherichia coli* (UPEC) exhibit reduced virulence potential. International Journal of Antimicrobial Agents *30*, 436–442.

Steigedal, M., Marstad, A., Haug, M., Damås, J.K., Strong, R.K., Roberts, P.L., Himpsl, S.D., Stapleton, A., Hooton, T.M., Mobley, H.L.T., et al. (2014). Lipocalin 2 imparts selective pressure on bacterial growth in the bladder and is elevated in women with urinary tract infection. J. Immunol. *193*, 6081–6089.

Szmelcman, S., and Hofnung, M. (1975). Maltose transport in *Escherichia coli* K-12: involvement of the bacteriophage lambda receptor. Journal of Bacteriology *124*, 112–118.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet. *5*, e1000344.

Wang, Y., Zhao, S., Han, L., Guo, X., Chen, M., Ni, Y., Zhang, Y., Cui, Z., and He, P. (2014). Drug resistance and virulence of uropathogenic *Escherichia coli* from Shanghai, China. J. Antibiot. *67*, 799–805.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.a. *99*, 17020–17024.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2005). Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization. Infection and Immunity *73*, 7657–7668.

Wu, X.R., Sun, T.T., and Medina, J.J. (1996). In vitro binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. Proc. Natl. Acad. Sci. U.S.a. *93*, 9630–9635.

Wurpel, D.J., Beatson, S.A., Totsika, M., Petty, N.K., and Schembri, M.A. (2013). Chaperone-usher fimbriae of *Escherichia coli*. PLoS ONE *8*, e52835.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research *18*, 821–829.

Chapter 3:

Functional role of the type 1 pilus rod structure in mediating host-pathogen interactions

By

Caitlin N. Spaulding[*], Henry L. Schreiber IV[*], Weili Zheng[*], Karen W. Dodson, Matt S. Conover, Fengbin Wang, Pontus Svenmarker, Areli Luna-Rico, Olivera Francetic, Magnus Andersson, Scott J. Hultgren, Edward H. Egelman

[*]Authors contributed equally to this work

3.1 Abstract

Uropathogenic *E. coli* (UPEC), which cause urinary tract infections (UTI), utilize type 1 pili, a chaperone usher pathway (CUP) pilus, to cause UTI and colonize the gut. The pilus rod, comprised of repeating FimA subunits, provides a structural scaffold for displaying the tip adhesin, FimH. We solved the 4.2 Å resolution structure of the type 1 pilus rod using cryo-electron microscopy. Residues forming the interactive surfaces that determine the mechanical properties of the rod were maintained by selection based on a global alignment of *fimA* sequences. We identified mutations that did not alter pilus production *in vitro* but reduced the force required to unwind the rod. UPEC expressing these mutant pili were significantly attenuated in bladder infection and intestinal colonization in mice. This study elucidates an unappreciated functional role for the molecular spring-like property of type 1 pilus rods in host-pathogen interactions and carries important implications for other pilus-mediated diseases.

3.2 Introduction

To mediate colonization of host and/or environmental habitats, Gram-negative bacteria encode a highly conserved family of adhesive pili called chaperone-usher pathway (CUP) pili. Notably, CUP pili are critical virulence factors in a wide range of pathogenic bacteria, including *Escherichia, Klebsiella, Pseudomonas, Haemophilus, Salmonella* and *Yersiniae* genera (Nuccio and Baumler, 2007). To date, 38 distinct CUP pilus types have been identified in *Escherichia* and *Shigella* genomes and plasmids, each of which is hypothesized to promote bacterial colonization of a distinct habitat (Nuccio and Baumler, 2007; Wurpel et al., 2013). Interestingly, single *Escherichia coli* genomes carry up to 16 distinct CUP operons, suggesting that the retention of an assortment of CUP pilus types by a single strain may be necessary to accommodate the complex lifecycle of *E. coli* (Wurpel et al., 2013).

Arguably, the best-studied CUP pili are those encoded by uropathogenic *E. coli* (UPEC), which is the causative agent of the majority of urinary tract infections (UTIs). UTIs affect 150 million people annually worldwide and are associated with significant morbidity and economic impact, with over $5 billion spent annually on treatment in the USA alone (Foxman, 2014). The current method for treating UTI is antibiotic therapy; however, ~25% of women who experience UTI will suffer from recurrent UTI (rUTI), despite receiving appropriate antibiotic treatment (Scholes et al., 2000). Further, an alarming number of UTIs are being caused by single and multi drug-resistant uropathogens, highlighting the urgent need to develop alternative treatments (Zowawi et al., 2015). UPEC that infect the urinary tract often originate from the host gastrointestinal tract. After being shed from the gut in the feces, UPEC can colonize peri-urethral or vaginal areas and subsequently ascend through the urethra to the bladder and/or kidneys, instigating UTI. In mice, type 1 pili, which promote binding to mannosylated proteins, play critical roles in both the gut and urinary tract. Recent work has revealed that type 1 pili help mediate UPEC intestinal colonization, thus promoting the establishment and/or maintenance of the UPEC reservoir in the gut that can eventually seed UTI (Spaulding et al., 2017). Upon entering the bladder, type 1 pili facilitate bacterial colonization and subsequent invasion into epithelial cells lining the bladder lumen (Martinez et al., 2000; Mulvey et al., 1998). Bladder invasion is a critical step in UPEC pathogenesis, allowing the bacteria to replicate in a niche protected from innate immune defense mechanisms, antibiotics, and expulsion during urination. UPEC that cannot invade the urothelium, like those lacking type 1 pili or its associated adhesin, FimH, are quickly cleared from the bladder, emphasizing the importance of type 1 pili mediated host-pathogen interactions on the fitness of UPEC during cystitis (Wright et al., 2005). After invading into a bladder cell, UPEC escape into the host cell cytoplasm and replicate to form

biofilm-like intracellular bacterial communities (IBCs) (Anderson et al., 2003; Justice et al., 2004) comprised of ~10^4 cells (Wright et al., 2007). Mouse models of UTI have revealed that while some mice are capable of self-resolving acute UPEC infection, others progress to chronic cystitis, which is characterized by persistent high titer bacteriuria (>10^4 CFU/ml) and high bacterial bladder burdens (>10^4 CFU) two or more weeks after inoculation (Hannan et al., 2010). In the absence of antibiotic treatment, chronic cystitis can also be observed in women (Ferry et al., 2004; Mabeck, 1972).

E. coli strains, including UPEC, are grouped into distinct clades (e.g., clades A, B1, B2, D, and E) based on their genetic relatedness (Tenaillon et al., 2010). While UPEC strains tend to be genetically heterogeneous the majority of UPEC strains isolated from women with UTI in the USA reside in the B2 clade (Schreiber et al., 2017). While the types of CUP pilus operons encoded in *E. coli* genomes varies between different clades and individual strains, the vast majority of sequenced *E. coli* strains, including nearly all sequenced UPEC clinical isolates, carry an intact copy of the type 1 pilus (Schreiber et al., 2017; Wurpel et al., 2013). Type 1 pili are encoded by the *fim* operon which, like the gene clusters encoding other CUP pili, encodes all the dedicated proteins necessary to assemble a mature pilus onto the bacterial surface, including: an outer-membrane pore-forming usher protein, a periplasmic chaperone protein, pilus subunits, and the tip adhesin protein. Most pilus tip adhesins, including the FimH adhesin, are made up of two domains, an N-terminal lectin domain, which is responsible for recognition and attachment to a ligand(s), and a C-terminal pilin domain, which connects the adhesin to the bulk of the pilus (Jones et al., 1992). Pilus subunits, including the pilin domain of the adhesin, are comprised of an incomplete immunoglobulin (Ig)-like fold, which lacks the C-terminal β-strand and require the action of the dedicated periplasmic chaperone for proper folding (Sauer et al., 2002). In a

process known as donor strand complementation (DSC), a periplasmic chaperone templates subunit folding by transiently providing one of its β-strands (Sauer et al., 1999). Chaperone-subunit complexes are then delivered to the outer membrane usher, which catalyzes pilus assembly *via* a reaction known as donor strand exchange (DSE). In DSE, the strand donated to a nascent subunit by the chaperone is replaced by the N-terminal extension (Nte) of an incoming subunit (Sauer et al., 2002).

Following this pattern, the type 1 pilus usher (FimD) and chaperone (FimC) help type 1 pili assemble into a composite pilus structure consisting of a short tip fibrillum made up of the adhesin protein (FimH) and two minor subunits (FimG and FimF) that is joined to the pilus rod, a homopolymer of ~1000 FimA subunits. Once extruded to the extracellular surface, the type 1 pilus coils into a rigid right-handed helical structure that is capable of unwinding into a flexible linear fiber (Abraham et al., 1992; Aprikian et al., 2011; Jones et al., 1995; Saulino et al., 2000). This ability to transition between a coiled, helical rod and a linear fiber has been proposed to allow the type 1 pilus to act as a "molecular spring" to maintain adherence in the face of fluid shear forces (Zakrisson et al., 2012). Specifically, we hypothesize that in the absence of urine voiding, the type 1 pilus rod is maintained in the coiled helical state that permits subsequent contact and invasion into bound epithelial cells. However, upon encountering the shear forces associated with micturition, the pilus extends to the linear form, absorbing the shear force and thus preventing the expulsion of the bacteria from the bladder.

Here, using high-resolution cryo-electron microscopy (cryo-EM), we solved the structure of the type 1 pilus rod. Residues involved in critical FimA-FimA interactions were identified and when mutated reduced the force required to unwind the helix, despite not altering the ability of the FimA protein to be incorporated into the pilus rod. When these mutations were introduced

into the chromosome of UTI89, a human UPEC clinical cystitis isolate, no changes in the overall expression of type 1 pili were observed and the ability of the bacteria to bind and agglutinate mannose-expressing guinea pig red blood cells *in vitro* was not altered. However, these mutations dramatically reduced the ability of UTI89 to establish an intestinal reservoir and cause acute and/or chronic cystitis, suggesting that the helical pilus rod has an important functional role, beyond serving as a platform to present FimH, in promoting colonization in the gut and infection of the bladder.

3.3 Results

3.3.1 Determination of the type 1 pilus rod structure

To characterize FimA polymers that form the type 1 pilus rod, we solved the cryo-EM structure of native type 1 pili appearing in a preparation of recombinantly expressed Type IV pili (T4P) from the *E. coli* K12 strain BW25113. As shown in **Figure 1A,** three types of filaments could be separated by eye: T4P, flagellar filaments, and a third class that were thicker and more rigid than T4P but thinner than the flagellar filaments. Sequencing the *fimA* PCR product showed that the encoded amino acid sequence was identical to the FimA protein from BW25113 (GenBank AIN34588.1) and MG1655 (GenBank NP_418734.1). There was no possibility of cross-contamination of the FimA and T4P filament images, as each has a very different helical symmetry. Averaged power spectra from FimA filaments (**Figure S1A**) showed no trace of the Type IV pilus spectrum (**Figure S1B**).

The FimA reconstruction had an overall resolution of 4.2 Å (**Figure S1C,D**), which is sufficient to build an atomic model of the structure (DiMaio et al., 2015). There were no ambiguities in threading the known FimA sequence through the density map (**Table S1),** even though the electron density observed on the outside of the rod is less well defined (**Figure 1B**).

As a result, the first two residues and the last residue of FimA, which are located close to each other on the outside of the rod, are not resolved in the density map. The 4.2 Å estimate of the resolution is consistent with the well-separated β-strands and the density present for certain bulky side chains in the central lumen (**Figure 1C-E**). Further, in the structure we observe the N-terminal donor strand of one subunit completing the β-sheet of the adjacent subunit**.** Overall, the type 1 pilus is a 70 Å diameter rod where each adjacent subunit rotates around the helical axis by 115° and translates along the axis by 7.7 Å (**Figure 2A**). If we label each subunit along the 1-start helix by N, a subunit N_0 interacts with six adjacent subunits (N_{-1},N_{-2},N_{-3},N_{+1},N_{+2},N_{+3}) (**Figure 2A,B**). The donor strand complementation involves mostly hydrophobic interactions. Most prominently, the hydrophobic amino acids Val5, Val10 and Phe12 in the N-terminal β-strand (Nte) of one subunit are inserted into the next subunit's hydrophobic groove created by a missing β-strand (**Figure S2**).

Further, residues corresponding to $FimA_{BW25113}$ 22-33, which are part of the interior surface of the rod helix, show structural differences between our cryo-EM rod structure and the previously solved crystal structures of FimA (Crespo et al., 2012) (**Figure S2A**). In the DSC interaction of the FimA-FimC complex FimA residues 22-25 form a β-strand interaction with residues 59-56; and residues 31-33 form β-strand interactions with the FimC chaperone residues 102-104 (Crespo et al., 2012). Further, within a self-complemented DSE structure FimA residues 29-33 form a β-strand interaction with the appended Nte residues (9-13) (Walczak et al., 2014). Within the cryo-EM rod FimA structure, residues 25-29 adopt a helical conformation and residues 30-33 form a β-strand interaction with residues 10-13 from the Nte of the next FimA subunit. These residues form the center strand within the hollow helical core, and thus the ability

of this region to adopt multiple conformations may be necessary in order for FimA to function as the major rod subunit.

3.3.2 Comparison between P and type 1 pili

The interface between subunits N_0 and N_{+3} is extremely important in maintaining the FimA helical rod. Previously, two atomic models (PDB ID: 2N2H and 2MX3) for the type 1 pilus have been generated using the same solid-state NMR data (Habenstein et al., 2015), with the helical symmetries for these models shown in **Table S2**. When subunit N_0 is superimposed between our cryo-EM model with a subunit in 2N2H and 2MX3, significant differences can be seen in the interface with subunit N_{+3} in the different models (**Figure S3).** The overall RMSD between C_α atoms of subunit N_{+3} in our cryo-EM model and 2N7H is 5.7 Å (**Figure S3A**), while it is 11.0 Å with 2MX3 (**Figure S3B**). Between 2MX3 and 2N2H, the overall RMSD is 5.6 Å for subunit N_{+3} when subunits N_0 are aligned (**Figure S3C**)**.**

Comparison of our type 1 pilus model with the P pilus structure (Hospenthal et al., 2016) shows an interesting difference: in the P pilus, residues 1-5 are fully ordered and make a 90° bend to form a "staple" which involves contacts with two other subunits (**Figure S4**). In contrast, the first two residues are disordered in our type 1 pilus rod, and the remaining three residues in this region project straight out of the structure and make no contacts with other subunits (**Figure S4**).

3.3.3 Conservation and Variability of FimA sequences

To explore the variation and evolution of FimA, we examined a set of 1,872 FimA protein sequences in the Ensembl Bacteria database (Kersey et al., 2016) (**Table S3**). After filtering to remove protein sequences that are expected to be non-functional and trimming of the signal sequence, the remaining 1,828 protein sequences were aligned, revealing that half of

FimA residues in the mature protein are invariant (79/161 residues) and that 83.9% are highly conserved (139/161 residues with >95% sequence identity) (**Figure 3A**). We found the FimA sequences to be significantly more variable than other parts of the type 1 pilus machinery. For example, the mature forms of the FimC chaperone and FimH adhesin (n= 1,760 and n=1,943, respectively; **Table S4**) obtained from the same database contain 97.6% and 97.5% highly conserved residues, respectively, as compared to 83.9% in FimA. In general, most *E. coli* genes evolve slowly (Lee et al., 2012), so the degree of divergence between alleles of FimA is abnormally high. This suggests that the selection forces acting on the FimA protein are different from the forces acting on the rest of the *fim* operon. Interestingly, the non-conserved residues within the FimA sequences are all located on the exterior surface of the solved cryo-EM model (**Figure 3B**), which is likely indicative of immune pressures selecting for antigenic diversification (Wildschutte et al., 2004). In particular, the first few amino acids of the N-terminus of the mature protein, disordered in our reconstruction, have some of the highest variability throughout the entire protein, suggesting that these amino acids do not play a functional role in the dynamics of unwinding in the type 1 pilus rod. In contrast, almost every residue involved in subunit-subunit interactions in the rod is either highly conserved or invariant (**Figure 3B,C**). This includes every residue lining the lumen (**Figure 3C**). The conservation of residues lining the lumen might be due to the transport of some product through this opening that is ~ 15 Å across (the height of the triangular lumen that rotates as one travels along the pilus). More likely, however, every residue in the lumen is either involved in making a subunit-subunit contact (including the DSE contacts with the Nte from an adjacent subunit) or packed tightly between residues that are making such contacts.

To examine the importance of FimA-FimA interactions on bacterial pathogenesis, we chose to investigate how mutations in the *fimA* sequence of BW25113 altered functional and mechanical characteristics of type 1 pili. In addition, to assess whether FimA mutants altered the type-1 mediated ability of UPEC to infect the bladder and colonize the gut, we made mutations in the chromosomal *fimA* gene of UTI89, a prototypical UPEC isolate. Thus, we do note that $FimA_{UTI89}$ differs from our cryo-EM rod $FimA_{BW25113}$ sequence in 16 residues, with all of the differences lying within highly variable residues located on the exterior of the pilus rod, including two additional amino acids at the N-terminus. It is the two additional N-terminal residues in UTI89 that generate a shift in the numbering of all other UTI89 residues by two when compared to $FimA_{BW25113}$.

3.3.4 Evidence of evolutionary pressures on FimA

To evaluate the evolutionary pressures shaping FimA and thus the type 1 pilus rod, we obtained gene sequences encoding FimA proteins analyzed above to measure the selection pressures acting on the gene. After removing incomplete and duplicate sequences from the analysis, we were able to compare a total of 191 unique *fimA* sequences (**Table S3**), which were then trimmed to remove the signal sequence. These sequences encoding the mature protein were examined for evidence of recombination, which can result from horizontal transfer of genetic sequences between distinct strains and confound evolutionary analyses that assume all sequences were vertically inherited (Anisimova et al., 2003). Phylogenetic trees were thus corrected for use in subsequent analyses (Kosakovsky Pond et al., 2006). We then assessed the ratio of the rates of nonsynonymous (*dN*) to synonymous nucleotide mutations (*dS*) in each codon in the alignment (*dN*/*dS*) to estimate the selection pressures acting on each residue in the FimA protein using a fixed-effects likelihood measurement algorithm (Kosakovsky Pond and Frost, 2005). In general,

a *dN*/*dS* ratio >1 is indicative of selection pressure favoring change in amino acid identity at the position (i.e., adaptive or positive selection) while a *dN*/*dS* ratio <1 indicates conservation at the codon (i.e., purifying selection or negative selection). Here, we found that 66 codons were under purifying selection, five codons were under adaptive selection, and an additional 27 codons were too conserved to be included in analysis (**Table S5**). The remaining 63 codons showed no evidence of statistically significant selection in either direction. All of the positively selected sites, N64, A109, T117, S120 and F138, encode residues located on the exterior surface of the FimA rod and away from any subunit-subunit interface (**Figure S5A**). The 66 codons under purifying selection combined with the 27 codons too conserved for analysis encode many of the highly conserved residues found throughout subunit-subunit interfaces with the subunits above, below, and alongside (**Figure 3B,C**). This strongly suggests that evolutionary pressures are working to conserve the identity of residues that interact within the helical rod, while diversifying amino acid residues that are exposed to the host and susceptible to immune recognition.

To more specifically determine if variation in FimA correlated with *E. coli* clades or pathogenic lifestyles, such as uropathogenicity, we examined the diversity and distribution of the *fimA* gene in a curated set of 67 *E. coli* genomes. This dataset included 21 distinct UPEC strains isolated from a cohort of women with frequent rUTI and 46 reference *E. coli* strains that included lab and commensal strains, as well as a variety of intestinal and extra-intestinal pathogens (Schreiber et al., 2017)(**Table S6**). The *fimA* gene was carried by the majority of *E. coli* strains analyzed (57/67); including nearly every UPEC strain (96.3% or 26/27 strains). Importantly, the *fimA* sequences from UPEC strains were spread through the phylogenetic tree, indicating that there was not a single variant of FimA found in all UPEC strains (**Figure S5B**).

To determine if these different variants were under different selection pressure, we measured the *dN*/*dS* ratio of each branch in this *fimA* phylogeny to see if there were branches that were under different selection pressures than the rest of the branches in the phylogeny (i.e., the "background" rate of selection)(Kosakovsky Pond et al., 2011). Here, we identified three branches with statistically significant evidence of episodic diversifying selection, including two branches carrying most of the clinical UPEC strains (branches labeled A and B) as well as enterohemorrhagic and enterotoxigenic *E. coli* strains in the branch labeled C (**Figure S5B**). This pattern of evolution is indicative of strong adaptive selection acting on some, but not all, branches in a phylogeny. Taken together, we find that *fimA* is much more diverse than other parts of the type 1 machinery, owing to high rates of nucleotide polymorphisms and genetic recombination, and that *fimA* has undergone repeated rounds of strong selective pressures that have conserved residues responsible for subunit-subunit interfaces while varying the external surface of the protein that is exposed to the host milieu**.**

3.3.5 Analysis of the effect of FimA mutations on pilus expression and function *in vitro*

To determine which FimA residues are required to form the helical rod, we constructed single amino acid codon mutations in the BW25113 *fimA* gene, which were subsequently cloned into the expression vector, pTRC99a (Amann et al., 1988). Residues were chosen for mutagenesis based on their positions in the atomic model and were changed to either Arginine or Glutamine to insert large charged residues that would promote disruption of FimA interactions without making the surface more hydrophobic. We expressed these variant *fimA* genes *in trans* in UTI89-LONΔ*fimA*, a strain with a chromosomal deletion of the *fimA* gene in a UTI89 strain where the phase-variable *fimS* promoter is locked in the ON orientation (LON) by altering the left inverted repeat necessary for promoter inversion (Kostakioti et al., 2012). This strain

transcribes the *fim* operon constitutively, which removes the possibility that differences in type 1 pili assembly/function are due to effects on phase variation by the mutants.

We assessed the function of the type 1 pilus in each *fimA* mutant by measuring the agglutination of guinea pig red blood cells (GP-RBCs). Most of our *fimA* mutants showed hemagglutination (HA) titers similar to the wild-type (WT) *fimA* (**Table S7**). However, several mutations (A25R, V32R, V65R, V85R, and P145R) abolished the production of adhesive pili (**Table S7**). This suggests that these mutations, which are located in areas of high conservation throughout the FimA monomer, may disrupt critical interactions with FimC or FimD or may not be able to correctly fold, thus preventing the assembly of the pilus. Notably, each of these residues was either under strong purifying selection pressure (A25, V85 and P145) or was too conserved for evolutionary analysis (codons for V32 and V65 displayed no synonymous mutations and extremely limited non-synonymous mutation rates), which emphasizes their importance in assembly of the type 1 pilus rod (**Figure 3A).** Expression of some mutant *fimA* genes, particularly V5R, E45R, E121R and A142R, resulted in a bacterial clumping phenotype when grown in static culture. D62R and D114R also instigated bacterial clumping, but to a lesser extent. Three of these substitutions, E45R, E121R and A142R, are found in residues with relatively high variability in our analysis of FimA sequences, with just 33.9%, 62.3%, and 52.1% sequence identity, respectively. Further, we found that two of these residues (V5 and D114) showed statistically significant evidence of strong purifying selection (**Figure 3A, Table S6**).

To measure the impact of $FimA_{BW25113}$ variants produced by the UTI89-LONΔ*fimA* strain on pili force-extension responses, we applied optical tweezers. Since type 1 pili are assembled from FimA subunits into a helical coil, they normally extend in three force phases: linearly increasing, constant force, linearly increasing – where the characteristic constant force originates

from sequential unwinding of subunits (region between the dashed lines in **Figure 4A**). The unwinding force directly relates to the strength of layer-to-layer interactions (e.g., between subunits N_0 and N_{+3}), thus any changes in these interactions caused by mutations would shift the magnitude of the force plateau. To investigate this, we first averaged the unwinding force of pili in the strain expressing WT FimA and subsequently measured the force response for each *fimA* mutant. The unwinding force of WT pili was 30.3 ± 0.1 pN (mean ± standard error (SE)) whereas all mutants showed a reduction in the force required for pilus unwinding (**Figure 4B**). In particular, the A22R variant showed the largest reduction in the unwinding force, 11.3 ± 0.2 pN. We can explain this by the fact that Ala22 in our rod model is tightly packed against Ala93 from another subunit and Val37 from a third subunit (**Figure 4C**). Thus, replacing a small alanine with a long arginine side chain is expected to disrupt this interface.

3.3.6 FimA mutants alter UPEC pathogenesis in the bladder and colonization of the gut

In the bladder, type 1 pili are required for binding and invasion of UPEC into superficial facet cells that line the bladder lumen and for the formation of intracellular bacterial communities (IBCs) during the first 6-18 hours of acute infection. This has been shown to be critical for ongoing infection in both humans and mouse models of cystitis (Spaulding and Hultgren, 2016). To determine if mutations in the FimA sequence altered UPEC pathogenesis in relevant mouse models, we constructed UTI89 strains with clean single codon mutations in the chromosomal *fimA* gene (with the phase-variable *fimS* promoter intact). Each mutation was made in the codon of a highly conserved amino acid positions of $FimA_{UTI89}$ (A24, D64, D116, and P134) with the exception of E123, which is a position that is highly variable amongst *E. coli* FimA sequences.

When compared to a strain with the reintegrated WT UTI89 *fimA* sequence, we found that each FimA mutant strain produced minimal changes to the levels of type 1 piliation (as assessed by electron microscopy analysis) and to the function of the type 1 pilus, as determined by the ability of bacteria to agglutinate GP-RBCs, indicating that these mutations do not prevent the function or expression of adhesive type 1 pili *in vitro* (**Figure S6**).

We next investigated how each mutant altered the kinetics of bladder infection during competitive infections with WT UTI89 over 28 days. In mice that developed chronic cystitis (defined as the development of persistent high titer ($>10^4$ cfu/ml) bacteriuria and high titer ($>10^4$ cfu/ml) bladder bacterial burdens at sacrifice >4 weeks post-infection), strains producing $FimA_{UTI89}$ variants with the D64R, E123R, D116R, and A24R substitutions were outcompeted, by up to 6 logs, by the reintegrated WT strain (**Figure 5A-F**). The FimA P134R variant had no effect on the ability of the strain to compete with the WT strain in chronically infected mice (**Figure 5B**). Mice that resolved their infections (defined as any animal whose urine or bladder titers dropped below 10^4 cfu/ml at least once during the 4 week infection) in this experiment are shown in **Figure S7**.

In mice infected with a single UTI89 strain, the WT and P134R variant caused chronic cystitis at similar rates (45% and 30%, respectively) (**Figure 5G, Figure S8**). However, the E123R and D116R mutant strains had reduced rates of chronic cystitis of 20% and 10%, respectively. Interestingly, 100% of mice infected with the A24R variant resolved their infections over the 4 week experiment, with half of the mice (10/20) exhibiting sterile urines by 10 dpi compared to just 15% (3/20) of mice infected with the WT strain at the same time point (**Figure S8**)**.** The defect in chronic infection caused by *fimA* variants is likely due to pathogenic deficiencies during acute UTI. Accordingly, we found that three *fimA* variants (A24R, E123R,

and D116R) significantly altered the ability of UTI89 to form IBCs at 6 hours post infection (hpi) (**Figure 5H**). Six of ten mice infected with the E123R variant and eight of ten of the mice infected with the D116R variant formed less than 15 IBCs. Even more strikingly, 70% of mice infected with the A24R variant formed no IBCs at 6 hpi and the other 30% formed 3 or less. This is in stark contrast to WT and P134R strains, which formed between 50-100 IBCs on average by 6hpi. We also observed that, on average, the IBCs formed at 12 hpi in E123R and D116R infected mice were smaller than those formed by the WT strain (**Figure 5I**). Interestingly, while mice infected with the P134R variant formed a similar number of IBCs as WT at 12 hpi the IBCs formed were smaller than those formed by the WT strain. No IBCs were detected at 12 hours with the A24R variant. The D64R variant was defective in chronic cystitis in both the competitive and single bladder infections, but no acute fitness defects were observed. Mice infected with D64R formed a similar number of IBCs as the WT strain at both 6 and 12 hpi and at 12 hpi the size and morphology of the IBCs was similar to that of WT infected mice, suggesting that the defect occurs at a later time-point. Accordingly, in mice singly infected with *fimA* D62R bacterial clearance is delayed, starting between 10-14 days.

In addition to playing a pivotal role in the urinary tract, a recent study found that type 1 pili also promote the establishment and/or maintenance of the UPEC intestinal reservoir (Spaulding et al., 2017). Deletion of the operon encoding the type 1 pilus or the FimH adhesin impedes intestinal colonization by UTI89 (Spaulding et al., 2017). Correspondingly, strains producing $FimA_{UTI89}$ variants D64R, D116R, and A24R also showed lower levels of intestinal colonization (by up to 2 logs) in the feces, cecum, and colon of mice compared to those colonized by WT *fimA*. This 2-log decrease mirrors the defect observed in UTI89Δ*fim* or UTI89Δ*fimH* strains, suggesting that these mutations prevent type 1 pilus-dependent gut

colonization in UTI89 (Spaulding et al., 2017). While not statistically significant, *fimA*$_{UTI89}$ P134R variant showed lower colonization (by up to one log) in the feces and cecum than the WT strain (**Figure 6**). Together, our data indicates that the sequence of the FimA major subunit is critical for pilus function in the bladder and gut and thus has a major impact on the outcome of UTI.

3.4 Discussion

3.4.1 Implications of advances in cryo-electron microscopy

Tremendous advances in cryo-EM within the past four years (Egelman, 2016; Subramaniam et al., 2016) largely driven by the introduction of direct electron detectors (Li et al., 2013) has meant that many complexes that were recalcitrant to crystallization can now be readily solved at near-atomic resolution by cryo-EM. In particular, it is exceedingly difficult to crystallize most helical polymers, as unless such a polymer has exactly two, three, four or six subunits per turn, it cannot be packed in any crystal space group so that all subunits are in equivalent environments. The type 1 pilin, FimA, has been extensively studied by x-ray crystallography and solution NMR, while the type 1 pilus has only been studied at high resolution by solid-state NMR (Habenstein et al., 2015). We show here that type 1 filaments, present as a background in a preparation of T4P, allow us to reach a near-atomic level of resolution and build an atomic model for the FimA rod. The ability of cryo-EM to separate out multiple conformations among particles (Gui et al., 2017; Vonck and Mills, 2017) or even of subunits within the same particle (Roh et al., 2017) has been one of the greatest strengths of cryo-EM, allowing for multiple states to be solved from the same micrographs. Biochemically heterogeneous preparations, such as when virions are a mixture of empty particles and those containing DNA (Dong et al., 2017) are now routinely sorted out into homogeneous structural

classes. We have taken advantage of that strength here to show that filament images which might otherwise have been discarded as a background contaminant can actually be used to build an atomic model. Our type 1 pilus rod model provides new insights into structure-function relationships in type 1 pili.

3.4.2 Type 1 pilus rod as a "molecular spring"

FimA is critical for proper assembly of the type 1 pilus and, as the major subunit that makes up the pilus rod, is also critical for the proper display of the FimH adhesin. However, here we have uncovered a more complex and previously unappreciated role for FimA and the type 1 pilus rod in host-pathogen interactions. We identified four mutations in $FimA_{UTI89}$ (A24R, E123R, D116R, and D64R) that expressed adhesive pili *in vitro* but reduced the ability of UTI89 to colonize the bladder, acutely and chronically. One of these mutants, $FimA_{UTI89}$ A24R, was severely attenuated during acute infection, forming almost no IBCs and thus was unable to chronically infect the mouse bladder.

The pilus rod is hypothesized to act as a "molecular spring" transitioning between a flexible, linear fiber and a coiled helix. This spring-like property is thought to prevent the pilus from breaking or detaching from the host surface by temporarily expanding to a linear form after encountering shear forces, which can occur in the bladder during urine voiding or in the gut during fecal or mucus shedding. Such a transition between the helical and unwound form of the FimA homopolymer has been predicted to significantly dampen the force experienced by the adhesion-receptor complex at the tip (Zakrisson et al., 2012). Since FimA-FimA interactions create the bulk of the pilus rod, mutations that reduce the stability of these protein-protein interactions can alter the ability of the pilus rod to withstand shear forces, which we discovered has detrimental effects on pilus function and pathogenesis. This is consistent with ~ 50% of the

residues in the mature FimA protein (79/161) being invariant (**Figure 3**), which includes almost every residue involved in a subunit-subunit interface in the rod. Pili formed with the WT $FimA_{BW25113}$ protein require more than twice the force to unwind than the pili formed by the $FimA_{BW25113}$ variant with an A22R mutation (~30pN vs. 11pN), likely due to steric clashes caused by this substitution affecting the tightly packed interface between subunits. Thus, when elongated during periods of urine flow, the pilus formed by the FimA A22R mutant variant may not have the strength to absorb and withstand the shear forces experienced during urination and thus allow the bacteria to be swept out of the bladder, preventing infection. In the gut, the pilus rod may play a similar role during mucus shedding. A recent study suggests that type 1 pili may promote UPEC colonization of the upper crypts (Spaulding et al., 2017). These bacteria would likely experience constant, low levels of shear force during mucus turnover and thus need the pilus to withstand some force to enable the bacteria to maintain their intestinal niche. Accordingly, three of the $fimA_{UTI89}$ mutants (A24R, D116R, and D64R) that were attenuated in bladder colonization also displayed significantly reduced intestinal colonization. However, in general, the phenotypes of the $fimA_{UTI89}$ mutants in gut colonization were not as severe as in the bladder.

Three mutations in $FimA_{UTI89}$ (D64R, D116R and E123R) resulted in a clumping phenotype when bacteria were grown *in vitro*. This clumping phenotype was also seen in the equivalent mutants in $FimA_{BW25113}$ (D62R; D114R, E121R). These residues are all located on the exterior surface of the rod structure and those tested by force extension ($FimA_{BW25113}$ D62R; D114R) do not have as strong an effect on the force needed for unwinding as $FimA_{BW2511}$ A22R. However, the clumping phenotype suggests that these mutations may alter side-to-side interactions between different pilus rods promoting pilus-pilus interactions within and between

bacteria, thus inhibiting the pilus mediated interactions needed *in vivo* for attachment, invasion and/or IBC formation. Interestingly, we do not observe a defect in bladder colonization in mice infected with $FimA_{UTI89}$ D64R until 10dpi, suggesting that normal rod function is needed during both chronic and acute infection.

3.4.3 Evolutionary pressures on the type 1 pilus rod

Together, our data suggest that the type 1 pilus rod may mediate colonization phenotypes through damping of shear forces or through other mechanism(s) outside of scaffold support for FimH-binding. Our analyses indicate that FimA is undergoing selection pressures due to as-yet undefined host-pathogen interactions, which may explain some colonization defects seen here. FimA displays patterns of episodic, divergent selection on surface exposed amino acids in a pattern that is similar, though much less robust, to what is seen in the flagellin protein in *Salmonella* (Li et al., 1994). In *Salmonella*, the flagellin subunits are split into several domains where the domains responsible for subunit-subunit assembly are highly conserved while the more external domains of the protein show high rates of variability (Galkin et al., 2008;Andersen-Nissen et al., 2005). The *Salmonella* flagellin protein is immunogenic, but the regions that induce inflammatory response are found in the conserved domains of the protein (Wildschutte et al., 2004). While it is known that flagellin are bound by Toll-like receptor 5 (TLR5) resulting in the induction of the innate immune response (Smith and Ozinsky, 2002) and that this recognition is targeted towards conserved features of flagellin (Andersen-Nissen et al., 2005), it is currently unknown which, if any, host immune receptor are capable of recognizing the FimA rod or subunits or which structural features are targeted by the host immune system. Further research is needed to fully elucidate the host pressures and responses that are influencing the evolution of the FimA rod structure.

Evolutionary and structural analysis of FimA, in combination with our *in vitro* and *in vivo* phenotyping, yielded several important insights into the selection pressures faced by UPEC as well as the evolutionary trajectories that pathogens follow to enhance their colonization of different host niches. Notably, we saw that different phenotypes caused by mutations in the FimA protein were associated with different classes of selection pressure. For example, the bacterial clumping phenotype is associated with mutation of three codons under purifying selection, but an equal number of the mutations that resulted in clumping are in codons with little to no evidence of selective pressure. In contrast, all mutations that failed to complement a *fimA* gene deletion were made in codons that are under very strong purifying selection or are too conserved for analysis. Together, the difference in selection pressure suggests that the mis-assembly of the pilus is much more harmful and/or toxic to *E. coli* than bacterial clumping. Further, the integration of evolutionary analysis with *in vivo* and *in vitro* functional analysis allowed us to decouple the selection pressures acting to preserve amino acid sites necessary for pilus assembly (such as P147 in $FimA_{UTI89}$) from the selection pressure maintaining the codons that were necessary for pilus function (such as A24 in $FimA_{UTI89}$). Given the intricacy of pilus assembly, one could expect that most of the codons under purifying selection would be related to pilus construction. Instead, we found that many codons under purifying selection are involved in keeping the force needed to unwind the type 1 rod within a narrow range. This leads to reasoning that a "weak" or “loose” FimA rod may be just as detrimental for *E. coli* as having no rod at all, at least in the eyes of evolution.

3.4.4 Summary

In summary, by combining structural studies, force spectroscopy, genetic analysis, and relevant mouse models of UTI and gut colonization, we conclude that the mechanical properties

of the type 1 pilus rod are essential for its functional role in mediating *E. coli* pathogenesis and persistence and appear to have been carefully "tuned" by evolution. Further studies of the hundreds of CUP pili encoded in Gram-negative bacteria are needed to further understand the unique and general aspects of the evolution of CUP pilus fibers. In addition, other bacterial pili, such as T4P, which have arisen independently of CUP pili but can play similar roles in pathogenesis, can also elongate under force (Biais et al., 2010) and thus it remains an interesting question as to how the physical properties of other pili have been selected for particular environments and how these properties impact bacterial pathogenesis.

3.5 Materials and Methods

3.5.1 Ethics statement

The Washington University Animal Studies Committee approved all procedures used for the mouse experiments described in the present study. Overall care of the animals was consistent with *The Guide for the Care and Use of Laboratory Animals* from the National Research Council and the USDA *Animal Care Resource Guide*.

3.5.2 Bacteria, cloning, mutagenesis

The BW25113 *fimA* gene sequence was cloned between the *EcoRI* and *BamHI* restriction sites in pTRC99A using standard PCR cloning techniques to create plasmid pTRC-fimA. Mutations were made within this plasmid using appropriate complementary primers to engineer codon changes in the template, pTRC-fimA, using *Pfx* polymerase and manufacturers instructions for PCR, followed by *DpnI* treatment of the resulting products to remove the methylated template before transformation into C600. Mutations were verified by sequencing. Mutant plasmids were transformed into UTI89-LON, Δ*fimA* for expression and functional studies as indicated.

In order to construct point mutations in the *fimA* allele in the UTI89 chromosome, the UTI89 *fimA* gene was deleted using a previously published technique that allows for flawless integration (Khetrapal et al., 2016). Briefly, *fimA* was deleted by homologous recombination using pSLC- 217 as a template and primers containing 50bp of homology to flanking regions of *fimA*. A deletion was then constructed using the previously described Red Recombinase method that would allow for reinsertion of constructs into the *fimA* site. Concurrently, a copy of UTI89 *fimA* was cloned into pTRC99a. Point mutations were then introduced into this construct using site directed mutagenesis. PCR fragments from confirmed mutants, and the WT allele, were reintegrated into the UTI89-LON, Δ*fimA* mutants constructed above at the original deletion site. Successful reintegration events were sequenced to confirm flawless integration and mutation presence.

3.5.3 Mouse studies

Animals were maintained in a single room in our vivarium. Prior to and after infection all animals received PicoLab Rodent Diet 20 (Purina) *ad libitum*. All animals were maintained under a strict light cycle (lights on at 0600h, off at 1800h). Mice were acquired from indicated vendors and randomly placed into cages (n=5 mice/cage) by employees of Washington University's Division of Comparative Medicine (DCM); no additional methods for randomization were used to determine how animals were allocated to experimental groups. Investigators were not blinded to group allocation during experiments.

For bladder infections, 6-week old female C3H/HeN mice were obtained from Envigo and were maintained in our vivarium for one week prior to infection. Bladder infections were performed via transurethral inoculation (Hung et al., 2009). UPEC strains were prepared for inoculation as described previously (Hung et al., 2009). Briefly, a single UTI89 colony was

inoculated in 20 mL of Luria Broth (LB) and incubated at 37°C under static conditions for 24 h. Bacteria were then diluted (1:1000) into fresh LB and incubated at 37°C under static conditions for 18-24 h. Bacteria were subsequently washed three times with PBS and then concentrated to ~1x10^8 CFU per 100 µL for intestinal infections and ~1x10^8 CFU per 50 µL for bladder infections. Bacteria were subsequently washed three times with PBS and then concentrated to ~1x10^8 CFU per 50 µL for bladder infections.

For intestinal colonization experiments, 6-week old female C3H/HeN mice were obtained from Envigo and were maintained in our vivarium for no more than 2 days prior to intestinal colonization. Mice received a single dose of streptomycin (1000mg/kg in 100 µL water by oral gavage (PO)) followed 24 h later by an oral gavage of ~10^8 CFU UPEC in 100 µL phosphate-buffered saline (PBS) (Spaulding et al., 2017). Bacteria were subsequently washed three times with PBS and then concentrated to ~1x10^8 CFU per 100 µL for intestinal infections.

In all cases, fecal and urine samples were collected directly from each animal at the indicated time points. Fecal samples were immediately weighed and homogenized in 1 mL PBS. Urine samples were immediately diluted 1:10 prior to plating. Mice were sacrificed via cervical dislocation under isofluorane anesthesia and their organs were removed and processed under aseptic conditions. Intestinal segments (cecum and colon) were weighed prior to homogenization and plating on LB supplemented with the appropriate antibiotic.

3.5.4 Enumeration of bladder intracellular bacterial communities (IBCs)

6-week old female C3H/HeN mice were given a transurethral inoculation with WT UTI89 or a *fimA* mutant strain. To accurately count the number of IBCs, mice were sacrificed 6 or 12 hours after infection. Bladders were removed aseptically, bi-sected, splayed on silicone plates and fixed in 4% (v/v) paraformaldehyde. IBCs, readily discernable as punctate violet spots,

were quantified by LacZ staining of bladder wholemounts or by immunofluorescence (Cusumano et al., 2011; Justice et al., 2006).

3.5.5 Hemagglutination assays (HA)

Bacteria were grown under type 1 pilus-inducing conditions (Greene et al., 2015), with appropriate antibiotics and .01-.02mM IPTG induction, if indicated. Pilus expression was assessed by hemagglutination assays (HA) as previously described (Greene et al., 2015) using bacterial cultures normalized to an optical density at 600 nm (OD_{600}) of 1 and guinea pig erythrocytes normalized to an OD_{640} of 2. The experiment was conducted in parallel in PBS with 2% w/v methyl-α-D-mannopyranoside.

3.5.6 Electron microscopy

Electron micrographs (EM) were taken of UTI89 or UTI89 isogenic mutants after growth under type 1 pilus-inducing conditions. A total of 300 bacterial cells were counted for each condition, and piliation on those cells was classified as bald (no pili), low (1 to 20 pili/cell), moderate (20 to 200 pili/cell), or abundant (>200 pili/cell).

3.5.7 Force extension experiments

For expression of type 1 pili the strains were grown in Luria Broth (LB) supplemented with carbenicillin (100 μg/mL) and IPTG (50μM), at 37°C overnight. The optical tweezers (OT) setup is built around an inverted microscope (Olympus IX71, Olympus, Japan) equipped with a high numerical aperture oil immersion objective (model: UplanFl 100X N.A. = 1.35; Olympus, Japan) and a 1292 x 964 pixel camera with a cell size of 3.75 x 3.75 μm (model: StingRay F-125, Allied Vision)(Mortezaei et al., 2013) (**Figure S9A**). The OT stands in a temperature controlled room with computers and controllers isolated from the room to reduce noise and vibrations. We use a continuous wave Nd:YVO_4 laser (Millennia IR, Spectra Physics, Santa Clara, CA)

operating at 1064 nm for trapping a single bacterium or microspheres. A probe laser (low power HeNe-laser operating at 632.8 nm) is merged with the trapping laser using a polarizing beam splitter cube (PBSC). The light from the probe laser is refracted by the trapped object and collected by the condenser and thereafter imaged onto a 2D position sensitive detector (PSD, L20 SU9, Sitek Electro Optics, Sweden). The PSD convert the incoming light to a photocurrent and thereafter to a voltage that is sent to a programmable low pass filter (SR640, Stanford research systems), later collected by a computer and processed with an in-house LabVIEW program. We minimized the amount of noise in the setup and optimized the measured time series using the Allan variance method described in (Andersson et al., 2011).

To prepare a sample we suspended bacteria in 1xPBS to a concentration (1:1000 of OD_{600} = 1) suitable for single cell analysis using optical tweezers (OT). Surfactant-free 2.5 μm amidine polystyrene microspheres (product no. 3-2600, Invitrogen, Carlsbad, CA) were similarly suspended in Milli-Q water, these microspheres were trapped and used as force probes. To mount bacteria and reduce the influence of surface interactions we prepared a 1:500 suspension of 9.5 mm carboxylate-modified latex microspheres (product no.2-10000, Interfacial Dynamics, Portland, OR) in Milli-Q. We dropped ten microliters of the microsphere-water suspension onto 24 x 60 mm coverslips (no.1, Knittel Glass, Braunschweig, Germany) and placed these in an oven for 60 min at 60°C to immobilize the microspheres to the surface. To firmly adhere bacteria to the microspheres, we added a solution of 20 mL of 0.01% poly-L-lysine (catalog no. P4832, Sigma-Aldrich, St. Louis, MO) to the coverslips, which, after 45 min incubation at 60°C, were stored until use.

To make a flow chamber, we added a ring of vacuum grease (Dow Corning, Midland, MI) around the area containing the poly-L-lysine-coated microspheres on one of the coverslips.

Carefully, we dropped a 3-mL suspension of bacteria and a 3-mL suspension of probe microspheres (surfactant-free 2.5 mm white amidine polystyrene latex microsphere, product no. 3-2600, Invitrogen, Carlsbad, CA) onto the area and sealed the flow chamber by placing a 20 x 20 mm coverslip (no.1, Knittel Glass) on top. Thereafter, we mounted the sample in a sample holder that is fixed to a piezo-stage (Physik Instrument, P-561.3CD stage) in the OT instrumentation. To get a reliable OT calibration parameter values we measured the temperature using a thermocouple in the sample chamber, 23.0°C ± 0.1°C and the suspension viscosity was assumed to only vary with temperature, thus, the viscosity was set to 0.932 mPas ± 0.002 mPas.

A free-floating bacterium was trapped by the optical tweezers run at low power to avoid cell damage. The bacterium was thereafter mounted on a large 9.5 μm microsphere coated with poly-L-lysine. We trapped a small free-floating 2.5 μm microsphere by the optical tweezers with normal power (a few hundreds of mW) and brought it close to (within tens of μm) but not in direct contact with, the bacterium. To calibrate the trap stiffness we used the Power spectrum method by sampling the microspheres position at 131,072 Hz and average 32 consecutive data sets acquired for 0.25 s each (Tolic-Norrelykke et al., 2006). Typically, the trap constant was found to be ~140 pN/μm for an output laser power of 800 mW. After calibration, the small microsphere was brought close to the bacterium in order to attachment a few pili with the microsphere (**Figure S9B**). To extend pili the piezo stage was typically moved at a constant speed of 10 nm/s and the sampling frequency was set to 10,000 Hz that was downsampled by 800. We controlled the piezo-stage and sampled the data using an in-house LabView program that is available upon request.

3.5.8 Pilus preparation for structural determination

To prepare the pilus extracts, bacteria of the *E. coli* strain BW25113 (Datsenko and Wanner, 2000) were inoculated by dense streaking on eight M9 minimal agar plates containing 0.5 % glycerol (vol/vol). After a 72-hour incubation at 30°C, bacteria were harvested in 30 mL of LB medium, vortexed vigorously for 5 min and passed 8 times through a 26-Gauge needle, to detach pili from the cells. Bacteria were removed at 4°C by three successive 10-min centrifugation steps at 16,000 x g. To collect the pili, cleared supernatants were centrifuged for 1 hr at 100,000 *g* in a cold Beckman Ti60 ultracentrifuge rotor. Pellet containing the crude pilus fraction was taken up in 200 mL of 50 mM HEPES, 50 mM NaCl pH 7.4, and maintained at 4°C for further analysis.

3.5.9 Cryo-electron microscopy data collection and image processing

3μL of sample was applied to glow discharged lacey carbon grids (TED PELLA, Inc., 300 mesh). Then the grids were plunge-frozen using a Vitrobot Mark IV (FEI, Inc.), and subsequently imaged in a Titan Krios at 300keV with a Falcon II direct electron detector (pixel size 1.05 Å /pixel). A total of 6,803 images, each of which was from a total exposure of 2 seconds dose-fractionated into seven chunks, were collected at a range of underfocus between 0.5~3μm. Images were motion corrected using MotionCorr (Li et al., 2013), and the program CTFFIND3 (Mindell and Grigorieff, 2003) was used for determining the defocus and astigmatism. Images with poor CTF estimation as well as defocus > 3μm were discarded. The SPIDER software package (Frank et al., 1996) was used for most other operations with the first two-chunk sums (containing a dose of ~ 20 electrons/ $Å^2$) of the motion-corrected image stacks. The CTF was corrected by multiplying the images from the first two-chunk sums with the theoretical CTF, which is a Wiener filter in the limit of a very poor signal-to-noise ratio (SNR). This both corrects the phases which need to be flipped and improves the SNR. The e2helixboxer

routine within EMAN2 (Tang et al., 2007) was used for boxing the filaments from the images. A total of 72,627 overlapping segments (384 px long), with a shift of 11 px between adjacent segments (~97% overlap), were used for the IHRSR (Egelman, 2000) reconstruction. With a featureless cylinder as a starting reference, 72,627 segments were used in IHRSR cycles until the helical parameters (axial rise and rotation per subunit) converged. Analysis of the population suggested that the axial rise was fairly fixed, but that the twist was variable. Using a reference-based sorting with models having a fixed rise but a variable twist, approximately 55% of the segments were excluded, having a twist outside of the range 114.4° to 115.6°. A sub-set of 32,726 segments were used for a few more cycles of IHRSR. The resolution of the final reconstruction was determined by the FSC between two independent half maps, generated from two non-overlapping data sets, which was 4.2 Å at FSC=0.143.

3.5.10 Model Building and Refinement

We used a previous FimA NMR model (PDB id: 2JTY, a single chain) as an initial template to dock into the cryo-EM map by rigid body fitting, and then manually edited the model in Chimera (Pettersen et al., 2004) and Coot (Emsley et al., 2010). We then used the combined model (1-19 and 21-159) as the starting template to re-build a single chain of the FimA protein using the RosettaCM protocol (Wang et al., 2015). Next, the full length model of FimA missing the first two residues and the last residue was iteratively refined by Phenix real-space refine (Adams et al., 2010) and manually adjusted in Coot. The refined single chain of the FimA model was then re-built by RosettaCM (Wang et al., 2015) with helical symmetry and refined by Phenix to improve the stereochemistry as well as the model map coefficient correlation. The FimA model was validated with MolProbity (Chen et al., 2010) and the coordinates deposited to the Protein Data Bank with the accession code XXXX. The corresponding cryo-EM map was

deposited in the EMDB with accession code YYYY. The refinement statistics are given in **Table S1**.

3.5.11 Bioinformatic analyses

The protein sequences of closely-related homologs of *E. coli* FimA, FimC, and FimH were obtained by individual searches the Ensembl Bacteria Genome database (Kersey et al., 2016) using the phmmer web-server (Finn et al., 2015) with a BLOSUM62 (FimA) or a BLOSUM90 (FimC and FimH) scoring matrix using the full-length *E. coli* UTI89 protein sequences as queries. The sequence matches were then filtered to remove low scoring hits and non-functional sequences (i.e., those predicted to lack critical sequence features such as complete signal sequences and/or C-terminal tyrosine residues in FimA and FimH). The signal sequences were trimmed from the remaining homologs using Geneious v 6.1.7 (Kearse et al., 2012) and the protein sequences were aligned with the MAFFT program using two iterations of the FFT-NS-i algorithm based on the PAM200 scoring matrix (Katoh and Standley, 2013). Conservation at each amino acid position was calculated using custom Python scripts and a sequence logo was created using the ggseqlogo package in R (R Core Team, 2017) using RStudio (RStudio Team, 2015).

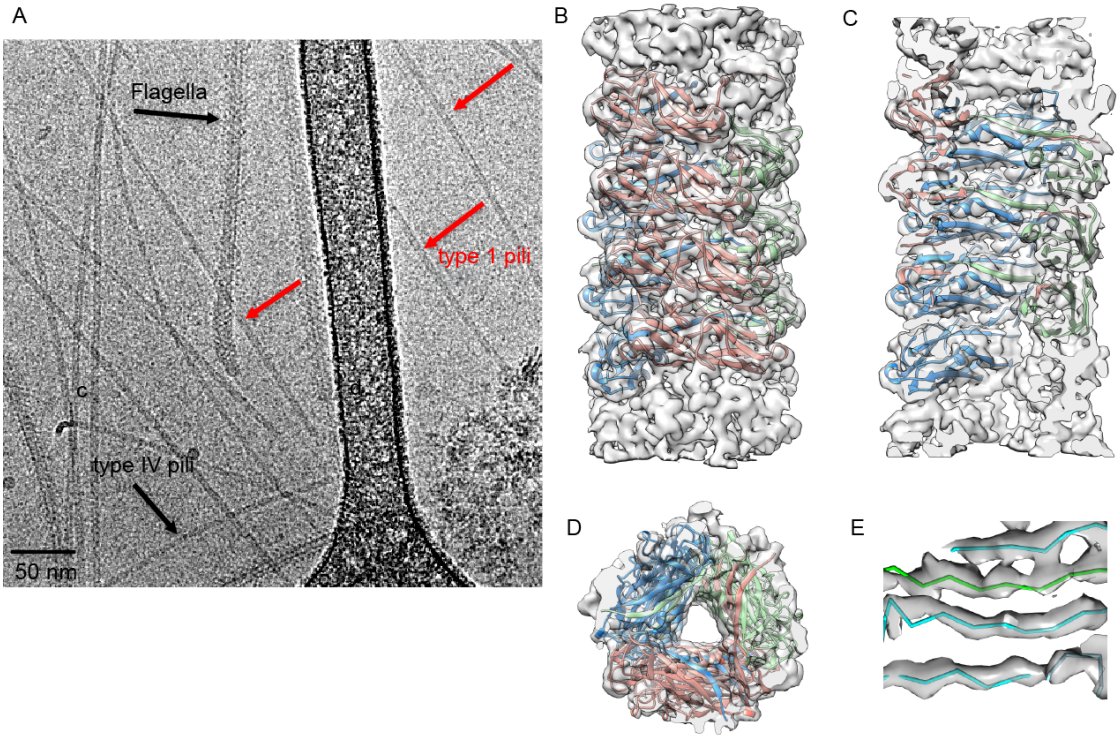
To estimate selection pressures on each codon in *fimA*, we obtained all available gene sequences encoding the protein sequences described above from the Ensembl Bacteria Genomes database (n=1,825, three were removed by submitter's request) using custom bash scripts (Table S3). A total of 191 unique sequences were identified using Geneious v 6.1.7 and trimmed to remove the signal sequence. Evolutionary model selection was performed using maximum likelihood ratio testing on the Datamonkey webserver (Delport et al., 2010; Pond and Frost, 2005), which identified the TIM2 model (model 010232) as the most likely model of nucleotide

substitution for the *fimA* homologs. These sequences were then were scanned for evidence of recombination using single breakpoint analysis (Kosakovsky Pond et al., 2006) and phylogenetic trees with a correction for a breakpoint found in codon 107 (position 320) were generated. These phylogenetic trees and evolutionary model were then used to measure the ratio of the rates non-synonymous (*dN*) to synonymous (*dS*) mutation in each codon (i.e., a *dN*/*dS* ratio) using a fixed-effects likelihood test to identify statistical significance (Kosakovsky Pond and Frost, 2005).

Using a collection of 67 curated, reference genomes from a previous study (Schreiber et al., 2017), we examined the carriage and phylogenetic context of *fimA* carriage using a BLAST-based search (Camacho et al., 2009,) with the UTI89 *fimA* gene as a query. Full-length sequences were extracted from the genomes using Geneious v 6.1.7, trimmed to remove signal sequences, and aligned using the MUSCLE program (Edgar, 2004). A phylogenetic tree was estimated using the RAxML program (Stamatakis, 2006) with the GTRCAT model and supported with 1000 bootstraps (Stamatakis et al., 2008). Evidence for episodic, diversifying selection was then identified using a random effects likelihood ratio test for each branch of the *fimA* phylogenetic tree (Kosakovsky Pond et al., 2011) using unique sequences from the genomes (32 duplicates removed, n=25). Branches showing statistically significant evidence for episodic, diversifying selection were then indicated on the corresponding branches of the phylogenetic tree constructed using RAxML.

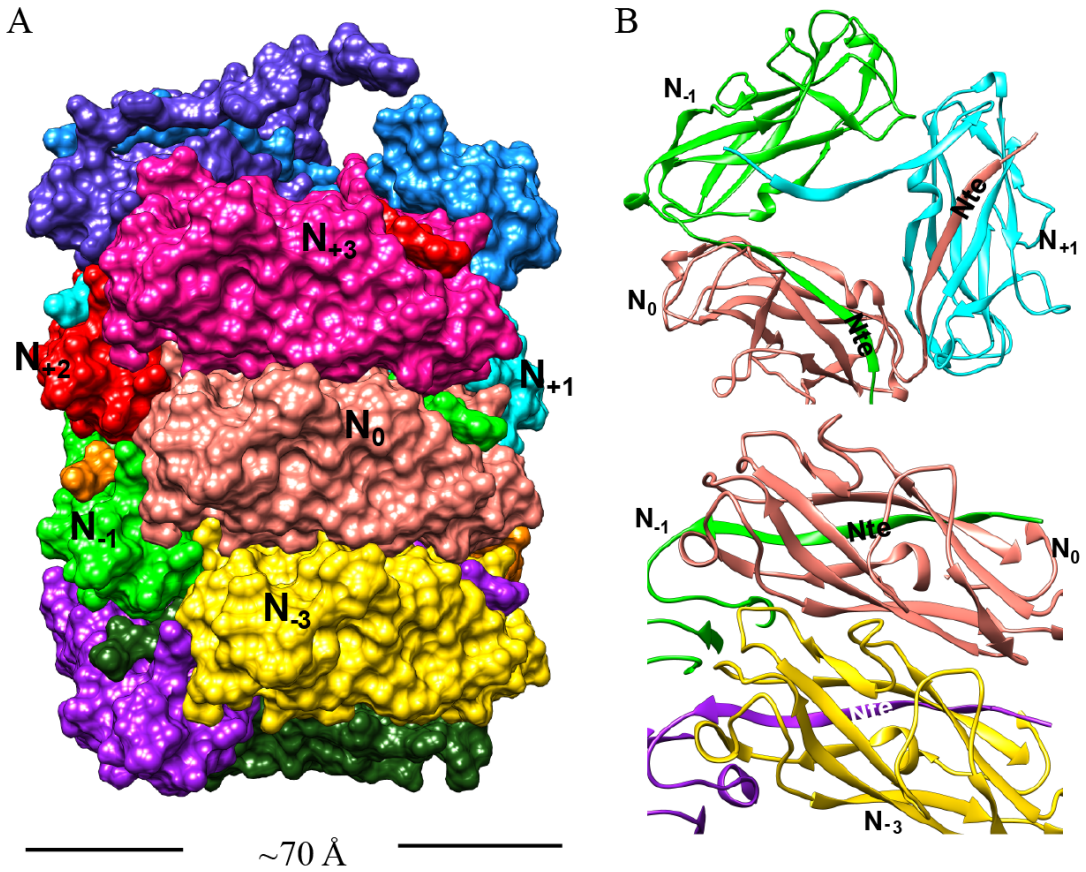
3.6 Figures

3.6.1 Figure. Cryo-EM Structure of Type 1 pili. 1



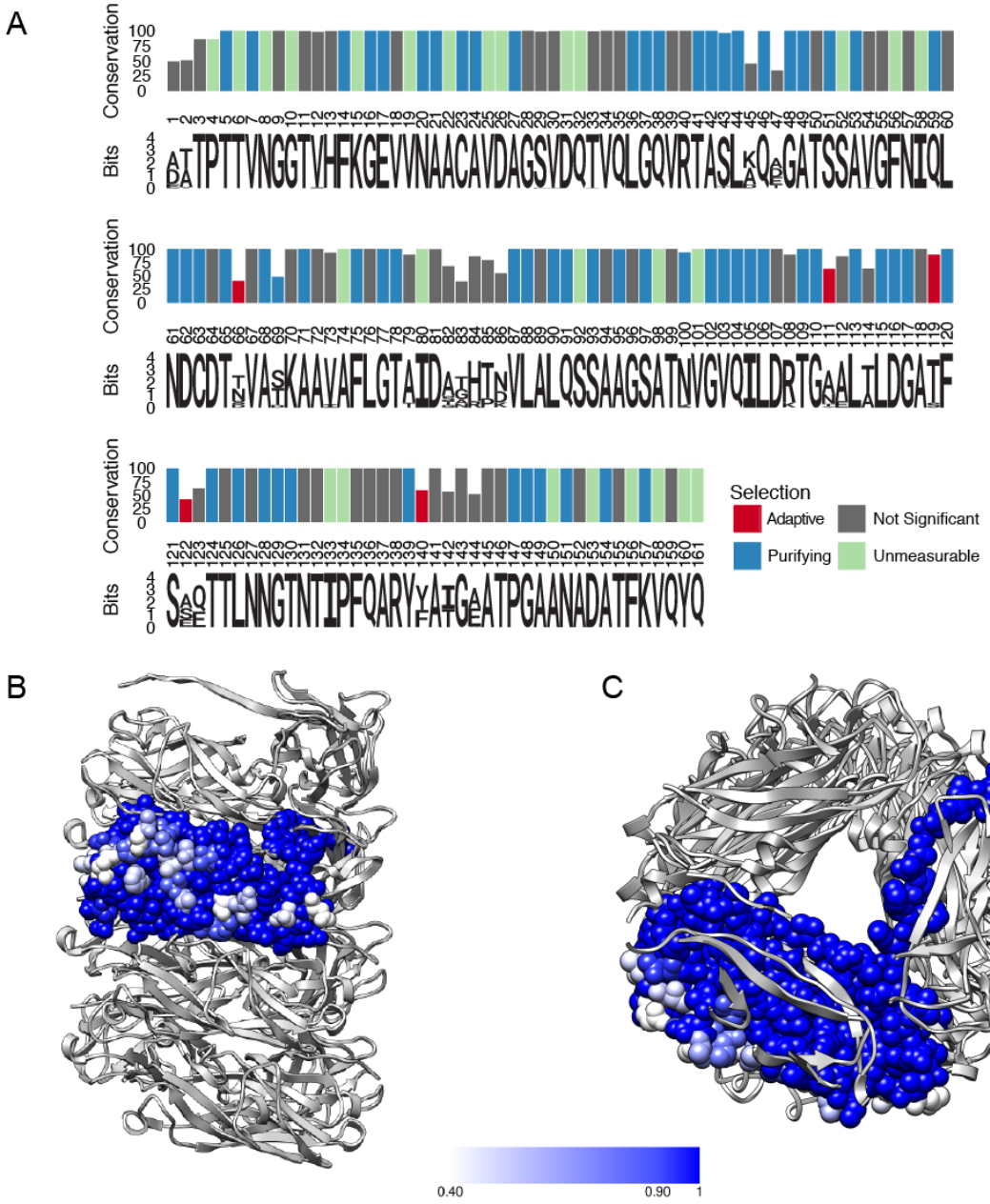
(A) An electron micrograph of type 1 pili in vitreous ice, surrounded by type 4 pili and flagellar filaments. (B) Side view, (C) interior view (the front half of the reconstruction has been removed) and (D) top view of overall reconstruction of FimA rod with subunits along the same left-handed 3-start helix colored in either blue, salmon or green. (E) Close-up view of donor strand complementation (DSC) in the central lumen showing that the β-strands in the reconstruction are very well resolved.

3.6.2 Figure 2. Subunit interface of the FimA rod.



(A) Surface view of FimA rod model with subunits numbered along the right-handed 1-start helix. Each subunit is in a different color. (B) Ribbon representation to illustrate the interface of subunit N_0 (in salmon) with neighboring subunits.

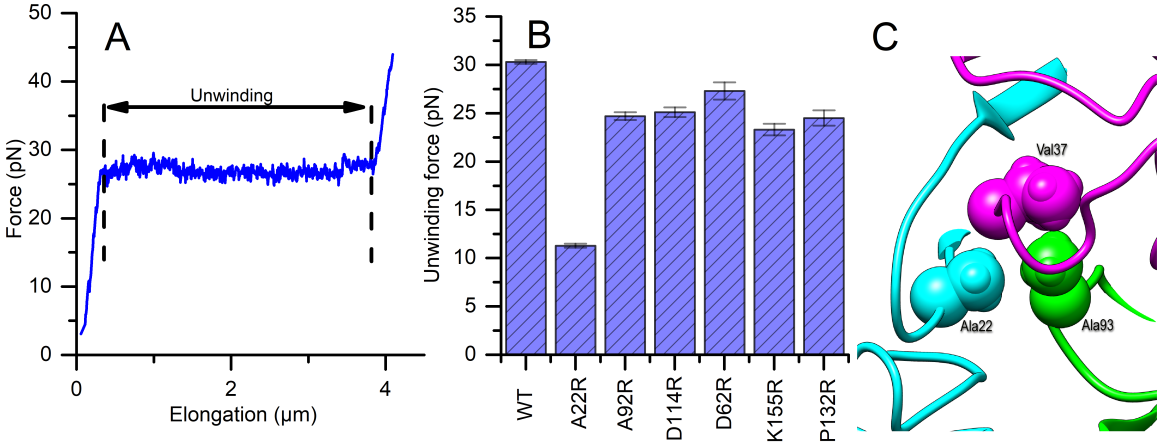
3.6.3 Figure 3. FimA conservation and variability.



(A) Conservation, consensus amino acid identity, and selection pressures on residues within the mature form of the FimA protein measured by alignment of 1,828 sequences. Numbering here is based on $FimA_{UTI89}$, which has two additional residues at the N-terminus of the mature protein relative to $FimA_{BW25113}$. Consensus amino acid residues are shown at each position. Height of the residues corresponds to their information content (bits) where the larger size indicates greater certainty that residue shown is the consensus residue at that position. The vertical bars represent the proportion of strains with the consensus amino acid (0-100%). Vertical bars are colored by selection pressure acting upon the amino acid position, with red and blue bars indicating residues with codons under adaptive and purifying selection, respectively. Green bars indicate codons with too little variability for evolutionary analysis and grey indicates amino acid positions that do not show statistically significant evidence for selection. (B,C) The degree of conservation for every FimA residue shown in (A) has been mapped onto a single subunit in our rod model. The absolutely conserved residues are in blue, the residues that are 40% conserved (the greatest degree of variability found) are in white, and light blue represents 90% conservation. Every residue facing the lumen is 100%

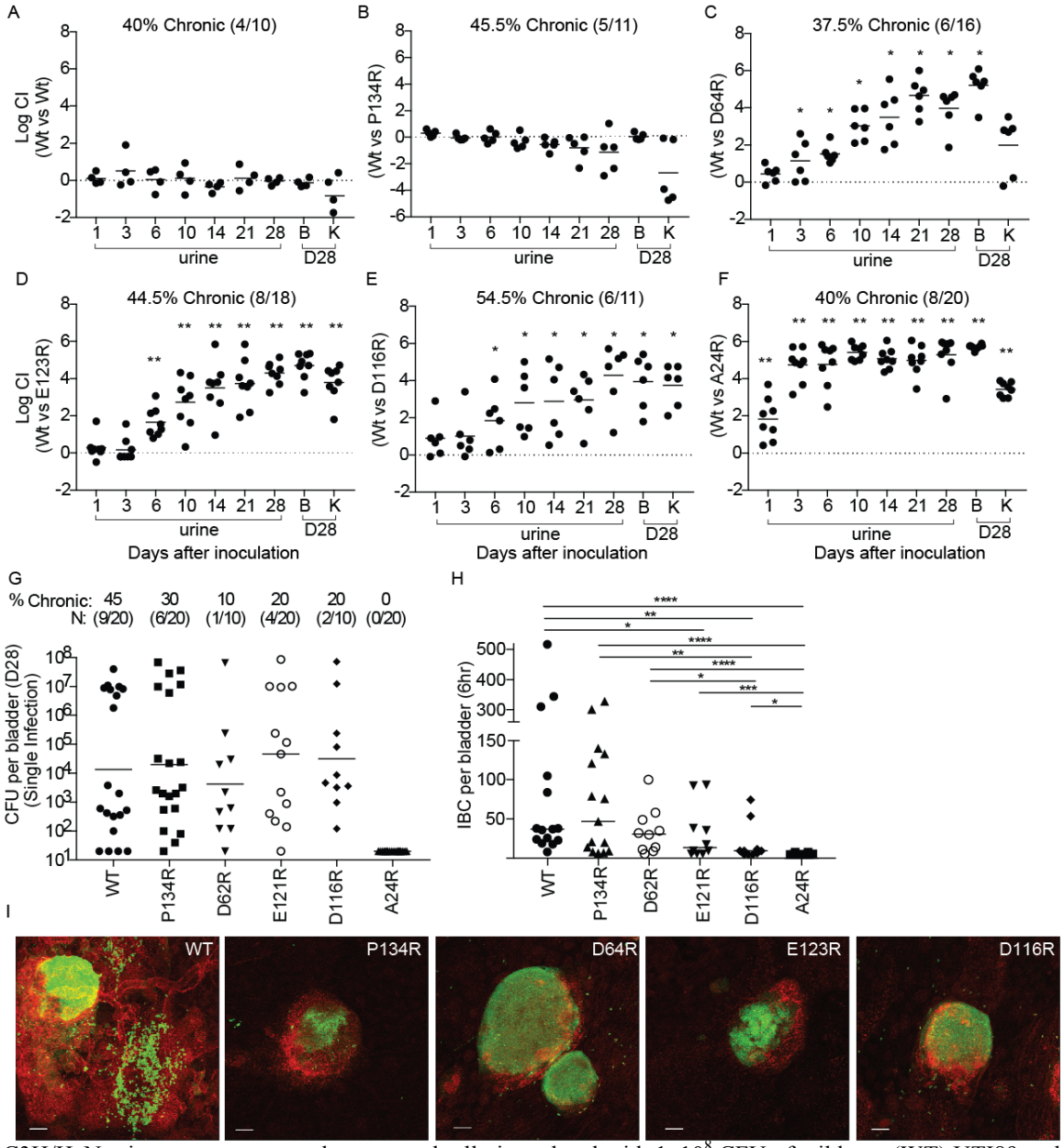
conserved. Overall, FimA displays strong signals of purifying selection resulting in conservation with hotspots of adaptive selection resulting in variability of surface residues

3.6.4 Figure 4. Mutations to FimA alter the force required to unwind the pilus.



(A) Force response of a single type 1 pilus. The force response is composed of three phases, elastic stretching of the shaft, unwinding of the shaft, and elastic stretching of individual subunits in an open coil. (B) Bar chart showing the average unwinding force of measured pili. The error bars represent the standard error of the mean. (C) Ala22 (cyan) plays an important role in the inter-subunit contacts through interacting with Val37 (magenta) from a second subunit and Ala93 (green) from a third subunit.

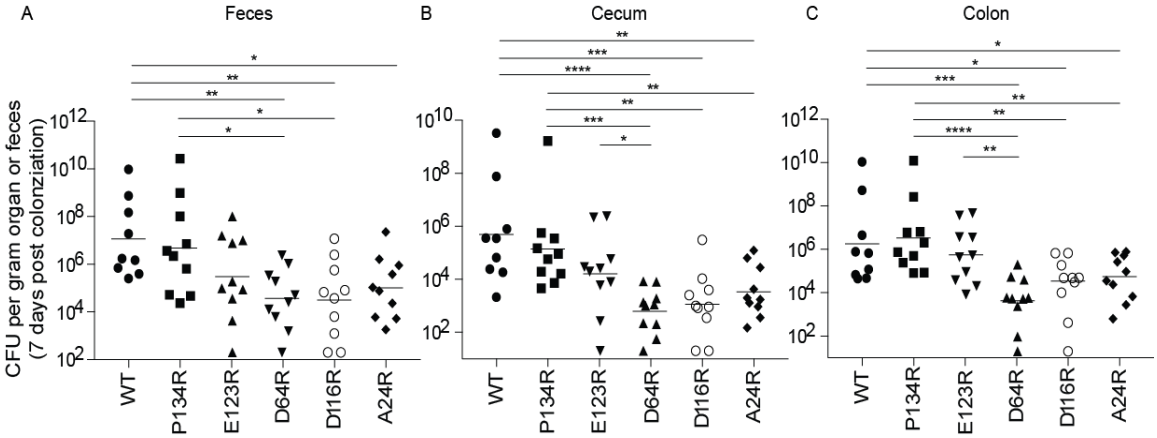
3.6.5 Figure 5. Point mutations in FimA alter UPEC pathogenesis in the bladder.



C3H/HeN mice were concurrently transurethrally inoculated with $1x10^8$ CFU of wildtype (WT) UTI89 and one of 5 isogenic UTI89 strains containing point mutations in the *fimA* gene. (A, B) Longitudinal urinalysis and examination of UTI89 titers in bladders and kidneys at time of sacrifice (28 dpi) revealed that, in chronically infected mice, the reintegrated FimA WT strain and P134R had no impact on the fitness of UTI89. (C-F) FimA proteins containing D64R, E123R, D116R, and A24R mutations had decreased fitness in chronically infected mice compared to the WT strain in competitive infections. (G) Mice infected singly, via transurethral inoculation, with D64R, E123R, D116R, and A24R mutations were also less likely to develop chronic UTI, as determined by longitudinal urinalysis and examination of UTI89 CFUs in bladders and kidneys at time of sacrifice (28 dpi). Bladders taken at 28 dpi after single infections are shown in panel G along with the percentage of mice that developed chronic UTI. (H) Number of IBCs formed at 6 hours post infection (hpi). (I) Immunofluencence images of IBCs at 12hpi for the indicated FimA mutant strains. No IBCs were detected in the A24R mutant at 12 hours. Abbreviations. B= bladder, K= kidney, CI= competitive index, IBC= intracellular bacterial community. Bars represent mean (A-F), geometric mean (G),

and median (H). *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 by Wilcoxon Signed Ranked test (A-F) or Mann Whitney U test (H). N=10, 2 replicates (A, G); N=11, 2 replicates (B, E); N=16, 3 replictates (C); N=18, 3 replicates (D); N=20 mice, 4 replicates (F, G); N=15, 3 replicates (H); N=4, 2 biological replicates (I). All replicates are biological.

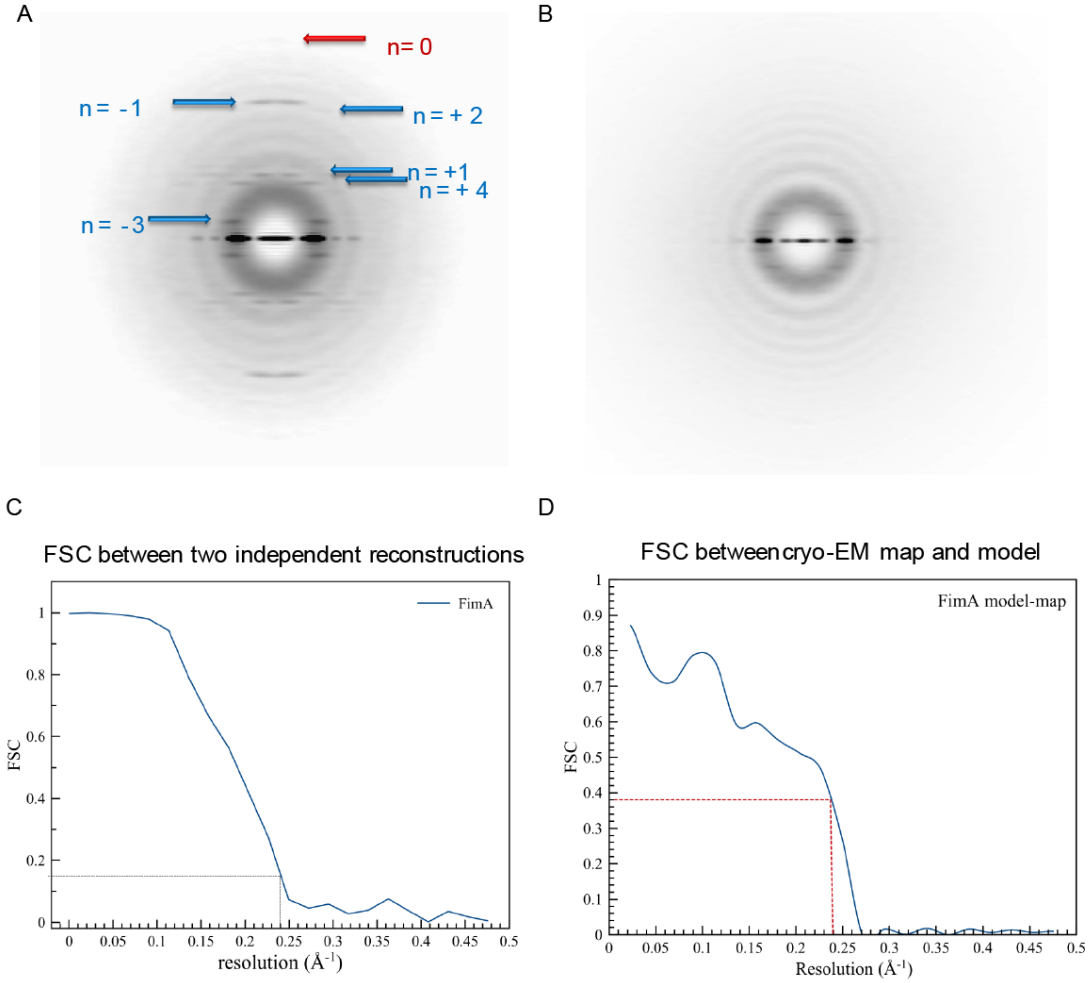
3.6.6 Figure 6. Mutations in FimA alter the ability of UTI89 to colonize the intestine.



C3H/HeN mice were orally gavaged with streptomycin, to disrupt colonization resistance, and one day later orally gavaged with $1x10^8$ CFU of WT UTI89 or one of 5 isogenic UTI89 strains with point mutations in *fimA*. Examination of CFU of each strain in the (A) feces, (B) cecum, or (C) colon of mice at 7 days post colonization revealed that strains expressing mutant FimA proteins (D116R, D64R, and A24R) had fitness defects compared to the WT strain. Bars represent geometric mean. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 by Mann Whitney U test. N=10, 2 biological replicates.

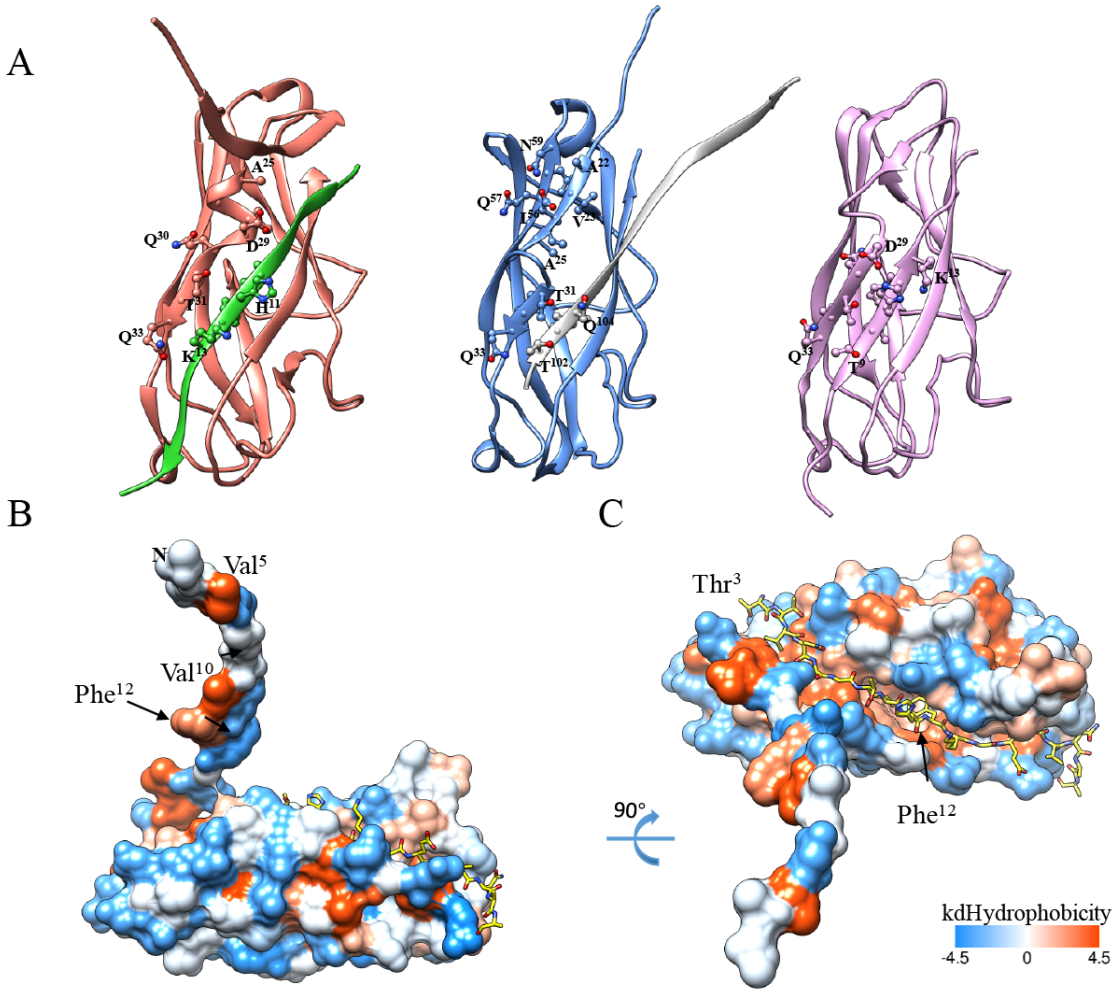
3.7 Supplemental Figures

3.7.1 Figure S1. Details of cryo-EM reconstruction of type 1 pili.



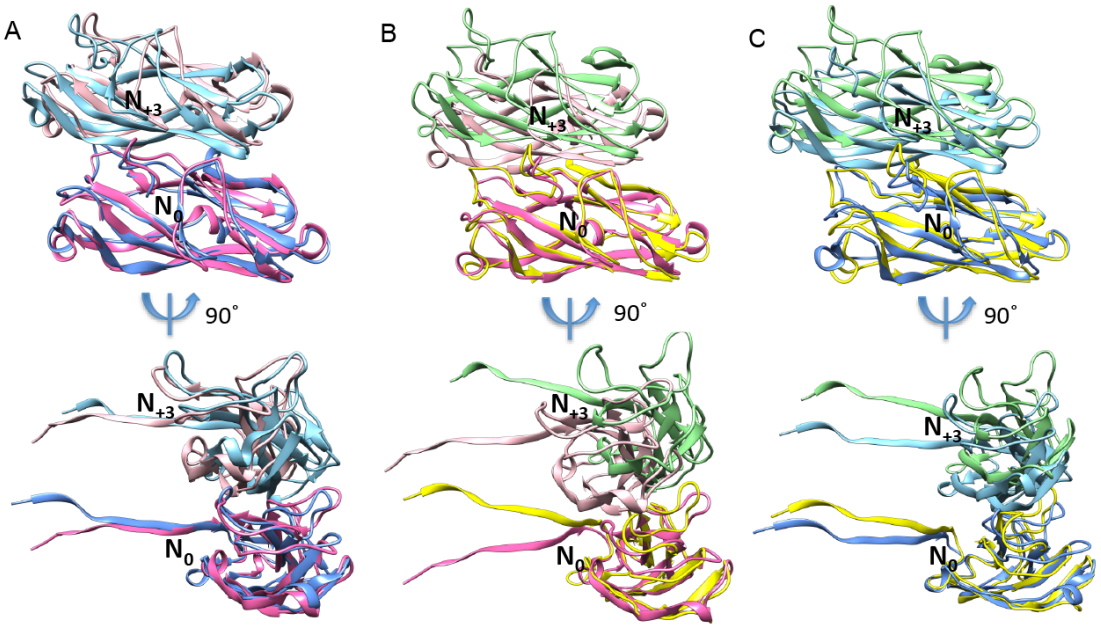
(A) Averaged power spectrum from the segments of type 1 pili, which shows the meridional at ~1/(7.5 Å) indicated by the red arrow. The blue arrow indicates layer lines containing Bessel orders +1, +2 ,+4, -1, and -3. (B) The power spectrum for the Type IV pili in the same preparation. (C) FSC plot generated from two independent half maps, each reconstructed from two independent dataset, shows a resolution of ~ 4.2 Å at FSC = 0.143. (D) FSC plot generated from the comparison of the cryo-EM map with the atomic model shows a resolution of ~4.2 Å at FSC = 0.38 (=$0.143^{1/2}$).

3.7.2 Figure S2. Nte inserts into the hydrophobic groove of the neighboring subunit.



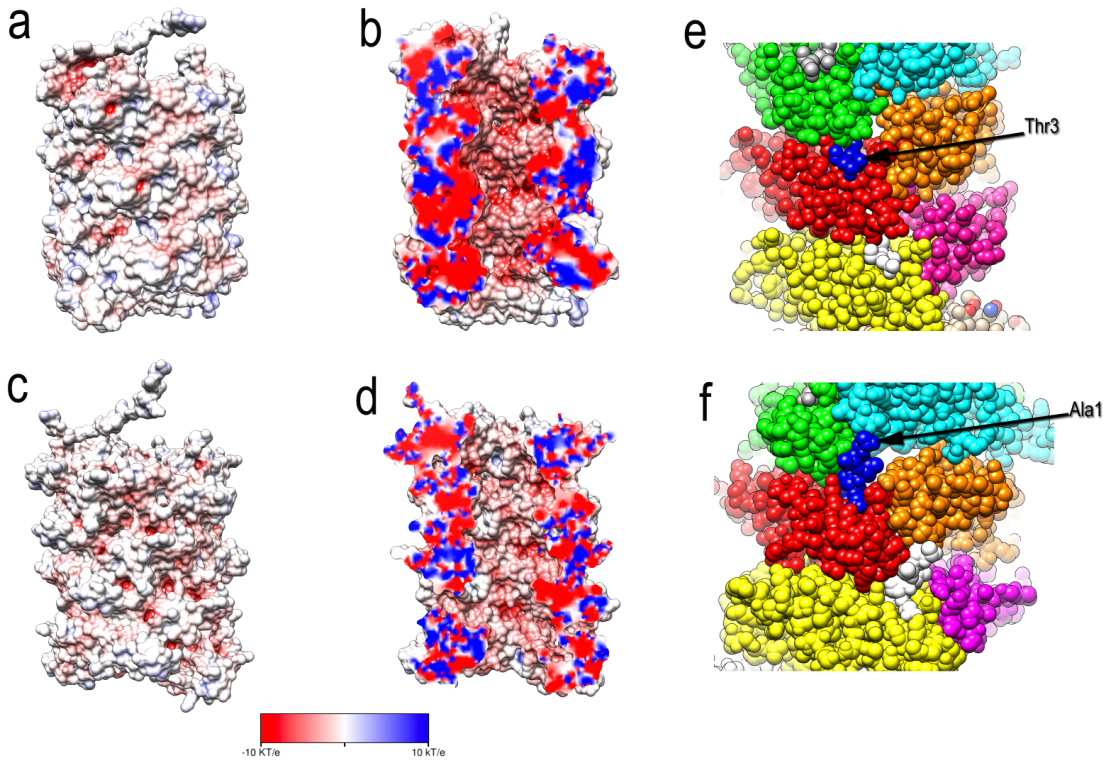
(A) Comparison of the DSC within the helical rod, within the FimA:FimC complex (PDB ID:3SQB), and in the self-complemented FimA monomer (PDB ID:2M5G). FimA in the helical rod is shown in salmon and Nte from the adjacent subunit is shown in green; FimA in FimA:FimC complex is shown in blue and FimC donor strand is shown in grey; FimA self-complemented monomer is shown in plum. (B,C) FimA rod subunit surface is colored according to hydrophobicity, with the adjacent subunit's Nte shown in a stick representation. Phe12 sits in a large hydrophobic pocket of the adjacent subunit.

3.7.3 Figure S3. Comparison of subunit N_0 and N_{+3} interfaces with previously deposited FimA pilus rod models.



(A) Our FimA pilus rod model and the deposited solid-state NMR FimA model 2N7H, when aligned with their N_0 subunit have an overall RMSD of 5.7 Å for subunits N_{+3}. (B) Our FimA pilus rod model and solid-state NMR FimA model 2MX3 when aligned with subunit N_0 have an overall RMSD of 11.0 Å for subunits N_{+3}. (C) FimA models 2N7H and 2MX3 when aligned with their N_0 subunit, the overall RMSD is 5.6 Å for subunit N_{+3}. The N_0 and N_{+3} subunits of our FimA model are colored in hot pink and pink, respectively. The N_0 and N_{+3} subunits of 2N7H are colored in dodger blue and light blue, respectively. The N_0 and N_{+3} subunits of 2MX3 are colored in yellow and green, respectively.

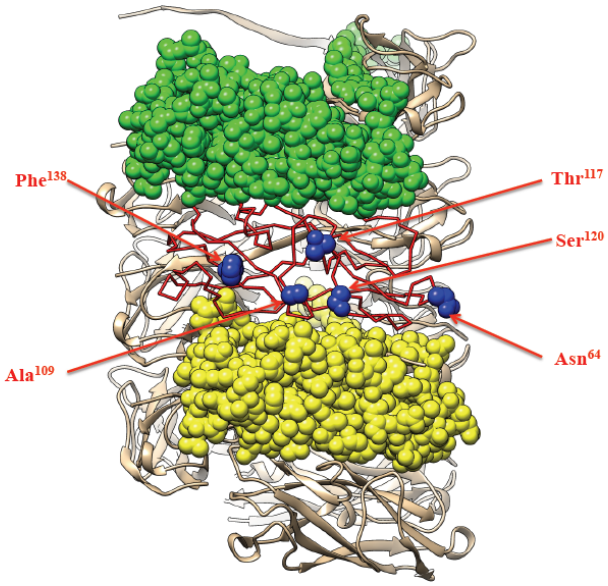
3.7.4 Figure S4. Comparison of type 1 pili and P pili.



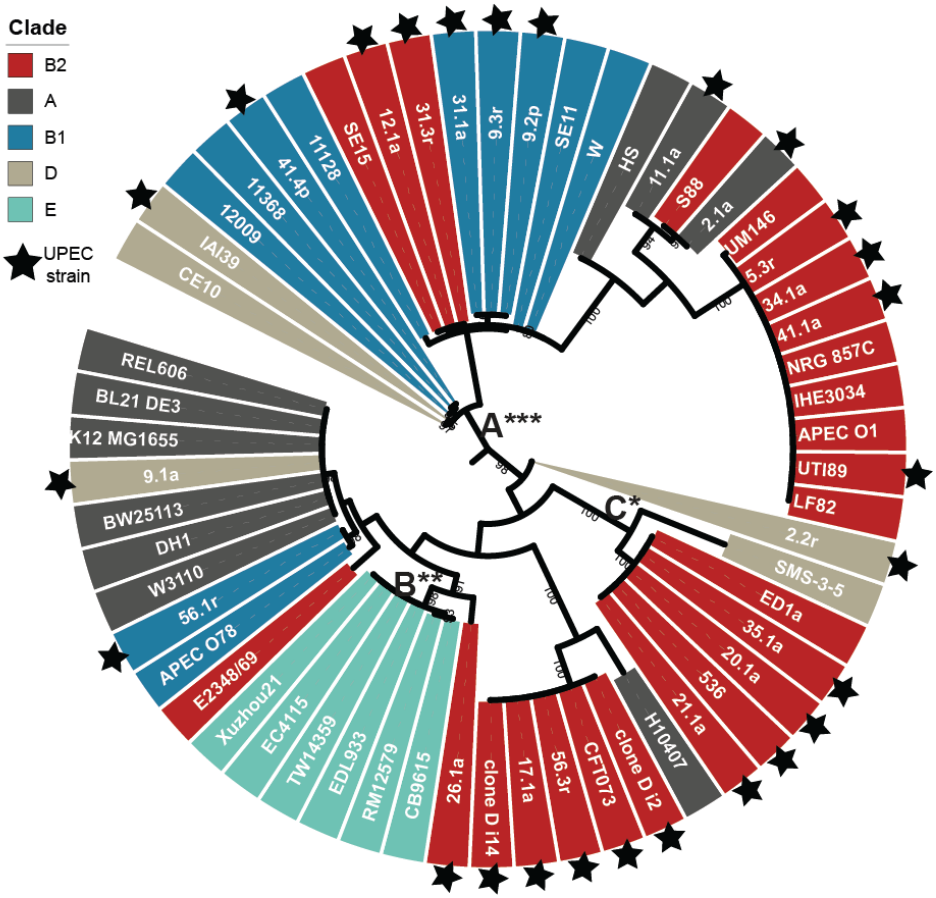
(A) The exterior and (B) central lumen of type 1 pili colored according to the electrostatic potential. (C) The exterior and (D) central lumen of P pili (PDB ID: 5FLU) colored according to the electrostatic potential. Red corresponds to a potential of -10 kT/e, and blue corresponds to a potential of 10 kT/e. (E,F) structures of type 1 and P pilus rods with each FimA subunit shown in a different color. (E) The N-terminus of FimA in the type 1 pilus projects out of the structure and makes no contacts with other subunits. (F) In contrast, the N-terminus in P pili (PDB 5FLU) forms a "staple" which involves contacts with two other subunits.

3.7.5 Figure S5. Location of FimA residues under positive selection for change in the helical rod.

A

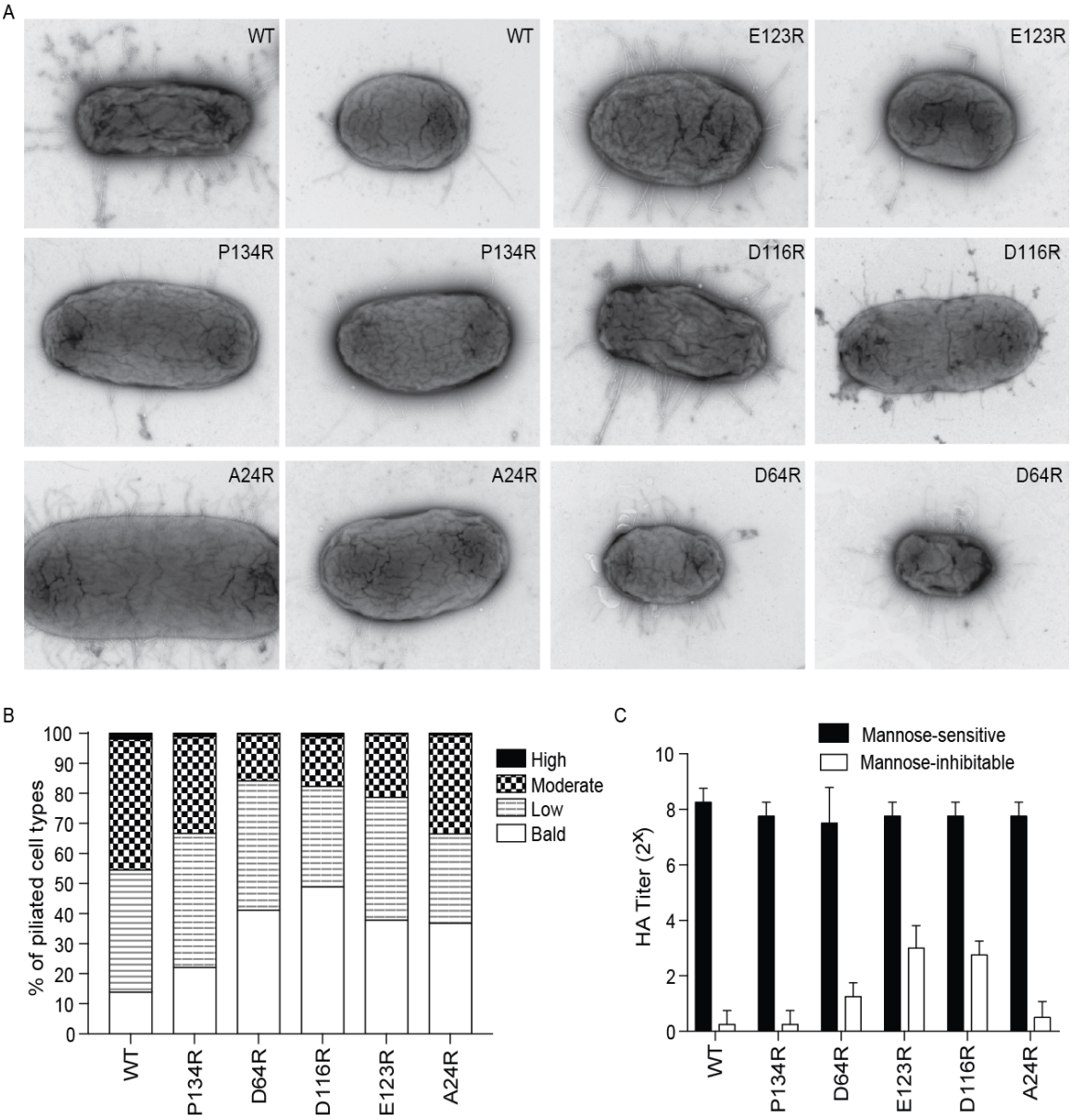


B



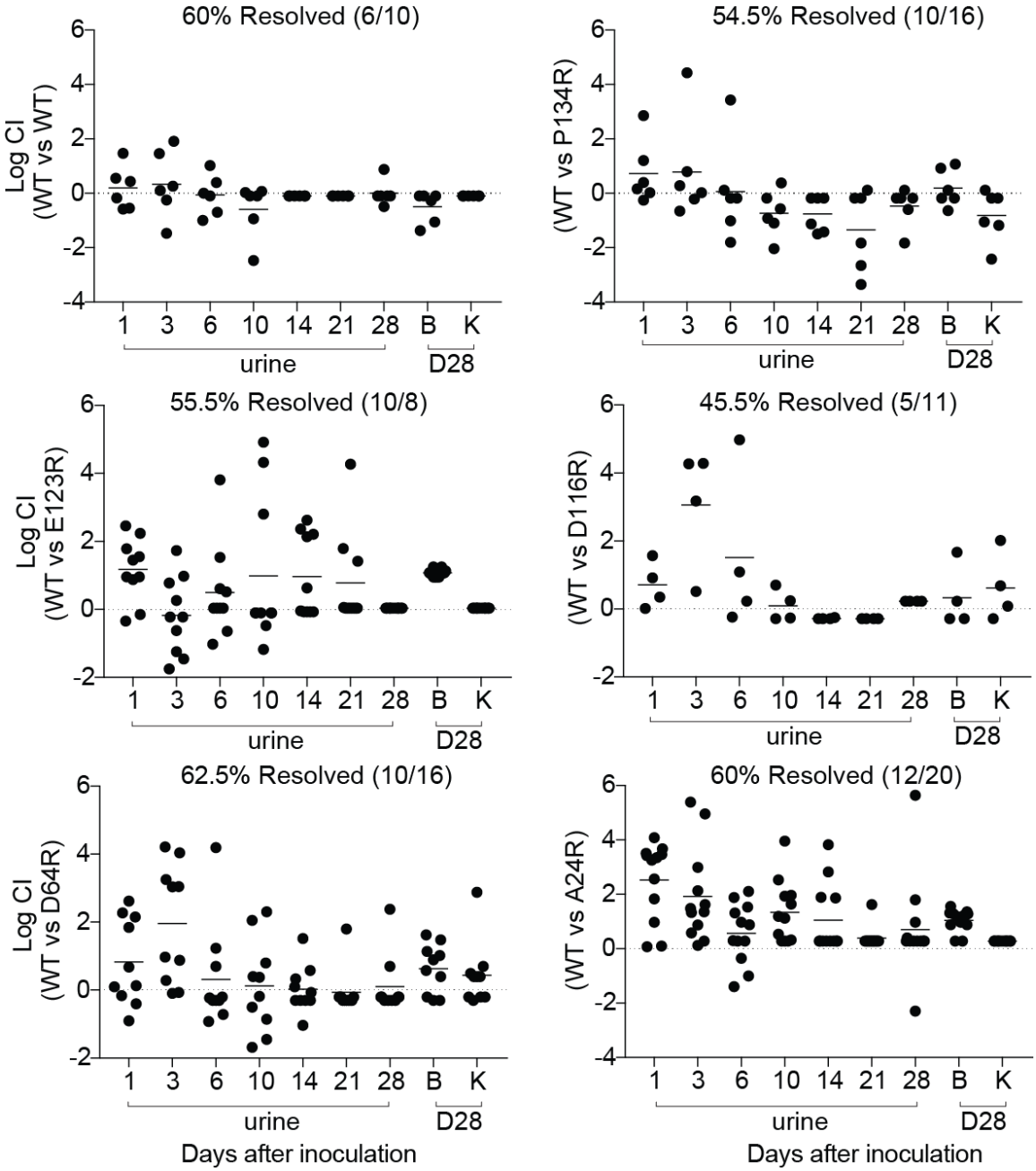
(A) Five residues in $FimA_{BW25113}$, Asn64, Ala109, Thr117, Ser120 and Phe138, under adaptive selection are all located on the outside of the pilus rod and are shown in blue spheres. (B) Phylogenetic analysis of the carriage and evolution of the *fimA* gene in a curated dataset of *E. coli* genomes. A total of 57 *fimA* sequences were aligned and used to construct a phylogenetic tree with bootstrap support (between 80 and 100%) indicated along internal branches. Taxon labels are colored according to their clade of origin and UPEC strains are indicated with a star. Measurement of the rates of non-synonymous (*dN*) to synonymous (*dS*) mutation for each branch identified three branches carrying many sequences from UPEC strains are under episodic, diversifying selection (labeled A, B, and C). Statistical significance for these measures are indicated by asterisks, with *$p \leq 0.05$, **$p \leq 0.01$, and ***$p < 0.001$ by likelihood ratio χ^2 (chi-squared) test.

3.7.6 Figure S6. Chromosomally integrated point mutations in *fimA* gene do not prevent the expression or function of type 1 pili *in vitro*.



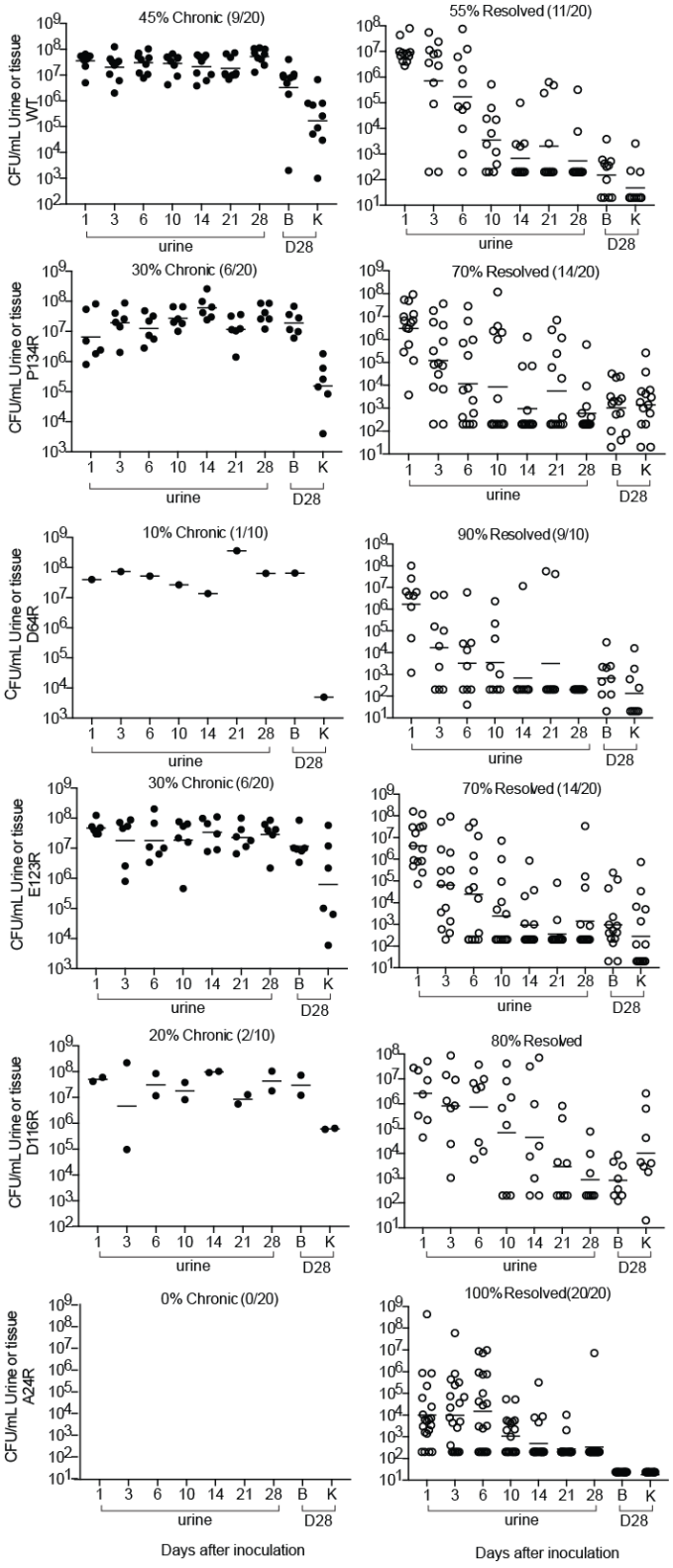
(A) Representative negative stain EM images of UTI89 strains producing WT FimA or one of the indicated mutant variants. (B) Percentage of bald, lowly-, moderately-, and highly-piliated cells in the different UTI89 populations. (C) The relative ability of the WT strain and each mutant to bind and agglutinate guinea pig red blood cells as assessed by hemagglutination titer analysis. Some mutants show increased binding to red blood cells in the presence of exogenous mannose compared to the WT strain, suggesting that some portion of the population of those strains may express other pilus types. Abbreviations: HA= hemagglutination assay. Bars represent mean ± SD (C). N= 600 cell counted per mutant, 2 biological replicates (B); N= 4 biological replicates (C).

3.7.7 Figure S7 CFU titers for mice that resolved competitive bladder infections shown in Figure 5.



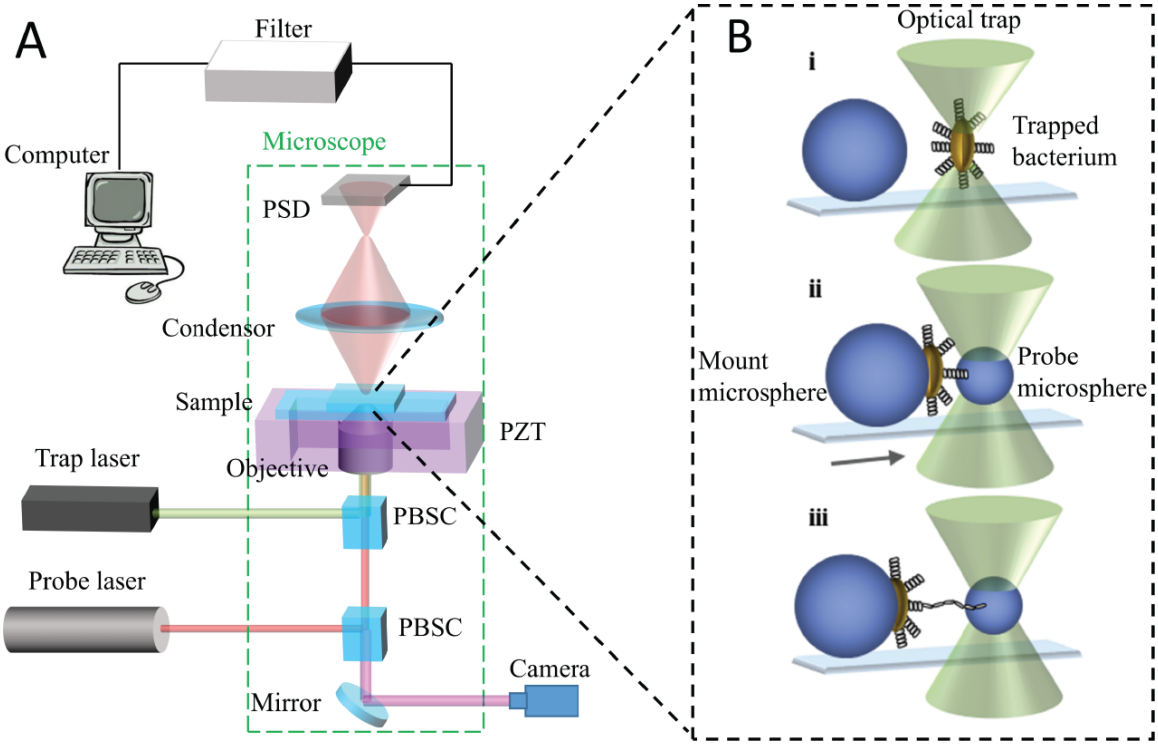
C3H/HeN mice were concurrently colonized, via transurethral inoculation, with $1x10^8$ CFU of wildtype (WT) UTI89 and one of 5 isogenic UTI89 strains containing point mutations in *fimA*. The resolution of infection in mice was determine by longitudinal urinalysis and examination of UTI89 titers in bladders and kidneys at time of sacrifice (28 dpi). The mice that resolved infection by both strains (percentage shown above each graph) are shown here. Chronically infected mice are shown in Figure 5. Abbreviations. B= bladder, K= kidney, CI= competitive index. Bars represent mean values.

3.7.8 Figure S8 CFU titers for mice that developed chronic UTI or resolved infection in single bladder infections shown in Figure UTI.



C3H/HeN mice were singly colonized, via transurethal inoculation, with $1x10^8$ CFU of wildtype (WT) UTI89 or one of 5 isogenic UTI89 strains containing point mutations in *fimA*. The status of infection (mice that were chronically infected or those that resolved infection) was determine by longitudinal urinalysis and examination of UTI89 titers in bladders and kidneys at time of sacrifice (28 dpi). Mice that developed chronic infection are shown on the left while those that resolved infection are shown on the right. Percentages of each outcome are given above each graph. Abbreviations. B= bladder, K= kidney. Bars represent geometric mean.

3.7.9 Figure S9. Schematic illustration of the optical tweezer setup



(A) The trap and probe laser beams are merged using a polarizing beam splitter cube (PBSC) positioned inside the microscope. The two laser beams are thereafter focused by the objective inside the sample chamber. The weak probe laser beam (μW), refracted by the object and collected by the condenser, is illuminating a PSD-detector that converts the incoming light to a photocurrent, which in turn is converted to a voltage signal. (B) Schematic illustration of a force spectroscopy experiment. A single cell or probe microsphere can be trapped by the laser. i) A single bacterium is mounted onto a large immobilized microsphere. ii) A probe microsphere is trapped and brought into contact with bacterial pili. iii) The bacterium and trapped microsphere are separated and an attached pilus is unwound.

3.8 Supplemental Tables

3.8.1 Table S1. Validation statistics for FimA model

Phenix Real-Space Refinement	
Map CC (around atoms)	0.68
MolProbity Score	**1.90**
Ramachandran outliers	0.00%
Ramachandran favored	92.21%
Ramachandran allowed	7.8%
Poor rotamers	0
RMSD bonds (Å)	0.00
RMSD bond angles (°)	0.85
All atom clashscore	7.75

3.8.2 Table S2. Helical parameters comparison within FimA models

Rotation(°)/ Translation(Å)	FimA cryo-EM model	Solid NMR FimA model PDB code ID:2N7H	Solid NMR FimA model PDB code ID:2MX3
N-N+1	115.0/7.7	111.5/7.2	111.5/8.7
N-N+2	130.0/15.4	137.0/14.4	137.0/17.5
N-N+3	15.1/23.0	25.5/21.6	25.5/26.2
N-N+4	99.9/30.7	86.0/28.8	86.0/34.9

3.8.3 Table S3. FimA sequences used in analysis of conservation and selection

Available in Appendix A

3.8.4 Table S4. List of FimC and FimH sequences

Available in Appendix A

3.8.5 Table S5. Codon-by-codon selection analysis in *fimA*

UTI89 Codon Position	AA Identity[a]	*dS*	*dN*	*dN/dS*[b]	Normalized *dN-dS*	dS (when dN=dS)	Log(L)	LRT	p-value[c]
1	49.18	0.063	1.639	25.906	1.092	1.193	-48.304	2.683	0.101
2	51.04	2.355	1.548	0.657	-0.559	1.818	-67.600	0.619	0.431
3	85.83	1.038	0.000	0.000	-0.719	0.264	-8.200	2.600	0.107
4	85.83	0.000	0.000	Undefined	0.000	0.000	-4.393	0.000	1.000
5	100.00	1.026	0.000	0.000	-0.711	0.224	-17.754	6.074	0.014
6	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
7	99.89	1.659	0.307	0.185	-0.936	0.738	-47.275	4.549	0.033
8	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
9	99.95	1.120	0.211	0.188	-0.630	0.538	-28.763	2.456	0.117
10	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
11	99.89	0.748	0.169	0.226	-0.401	0.373	-27.923	1.598	0.206
12	98.47	0.750	0.305	0.407	-0.308	0.434	-28.926	0.770	0.380
13	99.84	0.000	0.153	Infinite	0.106	0.099	-13.033	0.859	0.354
14	100.00	0.524	0.000	0.000	-0.363	0.098	-10.881	3.314	0.069
15	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
16	100.00	1.023	0.000	0.000	-0.709	0.298	-18.065	4.918	0.027
17	100.00	1.435	0.000	0.000	-0.994	0.232	-13.093	7.244	0.007
18	99.67	0.349	0.149	0.426	-0.139	0.208	-14.583	0.351	0.554
19	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
20	100.00	1.758	0.000	0.000	-1.218	0.467	-26.330	13.041	0.000
21	100.00	0.962	0.000	0.000	-0.666	0.418	-25.648	6.597	0.010
22	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
23	100.00	1.902	0.000	0.000	-1.317	0.589	-22.233	11.502	0.001
24	99.84	2.663	0.186	0.070	-1.716	1.090	-44.447	10.719	0.001
25	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
26	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
27	99.73	2.328	0.375	0.161	-1.353	1.094	-49.520	6.361	0.012
28	100.00	0.231	0.000	0.000	-0.160	0.110	-8.401	1.486	0.223
29	98.91	0.000	0.163	Infinite	0.113	0.111	-12.740	0.765	0.382
30	99.51	0.356	0.153	0.429	-0.141	0.213	-19.701	0.345	0.557
31	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
32	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
33	99.45	0.703	0.287	0.409	-0.288	0.446	-34.211	0.971	0.324
34	99.95	0.000	0.152	Infinite	0.106	0.106	-11.034	0.722	0.396
35	99.95	0.000	0.182	Infinite	0.126	0.151	-13.483	0.381	0.537
36	99.95	5.523	0.154	0.028	-3.719	1.592	-64.866	26.659	0.000
37	100.00	5.968	0.000	0.000	-4.135	2.022	-62.762	29.334	0.000
38	100.00	0.902	0.000	0.000	-0.625	0.151	-9.561	3.555	0.059
39	99.95	0.364	0.152	0.418	-0.147	0.215	-18.645	0.365	0.546
40	99.95	0.000	0.196	Infinite	0.136	0.126	-12.829	0.889	0.346
41	100.00	0.468	0.000	0.000	-0.325	0.177	-14.348	3.873	0.049
42	100.00	2.194	0.000	0.000	-1.520	0.738	-28.474	12.786	0.000

43	96.66	3.962	1.117	0.282	-1.970	1.930	-75.585	4.501	0.034
44	100.00	1.057	0.000	0.000	-0.733	0.343	-21.097	6.714	0.010
45	45.35	1.856	1.558	0.840	-0.206	1.639	-62.743	0.063	0.801
46	99.84	3.736	0.532	0.143	-2.219	1.042	-37.672	5.923	0.015
47	33.92	1.737	1.948	1.121	0.146	1.897	-69.047	0.028	0.867
48	100.00	0.587	0.000	0.000	-0.407	0.249	-14.333	3.420	0.064
49	100.00	1.565	0.000	0.000	-1.084	0.586	-28.212	9.789	0.002
50	100.00	0.228	0.000	0.000	-0.158	0.088	-11.113	1.908	0.167
51	100.00	0.716	0.000	0.000	-0.496	0.198	-12.678	5.088	0.024
52	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
53	100.00	2.502	0.000	0.000	-1.733	0.758	-26.816	13.594	0.000
54	99.29	0.490	0.152	0.311	-0.234	0.282	-21.021	0.972	0.324
55	99.95	0.000	0.210	Infinite	0.146	0.131	-11.593	0.940	0.332
56	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
57	100.00	0.287	0.000	0.000	-0.199	0.088	-11.240	2.349	0.125
58	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
59	100.00	4.015	0.000	0.000	-2.782	0.572	-21.855	15.131	0.000
60	100.00	0.336	0.000	0.000	-0.233	0.113	-13.037	2.186	0.139
61	100.00	0.497	0.000	0.000	-0.344	0.101	-8.123	3.156	0.076
62	100.00	0.548	0.000	0.000	-0.380	0.124	-8.273	2.947	0.086
63	100.00	0.961	0.000	0.000	-0.665	0.334	-17.194	6.239	0.012
64	99.89	0.494	0.498	1.009	0.003	0.497	-25.053	0.000	0.994
65	100.00	0.997	0.000	0.000	-0.691	0.363	-22.701	8.013	0.005
66	40.54	0.000	1.742	Infinite	1.207	1.275	-50.432	6.736	0.009
67	99.95	0.000	0.152	Infinite	0.105	0.106	-10.863	0.721	0.396
68	100.00	0.873	0.000	0.000	-0.605	0.340	-19.302	5.632	0.018
69	48.30	2.378	0.805	0.339	-1.089	1.300	-62.067	3.135	0.077
70	99.78	0.000	0.117	Infinite	0.081	0.100	-10.329	0.324	0.569
71	100.00	2.329	0.000	0.000	-1.613	0.901	-39.485	14.642	0.000
72	99.89	0.732	0.375	0.511	-0.248	0.495	-29.302	0.431	0.512
73	93.11	0.000	0.296	Infinite	0.205	0.209	-13.110	1.388	0.239
74	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
75	100.00	0.997	0.000	0.000	-0.690	0.266	-18.025	7.809	0.005
76	99.84	1.451	0.475	0.327	-0.676	0.750	-42.442	1.302	0.254
77	100.00	1.610	0.000	0.000	-1.115	0.548	-21.591	8.426	0.004
78	100.00	2.981	0.000	0.000	-2.065	0.750	-31.191	18.929	0.000
79	89.55	0.995	0.760	0.764	-0.163	0.846	-38.418	0.120	0.729
80	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
81	99.78	0.807	0.161	0.199	-0.447	0.344	-16.019	1.836	0.175
82	67.94	2.253	1.833	0.814	-0.291	1.964	-68.527	0.118	0.731
83	39.55	1.328	2.858	2.152	1.060	2.438	-91.627	1.439	0.230
84	86.27	0.000	0.630	Infinite	0.436	0.479	-26.950	2.189	0.139
85	79.27	1.395	0.632	0.453	-0.529	0.877	-38.898	1.213	0.271
86	54.70	2.053	1.088	0.530	-0.668	1.318	-59.849	0.806	0.369
87	99.95	1.561	0.152	0.098	-0.976	0.617	-38.208	6.431	0.011
88	100.00	1.761	0.000	0.000	-1.220	0.568	-27.461	11.210	0.001
89	99.95	0.362	0.186	0.514	-0.122	0.246	-15.404	0.215	0.643
90	100.00	1.745	0.000	0.000	-1.209	0.571	-29.423	11.064	0.001
91	99.73	1.866	0.180	0.096	-1.168	0.451	-22.969	3.755	0.053
92	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
93	100.00	2.664	0.000	0.000	-1.845	0.837	-38.287	15.569	0.000
94	99.95	0.356	0.186	0.522	-0.118	0.244	-20.268	0.206	0.650
95	99.73	2.793	0.378	0.135	-1.674	1.084	-43.630	7.000	0.008

96	99.95	1.156	0.211	0.182	-0.655	0.543	-27.518	2.564	0.109
97	99.89	1.257	0.138	0.110	-0.776	0.477	-33.071	5.127	0.024
98	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
99	99.95	0.301	0.144	0.478	-0.109	0.195	-16.761	0.265	0.607
100	93.44	3.100	0.647	0.209	-1.699	1.243	-63.405	7.075	0.008
101	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
102	100.00	1.515	0.000	0.000	-1.049	0.538	-21.654	8.137	0.004
103	100.00	0.507	0.000	0.000	-0.352	0.117	-10.753	2.931	0.087
104	99.95	4.058	0.180	0.044	-2.686	0.760	-32.143	10.069	0.002
105	100.00	2.024	0.000	0.000	-1.402	0.581	-34.746	16.994	0.000
106	100.00	1.062	0.000	0.000	-0.736	0.343	-23.299	6.755	0.009
107	99.95	0.292	0.154	0.527	-0.096	0.202	-16.447	0.200	0.655
108	89.22	0.000	0.506	Infinite	0.351	0.404	-29.254	1.786	0.181
109	99.95	1.875	0.137	0.073	-1.204	0.605	-35.536	8.274	0.004
110	100.00	6.311	0.000	0.000	-4.372	2.000	-61.316	31.635	0.000
111	63.02	0.000	1.959	Infinite	1.357	1.313	-46.972	8.616	0.003
112	86.27	1.071	1.117	1.042	0.031	1.105	-47.609	0.003	0.960
113	99.95	1.004	0.163	0.162	-0.582	0.439	-25.758	2.954	0.086
114	63.35	0.971	1.512	1.556	0.374	1.370	-55.605	0.341	0.559
115	100.00	0.660	0.000	0.000	-0.457	0.218	-15.465	4.421	0.036
116	100.00	2.610	0.000	0.000	-1.808	0.767	-28.083	16.959	0.000
117	100.00	2.377	0.000	0.000	-1.646	0.792	-39.861	12.731	0.000
118	99.95	0.000	0.181	Infinite	0.125	0.131	-8.948	0.635	0.426
119	89.33	0.000	0.591	Infinite	0.410	0.381	-29.730	3.465	0.063
120	100.00	0.839	0.000	0.000	-0.581	0.182	-12.807	6.039	0.014
121	100.00	1.000	0.000	0.000	-0.693	0.208	-12.444	6.190	0.013
122	42.40	0.638	2.456	3.849	1.259	1.743	-72.544	3.788	0.052
123	62.69	0.000	0.878	Infinite	0.609	0.714	-40.843	2.468	0.116
124	99.95	1.193	0.000	0.000	-0.827	0.376	-24.671	9.209	0.002
125	99.89	0.697	0.137	0.197	-0.388	0.345	-24.931	2.348	0.125
126	100.00	1.323	0.000	0.000	-0.917	0.436	-26.864	8.823	0.003
127	99.95	0.521	0.123	0.235	-0.276	0.198	-17.269	0.955	0.328
128	99.95	0.884	0.123	0.139	-0.528	0.346	-23.107	3.503	0.061
129	100.00	2.102	0.000	0.000	-1.456	0.769	-28.146	11.775	0.001
130	99.84	1.233	0.137	0.111	-0.760	0.528	-31.159	5.666	0.017
131	100.00	0.280	0.000	0.000	-0.194	0.085	-9.904	2.376	0.123
132	99.56	0.446	0.414	0.928	-0.022	0.426	-33.992	0.007	0.935
133	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
134	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
135	99.95	0.274	0.117	0.426	-0.109	0.164	-19.965	0.351	0.554
136	99.84	0.000	0.175	Infinite	0.121	0.144	-12.891	0.381	0.537
137	100.00	0.492	0.000	0.000	-0.341	0.130	-9.276	2.655	0.103
138	99.95	0.000	0.188	Infinite	0.130	0.121	-12.074	0.888	0.346
139	100.00	0.499	0.000	0.000	-0.346	0.119	-10.979	2.833	0.092
140	58.97	0.000	0.884	Infinite	0.613	0.683	-38.165	3.121	0.077
141	99.95	0.271	0.178	0.657	-0.064	0.215	-18.032	0.087	0.768
142	56.40	0.808	0.512	0.633	-0.206	0.607	-40.078	0.342	0.559
143	99.51	1.676	0.617	0.368	-0.734	1.015	-50.010	1.884	0.170
144	52.08	1.148	1.891	1.646	0.514	1.609	-60.516	0.583	0.445
145	100.00	0.276	0.000	0.000	-0.191	0.108	-10.155	1.875	0.171
146	99.89	0.000	0.137	Infinite	0.095	0.084	-11.525	0.977	0.323
147	99.62	11.326	0.337	0.030	-7.613	2.667	-81.398	40.731	0.000
148	100.00	5.532	0.000	0.000	-3.832	1.860	-46.874	29.087	0.000

149	99.95	5.407	0.178	0.033	-3.622	1.837	-68.581	23.781	0.000
150	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
151	100.00	0.527	0.000	0.000	-0.365	0.099	-8.505	3.327	0.068
152	99.95	0.486	0.178	0.366	-0.213	0.261	-17.902	0.482	0.488
153	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
154	100.00	1.333	0.000	0.000	-0.924	0.381	-18.178	7.466	0.006
155	99.95	0.000	0.137	Infinite	0.095	0.084	-11.524	0.977	0.323
156	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
157	100.00	0.849	0.000	0.000	-0.588	0.112	-10.636	4.027	0.045
158	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
159	99.95	0.000	0.175	Infinite	0.121	0.144	-12.900	0.381	0.537
160	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000
161	100.00	0.000	0.000	Undefined	0.000	0.000	0.000	0.000	1.000

[a]Amino acid position in alignment of 1,828 mature FimA sequences in Table S3

[b]Values of "Undefined" indicate codons where *dS* and *dN* = 0, preventing estimation of selection pressure (p-value = 1). Values of "Infinite" indicate codons where *dS* = 0 and *dN* >0

[c]p-values <0.1 are considered significant and are indicated in red text

3.8.6 Reference and UPEC strains used in analysis of *fimA* carriage

Strain	Patvhotype	Clade	BioProject (NCBI ID)	FimA Carriage
536	UPEC	B2	PRJNA16235	Yes
11128	EHEC	B1	PRJDA32513	Yes
11368	EHEC	B1	PRJDA32509	Yes
12009	EHEC	B1	PRJDA32511	Yes
55989	EAEC	B1	PRJNA33413	
2009EL-2050	EHEC	B1	PRJNA81097	
2009EL-2071	EHEC	B1	PRJNA81099	
2011C-3493	EHEC	B1	PRJNA81095	
ABU 83972	ABU	B2	PRJNA38725	
APEC O1	APEC	B2	PRJNA16718	Yes
APEC O78	APEC	B1	PRJNA184588	Yes
ATCC 8739	Commensal	A	PRJNA18083	
BL21(DE3)	Lab Strain	A	PRJNA20713	Yes
BW2952	Lab Strain	A	PRJNA33775	Yes
CB9615	EPEC	E	PRJNA42729	Yes
CE10	NMEC	D	PRJNA63597	Yes
CFT073	UPEC	B2	PRJNA313	Yes
clone D i14	UPEC	B2	PRJNA52023	Yes
clone D i2	UPEC	B2	PRJNA52021	Yes
DH1	Lab Strain	A	PRJDA52077	Yes
DH10B	Lab Strain	A	PRJNA20079	
E2348/69	EPEC	B2	PRJEA32571	Yes
E24377A	ETEC	B1	PRJNA13960	
EC4115	EHEC	E	PRJNA27739	Yes
ED1a	Commensal	B2	PRJNA33409	Yes
EDL933	EHEC	E	PRJNA259	Yes
H10407	ETEC	A	PRJEA42749	Yes
HS	Commensal	A	PRJNA13959	Yes
IAI39	UPEC	D	PRJNA33411	Yes
IHE3034	NMEC	B2	PRJNA43693	Yes
K12 MG1655	Commensal	A	PRJNA40075	Yes
LF82	AIEC	B2	PRJNA33825	Yes
NRG 857C	AIEC	B2	PRJNA41221	Yes
REL606	Lab Strain	A	PRJNA18281	Yes
RM12579	EPEC	E	PRJNA68245	Yes
S88	Commensal	B2	PRJNA33375	Yes
SE11	Commensal	B1	PRJNA18057	Yes
SE15	Commensal	B2	PRJDA19053	Yes
SMS-3-5	Environmental	D	PRJNA19469	Yes
TW14359	EHEC	E	PRJNA30045	Yes
UM146	AIEC	B2	PRJNA50883	Yes
UMNK88	ETEC	A	PRJNA42137	

UTI89	UPEC	B2	PRJNA16259	Yes
W	Lab Strain	B1	PRJNA48011	Yes
W3110	Lab Strain	A	PRJNA16351	Yes
Xuzhou21	EHEC	E	PRJNA45823	Yes
2.1a	UPEC	A	PRJNA269984	Yes
2.2r	UPEC	D	PRJNA269984	Yes
5.1a	UPEC	B1	PRJNA269984	
5.3r	UPEC	B2	PRJNA269984	Yes
9.1a	UPEC	D	PRJNA269984	Yes
9.2p	UPEC	B1	PRJNA269984	Yes
9.3r	UPEC	B1	PRJNA269984	Yes
11.1a	UPEC	A	PRJNA269984	Yes
12.1a	UPEC	B2	PRJNA269984	Yes
17.1a	UPEC	B2	PRJNA269984	Yes
20.1a	UPEC	B2	PRJNA269984	Yes
21.1a	UPEC	B2	PRJNA269984	Yes
26.1a	UPEC	B2	PRJNA269984	Yes
31.1a	UPEC	B1	PRJNA269984	Yes
31.3r	UPEC	B2	PRJNA269984	Yes
34.1a	UPEC	B2	PRJNA269984	Yes
35.1a	UPEC	B2	PRJNA269984	Yes
41.1a	UPEC	B2	PRJNA269984	Yes
41.4p	UPEC	B1	PRJNA269984	Yes
56.1a	UPEC	B1	PRJNA269984	Yes
56.3r	UPEC	B2	PRJNA269984	Yes

3.8.7 Table S7. *In vivo* and *in vitro* phenotypes of *fimA* mutants

Position (in $FimA_{BW25113}$)	AA Conservation	*dN/dS*[b]	p-value[c]	$FimA_{BW25113}$[d]	HA in PBS[e]	HA in Mannose[e]	Clumping Phenotype	Weaker Uncoiling	$FimA_{UTI89}$[f]	Ha in PBS[e]	HA in Mannose[e]	IBC Formation Defect	Bladder Colonization Defect	Gut Colonization Defect
-	-	-	-	WT	6	1			WT	8	0			
7 (5)	99.89	0	0.014	V5R	6	1	Yes							
24 (22)	99.84	0.07	0.001	A22R	6	1		Yes	A24R	8	0	Yes	Yes	Yes
27 (25)	99.73	0.161	0.012	A25R	3	1								
34 (32)	99.95	Infinite	0.396	V32R	1	1								
47 (45)	33.92	1.121	0.867	E45R	6	4	Yes							
64 (62)	99.89	1.009	0.994	D62R	5	3	Slight		D64R	7	1		Yes	Yes
67 (65)	99.95	Infinite	0.396	V65R	1	1								
87 (85)	99.95	0.098	0.011	V85R	1	1								
94 (92)	99.95	0.522	0.650	A92R	7	1		Yes						
116 (114)	100	0	0.000	D114R	6	3	Slight	Yes	D116R	8	3	Yes	Yes	Yes
123 (121)	62.69	Infinite	0.116	E121R	7	4	Yes		E123R	8	3	Yes	Yes	Slight
134 (132)	100	Undefined	1.000	P132R	6	1		Yes	P134R	8	0	Slight		
144 (142)	52.08	1.646	0.445	A142R	5	4	Yes							
147 (145)	99.62	0.03	0.000	P145R	0	0								
157 (155)	100	0	0.045	K155E	6	1		Yes						

[a]Codon positions are based on the global alignment shown in Figure 3a, which is the same as $FimA_{UTI89}$. The corresponding residue position in $FimA_{BW25113}$ is shown in parentheses.

[b]*dN*/*dS* ratios of "Infinite" indicate that dS = 0, while ratios of "Undefined" indicate that both dN and dS values = 0, thus preventing analysis. dN/dS >1 indicate adaptive selection and dN/dS <1 indicate purifying selection

[c]p-values <0.1 are considered significant and are indicated in red text.

[d]The indicated mutations were made in the plasmid pTRC99a- $FimA_{BW25113}$ and electroporated into UTI89ΔLIR, *fimA*- for *in vitro* analysis. Vector control (pTRC99a with no FimA) displayed an HA titer of 2 in PBS and 1 in Mannose

[e]"PBS" indicates phosphate buffered saline and "Mannose" indicates PBS with 2% w/v methyl-α-D-mannopyranoside, an inhibitor of type 1 pilus binding. Values represent the last well in a 1:2 serial dilution with a visible hemagglutination (HA) phenotype. Each value shown the median average of 3 biological replicates each with 2 technical replicates. Values with a defect in HA are indicated in red.

[f]The indicated mutations in the mature protein were made in the chromosomal copy of the *fimA* gene in UTI89 for analysis *in vivo*. Not all mutations tested in plasmids were tested in the corresponding *in vivo* analyses.

3.9 Acknowledgements

We would like to thank Wandy Beatty and Bryan Anthony of the WUSM Molecular Microbiology Imaging Facility.

3.9.1 Funding

This work was supported by grants from the NIH [GM122510 (E.H.E), AI048689 and DK064540 (S.J.H), 1F31DK107057 (C.N.S), and DK101171-02 (M.S.C)], the Swedish Research Council 621-2013-5379 (M.A.) and the ANR FiberSpace grant N°ANR-14-CE09-0004 (O.F.). A.L.R. was funded by the Pasteur Paris University PhD program. The cryo-EM work was conducted at the Molecular Electron Microscopy Core facility at the University of Virginia, which is supported by the School of Medicine and built with NIH grant G20-RR31199. The Titan Krios and Falcon II direct electron detector within the Core were purchased with NIH SIG S10-RR025067 and S10-OD018149, respectively.

3.9.2 Competing interests

The authors declare no competing interests, financial or non-financial.

3.8 References

Abraham, S.N., Land, M., Ponniah, S., Endres, R., Hasty, D.L., and Babu, J.P. (1992). Glycerol-induced unraveling of the tight helical conformation of *Escherichia coli* type 1 fimbriae. J Bacteriol *174*, 5145-5148.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological Crystallography *66*, 213-221.

Amann, E., Ochs, B., and Abel, K.J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene *69*, 301-315.

Andersen-Nissen, E., Smith, K.D., Strobe, K.L., Barrett, S.L., Cookson, B.T., Logan, S.M., and Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. Proc Natl Acad Sci USA *102*, 9247-9252.

Anderson, G.G., Palermo, J.J., Schilling, J.D., Roth, R., Heuser, J., and Hultgren, S.J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. Science *301*, 105-107.

Andersson, M., Czerwinski, F., and Oddershede, L.P. (2011). Optimizing active and passive calibration of optical tweezers. J Optics *13*, 044020.

Anisimova, M., Nielsen, R., and Yang, Z. (2003). Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics *164*, 1229-1236.

Aprikian, P., Interlandi, G., Kidd, B.A., Le Trong, I., Tchesnokova, V., Yakovenko, O., Whitfield, M.J., Bullitt, E., Stenkamp, R.E., Thomas, W.E.*, et al.* (2011). The bacterial fimbrial tip acts as a mechanical force sensor. PLoS Biology *9*, e1000617.

Biais, N., Higashi, D.L., Brujic, J., So, M., and Sheetz, M.P. (2010). Force-dependent polymorphism in type IV pili reveals hidden epitopes. Proc Natl Acad Sci USA *107*, 11358-11363.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological Crystallography *66*, 12-21.

Crespo, M.D., Puorger, C., Scharer, M.A., Eidam, O., Grutter, M.G., Capitani, G., and Glockshuber, R. (2012). Quality control of disulfide bond formation in pilus subunits by the chaperone FimC. Nature Chemical Biology *8*, 707-713.

Cusumano, C.K., Pinkner, J.S., Han, Z., Greene, S.A., Ford, B.A., Crowley, J.R., Henderson, J.P., Janetka, J.W., and Hultgren, S.J. (2011). Treatment and prevention of urinary tract infection with orally active FimH inhibitors. Science Translational Medicine *3*, 109-115.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA *97*, 6640-6645.

Delport, W., Poon, A.F., Frost, S.D., and Kosakovsky Pond, S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics *26*, 2455-2457.

DiMaio, F., Song, Y., Li, X., Brunner, M.J., Xu, C., Conticello, V., Egelman, E., Marlovits, T., Cheng, Y., and Baker, D. (2015). Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. Nature Methods *12*, 361-365.

Dong, Y., Liu, Y., Jiang, W., Smith, T.J., Xu, Z., and Rossmann, M.G. (2017). Antibody-induced uncoating of human rhinovirus B14. Proc Natl Acad Sci USA *114*, 8017-8022.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Egelman, E.H. (2000). A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy *85*, 225-234.

Egelman, E.H. (2016). The Current Revolution in Cryo-EM. Biophysical Journal *110*, 1008-1012.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica Section D, Biological Crystallography *66*, 486-501.

Ferry, S., Holm, S., Stenlund, H., Lundholm, R., and Monsen, T. (2004). The natural course of uncomplicated lower urinary tract infection in women illustrated by a randomized placebo controlled study. Scandinavian J Infectious Diseases *36*, 296-301.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. Nucleic Acids Res *43*, W30-38.

Foxman, B. (2014). Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect Dis Clin North Am *28*, 1-13.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Structural Biology *116*, 190-199.

Galkin, V.E., Yu, X., Bielnicki, J., Heuser, J., Ewing, C.P., Guerry, P., and Egelman, E.H. (2008). Divergence of quaternary structures among bacterial flagellar filaments. Science *320*, 382-385.

Greene, S.E., Hibbing, M.E., Janetka, J., Chen, S.L., and Hultgren, S.J. (2015). Human Urine Decreases Function and Expression of Type 1 Pili in Uropathogenic *Escherichia coli*. mBio *6*.

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., and Wang, X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res *27*, 119-129.

Habenstein, B., Loquet, A., Hwang, S., Giller, K., Vasa, S.K., Becker, S., Habeck, M., and Lange, A. (2015). Hybrid Structure of the Type 1 Pilus of Uropathogenic *Escherichia coli*. Angew Chem Int Ed Engl *54*, 11691-11695.

Hannan, T.J., Mysorekar, I.U., Hung, C.S., Isaacson-Schmid, M.L., and Hultgren, S.J. (2010). Early severe inflammatory responses to uropathogenic *E. coli* predispose to chronic and recurrent urinary tract infection. PLoS Pathogens *6*, e1001042.

Hospenthal, M.K., Redzej, A., Dodson, K., Ukleja, M., Frenz, B., Rodrigues, C., Hultgren, S.J., DiMaio, F., Egelman, E.H., and Waksman, G. (2016). Structure of a Chaperone-Usher Pilus Reveals the Molecular Basis of Rod Uncoiling. Cell *164*, 269-278.

Hung, C.S., Dodson, K.W., and Hultgren, S.J. (2009). A murine model of urinary tract infection. Nature Protocols *4*, 1230-1243.

Jones, C.H., Jacob-Dubuisson, F., Dodson, K., Kuehn, M., Slonim, L., Striker, R., and Hultgren, S.J. (1992). Adhesin presentation in bacteria requires molecular chaperones and ushers. Infection and Immunity *60*, 4445-4451.

Jones, C.H., Pinkner, J.S., Roth, R., Heuser, J., Nicholes, A.V., Abraham, S.N., and Hultgren, S.J. (1995). FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. Proc Natl Acad Sci USA *92*, 2081-2085.

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J., and Hultgren, S.J. (2004). Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. Proc Natl Acad Sci USA *101*, 1333-1338.

Justice, S.S., Lauer, S.R., Hultgren, S.J., and Hunstad, D.A. (2006). Maturation of intracellular *Escherichia coli* communities requires SurA. Infection and Immunity *74*, 4793-4800.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol *30*, 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C.*, et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C.*, et al.* (2016). Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res *44*, D574-580.

Khetrapal, V., Mehershahi, K.S., Chen, S., and Chen, S.L. (2016). Application and Optimization of relE as a Negative Selection Marker for Making Definitive Genetic Constructs in Uropathogenic *Escherichia coli*. Pathogens *5*.

Kosakovsky Pond, S.L., and Frost, S.D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol *22*, 1208-1222.

Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D., Delport, W., and Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol *28*, 3033-3043.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol *23*, 1891-1901.

Kostakioti, M., Hadjifrangiskou, M., Cusumano, C.K., Hannan, T.J., Janetka, J.W., and Hultgren, S.J. (2012). Distinguishing the contribution of type 1 pili from that of other QseB-misregulated factors when QseC is absent during urinary tract infection. Infection and Immunity *80*, 2826-2834.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci USA *109*, E2774-2783.

Li, J., Nelson, K., McWhorter, A.C., Whittam, T.S., and Selander, R.K. (1994). Recombinational basis of serovar diversity in Salmonella enterica. Proc Natl Acad Sci USA *91*, 2552-2556.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods *10*, 584-590.

Mabeck, C.E. (1972). Treatment of uncomplicated urinary tract infection in non-pregnant women. Postgraduate Medical Journal *48*, 69-75.

Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S., and Hultgren, S.J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO journal *19*, 2803-2812.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Structural Biology *142*, 334-347.

Mortezaei, N., Singh, B., Bullitt, E., Uhlin, B.E., and Andersson, M. (2013). P-fimbriae in the presence of anti-PapA antibodies: new insight of antibodies action against pathogens. Scientific Reports *3*, 3393.

Mulvey, M.A., Lopez-Boado, Y.S., Wilson, C.L., Roth, R., Parks, W.C., Heuser, J., and Hultgren, S.J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. Science *282*, 1494-1497.

Nuccio, S.P., and Baumler, A.J. (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol Mol Biol Rev *71*, 551-575.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pond, S.L., and Frost, S.D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics *21*, 2531-2533.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Roh, S.H., Hryc, C.F., Jeong, H.H., Fei, X., Jakana, J., Lorimer, G.H., and Chiu, W. (2017). Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. Proc Natl Acad Sci USA *114*, 8259-8264.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/

Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J., and Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. Science *285*, 1058-1061.

Sauer, F.G., Pinkner, J.S., Waksman, G., and Hultgren, S.J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. Cell *111*, 543-551.

Saulino, E.T., Bullitt, E., and Hultgren, S.J. (2000). Snapshots of usher-mediated protein secretion and ordered pilus assembly. Proc Natl Acad Sci USA *97*, 9240-9245.

Scholes, D., Hooton, T.M., Roberts, P.L., Stapleton, A.E., Gupta, K., and Stamm, W.E. (2000). Risk factors for recurrent urinary tract infection in young women. Journal Infectious Diseases *182*, 1177-1182.

Schreiber, H.L.t., Conover, M.S., Chou, W.C., Hibbing, M.E., Manson, A.L., Dodson, K.W., Hannan, T.J., Roberts, P.L., Stapleton, A.E., Hooton, T.M*., et al.* (2017). Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. Science Translational Medicine *9*.

Smith, K.D., and Ozinsky, A. (2002). Toll-like receptor-5 and the innate immune response to bacterial flagellin. Current Topics Microbiology and Immunology *270*, 93-108.

Spaulding, C.N., and Hultgren, S.J. (2016). Adhesive Pili in UTI Pathogenesis and Drug Development. Pathogens *5*.

Spaulding, C.N., Klein, R.D., Ruer, S., Kau, A.L., Schreiber, H.L., Cusumano, Z.T., Dodson, K.W., Pinkner, J.S., Fremont, D.H., Janetka, J.W*., et al.* (2017). Selective depletion of uropathogenic *E. coli* from the gut by a FimH antagonist. Nature *546*, 528-532.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688-2690.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol *57*, 758-771.

Subramaniam, S., Earl, L.A., Falconieri, V., Milne, J.L., and Egelman, E.H. (2016). Resolution advances in cryo-EM enable application to drug discovery. Current Opinion Structural Biology *41*, 194-202.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J Structural Biology *157*, 38-46.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. Nature Reviews Microbiology *8*, 207-217.

Tolic-Norrelykke, S.F., Schaffer, E., Howard, J., Pavone, F.S., Julicher, F., and Flyvberg, H. (2006). Calibration of optical tweezers with positional detection in the back focal plane. Review of Scientific Intruments *77*, 103101.103101-103101.103111.

Vonck, J., and Mills, D.J. (2017). Advances in high-resolution cryo-EM of oligomeric enzymes. Current Opinion Structural Biology *46*, 48-54.

Walczak, M.J., Puorger, C., Glockshuber, R., and Wider, G. (2014). Intramolecular donor strand complementation in the *E. coli* type 1 pilus subunit FimA explains the existence of FimA monomers as off-pathway products of pilus assembly that inhibit host cell apoptosis. J Molecular Biology *426*, 542-549.

Wang, R.Y., Kudryashev, M., Li, X., Egelman, E.H., Basler, M., Cheng, Y., Baker, D., and DiMaio, F. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. Nature Methods *12*, 335-338.

Wildschutte, H., Wolfe, D.M., Tamewitz, A., and Lawrence, J.G. (2004). Protozoan predation, diversifying selection, and the evolution of antigenic diversity in Salmonella. Proc Natl Acad Sci USA *101*, 10644-10649.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2005). Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization. Infection and Immunity *73*, 7657-7668.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. Cellular Microbiology *9*, 2230-2241.

Wurpel, D.J., Beatson, S.A., Totsika, M., Petty, N.K., and Schembri, M.A. (2013). Chaperone-usher fimbriae of *Escherichia coli*. PloS One *8*, e52835.

Zakrisson, J., Wiklund, K., Axner, O., and Andersson, M. (2012). Helix-like biopolymers can act as dampers of force for bacteria in flows. European biophysics journal : EBJ *41*, 551-560.

Zowawi, H.M., Harris, P.N., Roberts, M.J., Tambyah, P.A., Schembri, M.A., Pezzani, M.D., Williamson, D.A., and Paterson, D.L. (2015). The emerging threat of multidrug-resistant Gram-negative bacteria in urology. Nature Reviews Urology *12*, 570-584.

Chapter 4:

Selective depletion of uropathogenic *E. coli* from the gut by a FimH antagonist.

By

Caitlin N. Spaulding, Roger D. Klein, Ségolène Ruer, Andrew L. Kau, Henry L. Schreiber, Zachary T. Cusumano, Karen W. Dodson, Jerome S. Pinkner, Daved H. Fremont, James W. Janetka, Han Remaut, Jeffrey I. Gordon and Scott J. Hultgren

Nature

2017 Jun 22. 546(7659): 528-532.

doi: 10.1038/nature22972. PMID 28614296

Format Adapted for Dissertation

4.1 Background

Urinary tract infections (UTIs) caused by uropathogenic *Escherichia coli* (UPEC) affect 150 million people annually (Flores-Mireles et al., 2015; Foxman, 2014). Despite effective antibiotic therapy, 30–50% of patients experience recurrent UTIs (Foxman, 2014). In addition, the growing prevalence of UPEC that are resistant to last-line antibiotic treatments, and more recently to carbapenems and colistin, make UTI a prime example of the antibiotic-resistance crisis and emphasize the need for new approaches to treat and prevent bacterial infections (Mediavilla et al., 2016; Spaulding and Hultgren, 2016; Zowawi et al., 2015). UPEC strains establish reservoirs in the gut from which they are shed in the faeces, and can colonize the periurethral area or vagina and subsequently ascend through the urethra to the urinary tract, where they cause UTIs(Yamamoto et al., 1997). UPEC isolates encode up to 16 distinct chaperone-usher pathway pili, and each pilus type may enable colonization of a habitat in the host or environment (Wurpel et al., 2013). For example, the type 1 pilus adhesin FimH binds mannose on the bladder surface, and mediates colonization of the bladder. However, little is known about the mechanisms underlying UPEC persistence in the gut(Spaulding and Hultgren, 2016). Here, using a mouse model, we show that F17-like and type 1 pili promote intestinal colonization and show distinct binding to epithelial cells distributed along colonic crypts. Phylogenomic and structural analyses reveal that F17-like pili are closely related to pilus types carried by intestinal pathogens, but are restricted to extra-intestinal pathogenic *E. coli*. Moreover, we show that targeting FimH with M4284, a high-affinity inhibitory mannoside, reduces intestinal colonization of genetically diverse UPEC isolates, while simultaneously treating UTI, without notably disrupting the structural configuration of the gut microbiota. By selectively depleting intestinal UPEC reservoirs, mannosides could markedly reduce the rate of UTIs and recurrent UTIs.

4.2 Contribution to findings

In this Chapter, I present a large, collaborative project describing the functional role for chapterone-usher pathway (CUP) pili in mediating gut colonization by uropathogenic *E. coli* (UPEC). Specifically, we showed that both type 1 and F17-like pili are critical factors in mediating UPEC colonization of the mouse gut. Further, we show the ability of a small molecule inhibitor of type 1 pili, so-called mannosides, are capable of disrupting this colonization without disrupting the rest of the gut microbial community. Within this project, my contribution was an investigation of the carriage and evolution of F17-like pili in clinical UPEC and the association of this pilus system with recurrent urinary tract infections (rUTI). Briefly, I found that the *ucl* operon, which encodes F17-like pili, is found almost exclusively within clade B2 strains of *E. coli*, which cause the majority of UTIs in the U.S.A (Schreiber, 2017). Importantly, I found that F17-like pili are further enriched in UPEC strains causing recurrent UTI. In addition, I found that the *ucl* operon is most closely related to CUP pili operons found in other bacteria that colonize the gut, specifically *Proteus mirabilis*, and that *E. coli* most likely obtained the *ucl* operon through horizontal gene transfer from these gut colonizers. Taken together, these data suggest that UPEC have coopted F17-like pili to enable their persistence within the gut reservoir of their hosts, thus enhancing their ability to cause recurrent UTI.

4.3 Results

4.3.1 Influence of chaperone-usher pathway (CUP) pili on gut colonization

The genome of UTI89, a human cystitis isolate, contains nine distinct functional chaperone-usher pathway (CUP) pili. To determine whether any of these CUP pili promote intestinal colonization, we used a strep- tomycin mouse model of UPEC intestinal colonization(Kaiser et al., 2012) to co-colonize C3H/HeN mice with wild-type UTI89 and one of

nine mutant strains, each lacking a single CUP operon (**Figure S1**). Deletion of each of the seven operons *yfc*, *yeh*, *yad*, *pap*, *sfa*, *yqi* and *mat* had no effect on UTI89 intestinal fitness compared to the isogenic wild-type strain (**Figure 1a–g**). However, deletion of the *fim* or *ucl* pilus operons, which encode type 1 or F17-like pili, respectively, produced significant defects in colonization (up to 100- and 1,000-fold, respectively; Figure 1h, i). Loss of FimH, the type 1 pilus adhesin, mirrored the defect caused by deletion of the full type 1 pilus operon (**Figure S2a**). Deletion of both pilus types in a single strain produced a fitness defect greater than either individual deletion alone, suggesting that these two pilus types do not have redundant roles (**Figure 1j, k**).

In a mouse model, type 1 pilus-mediated binding to mannosylated receptors is indispensable for bladder colonization and invasion of urothelial cells lining the bladder lumen2,5. Once inside urothelial cells, a single bacterium rapidly divides, forming an intracellular bacterial community (IBC) (Flores-Mireles et al., 2015; Spaulding and Hultgren, 2016). Furthermore, UPEC can access under- lying transitional cells, forming quiescent intracellular reservoirs (Flores-Mireles et al., 2015; Spaulding and Hultgren, 2016). Mutations in *fimH* abolish the ability of UPEC to colonize the bladder, form IBCs and quiescent intracellular reservoirs (Flores-Mireles et al., 2015; Spaulding and Hultgren, 2016; Wright et al., 2007). By contrast, no role was observed for F17-like pili in the rate or severity of bladder infection after individual or concurrent transurethral inoculations of UTI89 and UTI89Δ*ucl* strains into the bladders of female C3H/HeN mice (**Figure S3**). Differences between mouse and human bladders or the overexpression of F17-like pili *in vitro* may account for the inconsistency with another study that showed a role for F17-like pili in binding to desquamated epithelial cells obtained from human urine (Wurpel et al., 2016).

4.3.2 Gut binding by FimH and UclD

The *fim* and *ucl* operons encode two-domain tip adhesin proteins, FimH and UclD, respectively. The adhesin lectin domain contains the ligand-binding site, and the pilin domain joins the adhesin to the pilus rod(Spaulding and Hultgren, 2016). Purified FimH lectin domain ($FimH^{LD}$) bound to more differentiated epithelial cells located in the upper portion of crypts and in 'surface epithelial cuffs' (the colonic homologues of small intestinal villi) (**Figure 1l**). FimH binding was prevented by pretreating tissue sections with peptide-*N*-glycosidase F (PNGase F), which cleaves *N*-linked oligosaccharides. $FimH^{LD}$ also bound to Caco-2 cells (an immortalized human enterocyte-like cell line derived from colorectal carcinoma); binding was inhibited by D-mannose and a high-affinity mannose analogue (mannoside), M4284 (**Figure S2b**)(Jarvis et al., 2016). The UclD lectin domain ($UclD^{LD}$) also bound colonic epithelial cells in tissue sections; binding was inhibited by pretreating tissue sections with *O*-glycosidase, an enzyme that cleaves *O*-linked oligosaccharides, suggesting that the UclD ligand is contained within an *O*-glycan (**Figure 1m**).

4.3.3 Evolution of the *ucl* operon

CUP pili are highly conserved throughout Proteobacteria and are assembled by dedicated chaperone-usher assembly machines encoded by each respective CUP operon along with the various subunit types comprising the pilus fibre (Spaulding and Hultgren, 2016; Wurpel et al., 2013). The sequence identity between usher genes of distinct CUP pilus types is greater than the identity of genes that encode other CUP pilus proteins and thus can be compared to deter- mine evolutionary relationships of CUP pili among Proteobacteria (Nuccio and Bäumler, 2007; Wurpel et al., 2013). A homology search of a database of γ-Proteobacteria genomes revealed that the UTI89 F17-like usher gene sequence (*uclC*) shared highest identity with other *E. coli uclC* sequences and with orthologous usher sequences of *Proteus mirabilis*, a bacterium

that can colonize the gut, and *Salmonella enterica*, an intestinal pathogen (**Table S1**). The *uclC* usher gene was also closely related to usher genes in F17 (thus the derivation of the name, F17-like), *pVir99* and *ECs1278* pili. A phylogenetic analysis showed clustering of these *E. coli* and *Proteus* species ushers into a distinct sub-branch within the broader F17 group usher phylogeny, suggesting that they share a common ancestor (**Figure S4**). F17-like pili are present in only 10% of *E. coli* strains; these strains are almost exclusively in the B2 clade, which contains most extraintestinal pathogenic *E. coli* (ExPEC) and UPEC strains (Wurpel et al., 2016; Zhang et al., 2002). By contrast, F17 and ECs1278 pili are found in the intestinal pathogens, enterotoxigenic *E. coli* (ETEC) and enterohemorrhagic *E. coli* (EHEC), respectively, which are specific to clades A, E and B1 (Low et al., 2006; Richards et al., 2014). These findings suggest that UPEC strains in clade B2 may have acquired the *ucl* operon from a different species and retained this factor to facilitate its residency in the gut. No binding of $UclD^{LD}$ was observed to *N*-acetylglucosamine (GlcNAc), the ligand bound by F17 pili, indicating that F17-like pili bind a distinct ligand from that of F17 pili (Merckel et al., 2003)(**Figure 1n**). Notably, although the amino acid sequence of the full-length UclD adhesin has diverged from that of the F17 adhesin F17G, it is almost invariant across all strains encoding it (more than 99%), suggesting that there is a single, distinct ligand for UclD adhesins (Wurpel et al., 2016).

4.3.4 Carriage of F17-like pili in uropathogenic *E. coli*

A comparative genomic analysis of 43 strains isolated from a cohort of 14 women at the time of initial presentation with acute UTI, or during subsequent recurrent UTIs, revealed that 14 of the recurrent UTI events were caused by B2 strains (Czaja et al., 2009; Schreiber et al., 2017b)(**Table S2**). Of these 14 strains, 13 encoded F17-like pili (approximately 93%) (**Figure S5**). By contrast, F17-like pili have been found in less than 50% of all B2 strains (Wurpel et al.,

2016)(*E. coli* reference collection; ECOR), suggesting that F17-like pili might be associated with UPEC persistence in women with recurrent UTI owing to their ability to promote maintenance of a UPEC intestinal reservoir.

4.3.5 Crystal structure of UclD^{LD}

To further characterize F17-like pili, we solved two X-ray crystal structures of UclD^{LD}. The structures in the $P2_1$ (green) and $P2_12_12_1$ (grey) space groups were resolved to 1.05 Å and 1.60 Å resolution, respectively, and are nearly identical (**Figure 2a** (left) and **Table S3**). Despite low primary sequence identity (~25%), the structural characteristics of UclD^{LD} and the F17 adhesin F17G^{LD} are conserved (Merckel et al., 2003)(**Figure 2a–c**). This includes the presence of a transverse putative binding site in UclD^{LD} located at a similar position to the GlcNAc-binding site on F17G (Merckel et al., 2003)(**Figure 2a** (right)). Structural and sequence alignments reveal two large conserved insertions in UclD relative to F17G (**Figure 2c**), whose direct proximity to the putative binding pocket suggests that they are involved in UclD receptor binding. The six conserved residues that make up the candidate binding pocket in UclD are chemically distinct from their F17G counterparts (**Figure 2b, c**), providing further evidence that UclD binds a distinct ligand.

4.3.6 FimH antagonists block UPEC colonization of the gut

In light of the role of FimH and UclD in gut colonization, we conducted a study designed to reduce the UPEC intestinal reservoir with an adhesin-directed therapeutic. M4284 is a high-affinity biphenyl mannoside whose binding affinity for FimH is approximately 100,000-fold higher than the natural sugar D-mannose (Han et al., 2012; Jarvis et al., 2016). Pharmacokinetic analysis revealed that M4284 concentrations remain high in the faeces of mice for up to 8 h after an oral dose (**Figure 3a**). Treating mice that were colonized by UTI89 with three doses

(administered by oral gavage) of M4284, a regimen that successfully treats UPEC UTI in mice (Cusumano et al., 2011), significantly reduced levels of UTI89 in the faeces, caecum and colon compared to those treated with a vehicle control (**Figure 3b, c**). Treating mice with additional M4284 doses further reduced the UTI89 population, and the overall number of UPEC continued to be lower in M4284-treated mice after termination of treatment (**Figure S6**). While D-mannose blocks FimH binding *in vitro*, treating mice with D-mannose did not alter UTI89 levels *in vivo* (**Figure S2b** and **Figure 3d**).

The infectious dose required to cause cystitis in 50% of mice (ID_{50}) in the UTI mouse model is 10^5 colony-forming units (CFU) (Rosen et al., 2008). Furthermore, decreasing the dose of UPEC introduced into the bladder from 10^8 to 10^6 CFU significantly reduced the rate of UTI, suggesting that the 1–1.5 log (or 90 – 95%) mannoside-driven reduction in faecal UPEC levels would reduce the numbers of bacteria available to access the urinary tract and probably reduce the rate of UTI and/or recurrent UTI (**Figure S7**). Indeed, we found that M4284 simultaneously reduces UTI89 levels in the gut and urinary tracts of mice that were concurrently colonized with UTI89 in the gut and bladder (**Figure 3e–g**).

4.3.7 FimH antagonists do not disrupt the gut microbiota and are broadly effective

We sequenced bacterial 16S rRNA gene amplicons generated from faecal samples of C3H/HeN mice that had not been given streptomycin or infected with UPEC but were treated with three doses of M4284 or vehicle alone. We found that M4284 produced no significant changes in the overall phylogenetic configuration of the microbiota as judged by the unweighted UniFrac dissimilarity metric, in contrast to the significant perturbations produced by treatment with ciprofloxacin, a fluoroquinolone antibiotic (**Figure 4a** and **Figure S8a**). Using this UniFrac metric, we found that M4284 treatment did not produce significant perturbations in bacterial

community structure in mice pretreated with streptomycin and then colonized with UTI89 (**Figure S9**). We concluded that M4284 can function to selectively extirpate UPEC from the gut in our preclinical model. Notably, although most Enterobacteriaceae carry the *fim* operon (Jones et al., 1995; Schreiber et al., 2017a; Wurpel et al., 2013), M4284 treatment does not significantly affect the abundance of intestinal Enterobacteriaceae (**Figure S8a, b**), suggesting that these bacteria may not be expressing type 1 pili during M4284 exposure or that they reside within inaccessible intestinal habitats.

M4284 treatment of mice colonized with three additional genetically diverse UPEC clinical isolates (EC958 (Totsika et al., 2011), 41.4p (Schreiber et al., 2017b) and CFT073(Welch et al., 2002)), reduced the levels of each UPEC strain by a similar percentage in the faeces, caecum and colon (**Figure 4b–f** and **Figure S10a**). Furthermore, we found that M4284 treatment reduced UTI89 levels in C3H/HeN and C57BL/6 mice from different vendors, containing distinct gut microbial communities. In each case tested, the percentage reduction in UPEC levels in caecum, colon and faeces did not vary significantly between the different treatment groups (**Figure 4f–h** and **Supplementary Figs 8c, 10b**). We concluded that M4284 treatment has activity against different UPEC strains in different host genetic backgrounds and gut microbial community contexts.

4.4 Discussion

As the prevalence of antibiotic-resistant pathogens continues to rise, the need to develop highly targeted/specific therapeutic approaches has gained increased urgency (Galtier et al., 2016; Yao et al., 2016). Furthermore, an increasing number of studies are finding that disruption of the gut microbiota by orally administered antibiotics, especially during childhood, may affect its functional properties in ways that are deleterious to the host, not only in the short term but

also for more protracted periods of time(Cox et al., 2014; Dethlefsen and Relman, 2011). Therefore, developing therapeutic agents, such as mannosides, that specifically target a pathogen without disrupting the remainder of a microbial community has important ramifications not only for UPEC but also potentially for other infections, including those caused by enteropathogens. In addition, the identification of genes involved in UPEC intestinal colonization may provide a method by which patients with UTIs could be stratified for epidemiological studies of risk for recurrent disease as well as for proof-of-concept clinical studies of the efficacy of CUP-directed treatment regimens.

4.5 Materials and Methods

4.5.1 Ethics statement

The Washington University Animal Studies Committee approved all procedures used for the mouse experiments described in the present study. Overall care of the animals was consistent with The Guide for the Care and Use of Laboratory Animals from the National Research Council and the USDA Animal Care Resource Guide. For collection of colonic tissues for adhesion binding studies, mice were euthanized according to institutional, national and European animal regulations, using protocols that were also approved by the animal ethics committee of Ghent University.

4.5.2 Bacterial strains

CUP operon and adhesin deletions in UTI89 were engineered by replacing the gene(s) of interest with antibiotic-resistance markers using the λ Red Recombinase system (Datsenko and Wanner, 2000). Earlier reports described wild-type UTI89 and its isogenic *fim* and *fimH* mutants (Rosen et al., 2008; Wright et al., 2007) as well as EC958 (Totsika et al., 2011), 41.4p (Schreiber et al., 2017b) and CFT073 (Welch et al., 2002).

4.5.3 Colonization of mice with UPEC strains

Six-week-old female C3H/HeN mice were obtained from Envigo or Charles River Laboratories (CRL). Six-week-old female C57BL/6 mice were also obtained from Envigo. Animals were maintained in a single room in our vivarium for no more than 2 days before treatment. Before and after treatment all animals received PicoLab Rodent Diet 20 (Purina) *ad libitum*. All animals were maintained under a strict light cycle (lights on at 06:00, off at 18:00). For competitive infections, if a phenotype was observed after testing five mice (1 biological replicate), the experiment was repeated 1–2 times (total of $n = 10–16$ mice, 2–3 biological replicates). For 16S rRNA analyses, $4 – 5$ mice were examined (1 biological replicate). For all other experiments, $9 – 16$ mice were tested and the experiment was repeated 2–3 times (2–3 biological replicates). Exclusion criteria for mice were pre-established; (i) both introduced strains in competitive infections became undetectable during the course of a 14-day experiment, and (ii) mice died or lost more than 20% of their body weight. No mice in this study met these criteria. Mice were acquired from indicated vendors and randomly placed into cages ($n = 5$ mice per cage) by employees of Washington University's Division of Comparative Medicine; no additional methods for randomization were used to determine how animals were allocated to experimental groups. Investigators were not blinded to group allocation during experiments.

Animals received a single dose of streptomycin (1,000 mg kg^{-1} in 100 µl water by oral gavage) followed 24 h later by an oral gavage of approximately 10^8 CFU UPEC in 100 µl PBS. Bladder infections were performed via transurethral inoculation[30]. UPEC strains were prepared for inoculation as described previously (Hung et al., 2009). In brief, a single UTI89 colony was inoculated in 20 ml of Luria Broth (LB) and incubated at 37 °C under static conditions for 24 h. Bacteria were then diluted (1:1,000) into fresh LB and incubated at 37 °C under static conditions for 18–24 h. Bacteria were subsequently washed three times with PBS and then concentrated to

approximately 1×10^8CFU per 100 µl for intestinal infections and 1×10^8 CFU per 50 µl for bladder infections.

In all cases, faecal and urine samples were collected directly from each animal at the indicated time points. Faecal samples were immediately weighed and homogenized in 1 ml PBS. Urine samples were immediately diluted 1:10 before plating. Mice were euthanized via cervical dislocation under isofluorane anaesthesia and their organs were removed and processed under aseptic conditions. Intestinal segments (caecum and colon) were weighed before homogenization and plating on LB supplemented with the appropriate antibiotic.

4.5.4 Enumeration of bladder intracellular bacterial communities

Six-week-old C3H/HeN mice were given a single oral dose of either M4284 (100 mg kg^{-1}) or vehicle control (10% cyclodextrin) 30 min before transurethral inoculation with UTI89. To count accurately the number of IBCs, mice were euthanized 6 h after infection. Bladders were removed aseptically, bisected, splayed on silicone plates and fixed in 4% (v/v) paraformaldehyde. IBCs, readily discernable as punctate violet spots, were quantified by LacZ staining of bladder wholemounts (Cusumano et al., 2011; Justice et al., 2006).

4.5.5 Immunofluorescence studies

The protocols used for immunohistochemical analysis are based on a previous study (Johansson and Hansson, 2012). After euthanization of 6-week-old, female C57BL/6 mice (supplied by VIB-Ghent University breeding program, Belgium), segments of colon were fixed in methanol-Carnoy for a minimum of 3 h at room temperature. The fixed tissues were then embedded in paraffin and 4-µm-thick sections were cut and placed on glass slides. Slides were de-paraffinized and re-hydrated by incubating them in xylene, isopropanol, 100% ethanol and finally 70% ethanol (each step involving a 3 min incubation in the reagent followed by another

3 min incubation in fresh reagent). Slides were subsequently rinsed in tap water and PBS, placed in blocking buffer (5% fetal calf serum prepared in PBS) at room temperature for 30 min, and then incubated with rabbit polyclonal antibodies to Muc2 (1:2,000; mucin 2 (H-300), Santa Cruz Biotechnology) for 2 h. After three washes with PBS, slides were incubated with a goat anti-rabbit Dylight-488 labelled secondary antibody (1:1,000 dilution, ThermoFisher 35553) in blocking buffer for 1 h at room temperature. Slides were washed three times with PBS before counterstaining with bis-benzimide (Hoechst dye) (1:1,000 in PBS) for 10 min at room temperature. Finally, slides were incubated with $FimH^{LD}$ or $UclD^{LD}$ ($P2_1$) protein, labelled with NHS 650 nm Dylight, in blocking buffer at 4 °C overnight. Before staining, sections were treated with *O*-glycosidase (NEB) or PNGase F (Sigma) at 37 °C using buffers and protocols supplied by the manufacturer. Slides were washed subsequently with PBS before treatment with fluoro-mounting medium (*n*-propyl gallate in glycerol) and viewing under a confocal microscope (Leica Microsystems LAS-AF-TCS SP5) using a 20 × 125 objective.

4.5.6 Mannoside treatment

D-mannose or the mannoside M4284 (which has been characterized in a previous study (Jarvis et al., 2016)), were diluted in vehicle (water and 10% cyclodextrin, respectively) and administered to 6-week-old C3H/HeN mice at a dose of 100 mg kg^{-1}. Control animals were treated with water or 10% cyclodextrin alone. Unless stated otherwise, three doses of M4284, cyclodextrin, or D-mannose were given via oral gavage over 24 h, with doses administered 8 h apart. Mice were euthanized and intestinal tissues were processed for analysis of viable bacteria (CFU) 8 h after the last dose, unless otherwise noted. To test the effect of M4284 on intestinal UPEC titres after treatment was terminated, mice were euthanized 5 days after the last dose of mannoside. To test the effect of additional doses on M4284 treatment on UPEC titres, mice were

given five doses of mannoside; the first three doses were administered 8 h apart, followed 12 h later by the fourth dose, and 24 h later by the fifth dose. Mice were euthanized 24 h after the fifth dose.

4.5.7 Carriage of F17-like pili

We examined 43 available UPEC isolates (Table S2). These isolates originated from a clinical study of 14 women who experienced at least two episodes of a UTI (an initial UTI and one or more recurrent UTIs) during the 90-day study window(Czaja et al., 2009). The isolates used in this work were sequenced in a previous study (Schreiber et al., 2017b)(Bioproject ID PRJNA269984) and include (i) 14 isolates collected at enrollment, (ii) 18 isolates collected during recurrent UTI (10 women experienced a single recurrent UTI while four women experienced two recurrent UTI events), and (iii) 11 isolates collected in the days leading up to a recurrent UTI.

The distribution of the F17-like operon in these clinical *E. coli* isolates was determined using BLAST and the F17-like operon from UTI89 as the query sequence. A 'hit' was considered as any genome sequence that matched the entire length of the query sequence with more than 75% identity. As a control to prevent false negatives in the BLAST search of draft genomes, DNA sequencing reads from each clinical UPEC isolate were mapped against a reference sequence constructed by concatenating all the *ucl* genes with 100 N-separators using Geneious v6.1.7 (Kearse et al., 2012).

4.5.8 Phylogenetic analyses and sequence alignments

Amino acid alignments of full-length UTI89 UclD, *P. mirabilis* UcaD, *S. enterica* UclH, and ETEC F17G were conducted using the MAFFT L-INS-i iterative refinement method and the default BLOSUM62 scoring matrix (Katoh et al., 2005)(Table S1). MAFFT collected up to 100

homologues with E values of less than 1×10^{-10} to each sequence from the SwissProt database to improve alignment accuracy. Homologues are automatically removed from the final alignment. The alignment was visualized using Geneious[33]. A homology search of the coding sequence database of the European Nucleotide Archive (ENA) was conducted using the Basic Local Alignment Search Tool (BLAST)(Altschul et al., 1990) using the UTI89 *uclC* (ENA accession ABE10308) and EDL933 *ECs1278* (ENA accession AIG67653) as queries. Sequences that matched either gene sequence with more than 50% identity were downloaded and then filtered to remove partial hits (less than 80% length of query sequence) and sequences with nonsense mutations, which resulted in a total of 659 sequences (Table S1). Duplicate sequences were then removed, resulting in a list of 122 unique, representative sequences. These sequences were then aligned with the UTI89 *fimD* usher sequence (ENA accession ABE10417) as an outgroup using the MAFFT MAFFT L-INS-i alignment method and the 200PAM scoring matrix(Katoh et al., 2005): The phylogenetic relationship between gene sequences was then estimated using RAxML v8.1.3 with the GTRCAT model (Stamatakis et al., 2005; Stamatakis, 2006)and supported with 1000 bootstrap replicates; the tree was visualized using the tool interactive Tree of Life (iTOL) v3 (Letunic and Bork, 2016).

4.5.9 ELISA targeting FimH

Caco-2 cells (ATCC number HTB-37) were cultured in minimum essential medium (MEM) supplemented with 20% FBS. Cell cultures tested negative for mycoplasma. Cells were split into 48-well plates, grown to 100% confluence and then fixed with paraformeldahyde for 15 min followed by treatment with blocking buffer (PBS containing 2% BSA) for 2 h. A truncated FimH, corresponding to residues 1–178 of the mature FimH adhesin ($FimH^{LD}$), expressed in *E. coli* and purified as described previously (Kalas et al., 2017), was serially diluted

in blocking buffer and incubated with the fixed Caco-2 cells for 1h at room temperature. To test the effect of D-mannose or M4284 on FimH binding, 0.2 mg ml^{-1} $FimH^{LD}$ was pre-incubated for 5 min in the presence or absence of 1 mM D-mannose (Sigma-Aldrich) or 1 mM M4828 (in 20 mM Tris, pH 8.0, or 20 mM Tris plus water or 10% cyclodextrin, respectively) before serial dilution and incubation. Wells were washed four times with PBS and 0.05% Tween 20 (PBST) before incubation with a polyclonal rabbit anti-T3 antibody against $FimH^{LD}$ (generated against FimH residues 1–165; ref. 37) for 1 h at room temperature. After another series of four washes, secondary antibody (goat anti-rabbit Ig conjugated to horseradish peroxidase; ThermoFisher, 32460) was incubated with the cells for 1 h at room temperature (24 °C) before washing in PBST. Plates were developed with the BD OptEIA TMB substrate reagent kit for 5 min at room temperature (24 °C) before quenching with 1 M H_2SO_4. Binding was assessed by measuring the absorbance at 450 nM on a TECAN infinite 2 PRO plate reader. Wells lacking protein were used as control. All conditions were examined in quadruplicate.

4.5.10 Effect of antibiotic exposure on the microbiota

Six-week-old female C3H/HeN mice from Envigo and CRL were subjected to the following treatments: (i) none (naive control mice, untreated), (ii) three doses of M4284 (100 mg kg^{-1}; in 10% cyclodextrin) or 10% cyclodextrin given 8 h apart, or (iii) ciprofloxacin (two doses of 15 mg kg^{-1} given 12 h apart). All doses were given via oral gavage. Five mice were included for each treatment type (1 biological replicate) and faecal samples were collected before treatment and 24 h after the last dose of each treatment. Another group of four C3H/HeN mice from Envigo (1 biological replicate) were pretreated with streptomycin and colonized with UTI89 before receiving treatment with either three doses of M4284 (100 mg kg^{-1}; in 10%

cyclodextrin) or with 10% cyclodextrin alone; faecal samples were collected before treatment but after exposure to streptomycin and UTI89 and 24 h after the last dose of each treatment.

4.5.11 Bacterial 16S rRNA sequencing

DNA was extracted by bead beating in extraction buffer (200 mM Tris, pH 8.0, 200 mM NaCl, 20 mM EDTA), 210 µl of 20% SDS and 500 µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1). This crude DNA extract was purified (Qiaquick PCR purification kit) and PCR used to generate amplicons from the V4 region of bacterial 16S rRNA genes using primers and cycling conditions described previously (Caporaso et al., 2011). Amplicons were pooled in equimolar ratios and sequenced on an Illumina MiSeq instrument (paired-end 250-nucleotide reads). Paired V4-16S rRNA sequences were merged using FLASH software (Magoč and Salzberg, 2011), demultiplexed, and reads clustered into 97%ID OTUs (2013 Greengenes OTU reference database; QIIME version 1.9.0 (Rideout et al., 2014)). A custom database using modified NCBI bacterial taxonomy was used to train the Ribosomal Database Project (RDP) version 2.4 classifier and assign taxonomy to picked OTUs (Kau et al., 2015). The resulting OTU table was filtered to include only OTUs found in at least two samples at greater than or equal to 0.1% relative abundance.

4.5.12 F17-like constructs and purification

For the $P2_1$ UclD^{LD} construct, the first 197 amino acids of the mature UclD adhesin protein were cloned into pDEST14 using Gateway technology (Invitrogen), resulting in plasmid pUclD^{AD}. Expression was induced with 1 mM IPTG. Periplasmic extracts were prepared by resuspending bacterial pellets in 20 mM Tris, 20% sucrose, pH 8 (4 ml per gram of pellet). Subsequently, 40 µl of 0.5M EDTA and 10 mg ml^{-1} lysozyme were added per gram of pellet and the suspension was incubated on ice for 30 min. This step was followed by addition of 40 µl of

2.5 M $MgCl_2$ per gram of cell pellet and incubation on ice for 5 min. Cells were spun at 15,000*g* and the supernatant was saved as the periplasmic extract. The extract containing the UclD lectin domain was dialysed against 20 mM HEPES, pH 7, passed over a SP FF cation exchange column (GE Healthcare) and bound material eluted with 20 mM HEPES, pH 7, 1 M NaCl. Pooled fractions containing UclD lectin domain were then applied to a Phenyl Hi Trap column (GE Healthcare) after addition of 1 M ammonium sulfate. Elution was performed using 20 mM HEPES, pH 7.

To generate purified $UclD^{LD}$ for the $P2_12_12_1$ space group, DNA from the UTI89 *uclD* gene encoding the N-terminal 217 amino acids of the protein were cloned into pTRC99a with a C-terminal 6-His tag. This construct was expressed in the periplasm of *E. coli* DL41(DE3), a methionine auxotroph strain suitable for expression of native or selenomethionine-labelled protein. Periplasmic extracts were first dialysed against PBS supplemented with 250 mM NaCl, then bound to a cobalt (Goldbio) column; bound proteins were eluted with PBS containing 250 mM NaCl, and 250 mM imidazole. Pooled fractions were dialysed into 20 mM MES, pH 5.8, bound to an HR16/10 Mono S cation exchange column (GE Healthcare), and eluted with 300 mM NaCl. After cleavage of the periplasmic localization sequence, the mature form of $UclD^{LD}$–6×His contained 203 amino acids.

Selenomethionine-labelled protein was purified using the same protocol, but all buffers were supplemented with 2 mM β-mercaptoethanol and 1 mM EDTA to prevent oxidation. EDTA was omitted from the periplasmic dialysis buffer to prevent chelation of immobilized cobalt.

4.5.13 Crystallization and structure determination

For the $P2_1$ $UclD^{LD}$ structure solved in the $P2_1$ space group, $UclD^{LD}$ (15 mg ml^{-1}) was crystallized using sitting drop vapour diffusion against a solution containing 16% PEG 4000,

0.1 M Tris-HCl, pH 8.5, 0.2 M magnesium chloride. UclD crystals were flash cooled to 100 K in a solution containing 16% PEG 4000, 0.1 M Tris-HCl, pH 8.5, 0.2 M magnesium chloride and 30% glycerol. Data were collected at beamline ID29 (ESRF, Grenoble, France) to 1.05 Å resolution. Data were indexed and processed with XDS (Kabsch, 2010), scaled and merged using SCALA in the CCP4 suite (Evans and Murshudov, 2013). Data and refinement statistics can be found in Table S3.

For the UclD^{LD} structure solved in the $P2_12_12_1$ space group, UclD^{LD} (10 mg ml^{-1} 10 mM MES pH 5.8) was crystallized by the hanging drop vapour diffusion method against a well solution containing 0.1 M potassium phosphate (monobasic), 0.2 M potassium iodide and 20% PEG 3350. One microlitre of the protein solution was mixed with 1 µl well solution and incubated at 18 °C. Crystals were harvested and transferred to a solution containing 0.1 M potassium phosphate (monobasic), 0.2 M potassium iodide and 20% PEG 3350 supplemented with 20% glycerol before being flash-frozen in a bath of liquid nitrogen. Data were collected at beamline 4.2.2 (ALS Berkeley) to 1.6 Å resolution. Data were indexed and processed with XDS(Kabsch, 2010), scaled and merged AIMLESS in the CCP4 suite(Evans and Murshudov, 2013) and phased with the Single anomalous dispersion (SAD) method using phenix.autosol, and refined with phenix.refine (Adams et al., 2012). Data and refinement statistics can be found in Table S3. r.m.s.d. values were calculating using the DALI server (Holm and ½m, 2010). Structural alignments were performed in PROMALS3D using the default settings. Secondary structure assignments for UclD^{LD} were completed using DSSP.

4.5.14 Differential scanning fluorimetry

Purified UclD^{LD} (1.4 µg per well) was incubated with 5× Sypro orange fluorescent dye in 20 mM Tris, pH 8.0, with or without 10 mM monosaccharide in a total volume of 70 µl. Samples

were heated from 20 °C to 100 °C in 30-s/0.5 °C increments using a Bio-Rad C1000 thermocycler with CFX96 RT–PCR attachment. The reported melting temperatures were determined by the inflection point of the sigmoidal graph.

4.5.15 Data availability

Bacterial V4-16S rRNA data sets have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB19121. Sequences used to examine the carriage of F17-like pili in clinical recurrent UTI isolates were previously published[18] and are deposited in the NCBI under the BioProject accession PRJNA269984. Crystallography data have been deposited in the Protein Data Bank (PDB) under accession codes 5NWP ($P2_1$) and 5VQ5 ($P2_12_12_1$). All other data are available from the corresponding author upon reasonable request.

4.5.16 Code availability

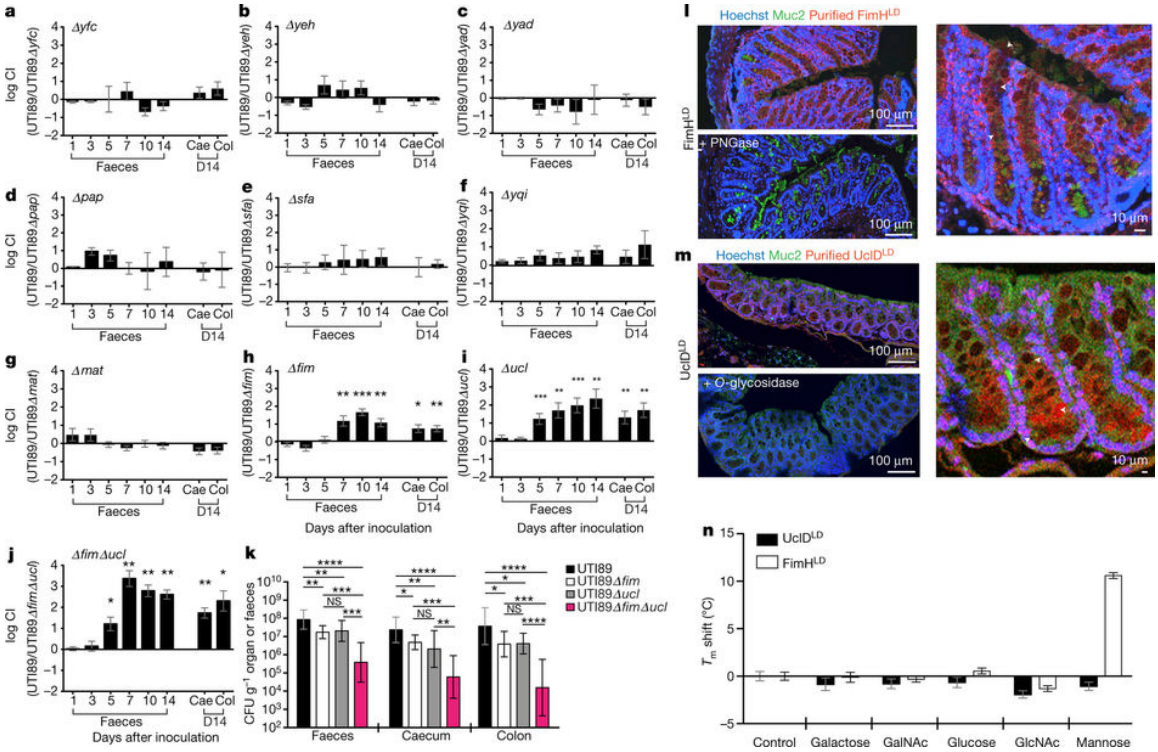
No new code was generated for this study. All software was obtained from publicly available sources; papers describing the software are cited in the text.

4.5.17 Statistical analysis

No statistical methods were used to predetermine sample size. The statistical significance of differences between groups in experiments (excluding competitive infections) was determined by a Mann–Whitney *U* test. The competitive index was defined as: (CFU output strain A/CFU output strain B)/(CFU input strain A/CFU input strain B). For competitive infections, statistical significance was determined by a Wilcoxon signed-rank test. Statistical analyses were performed using Graphpad Prism 7.

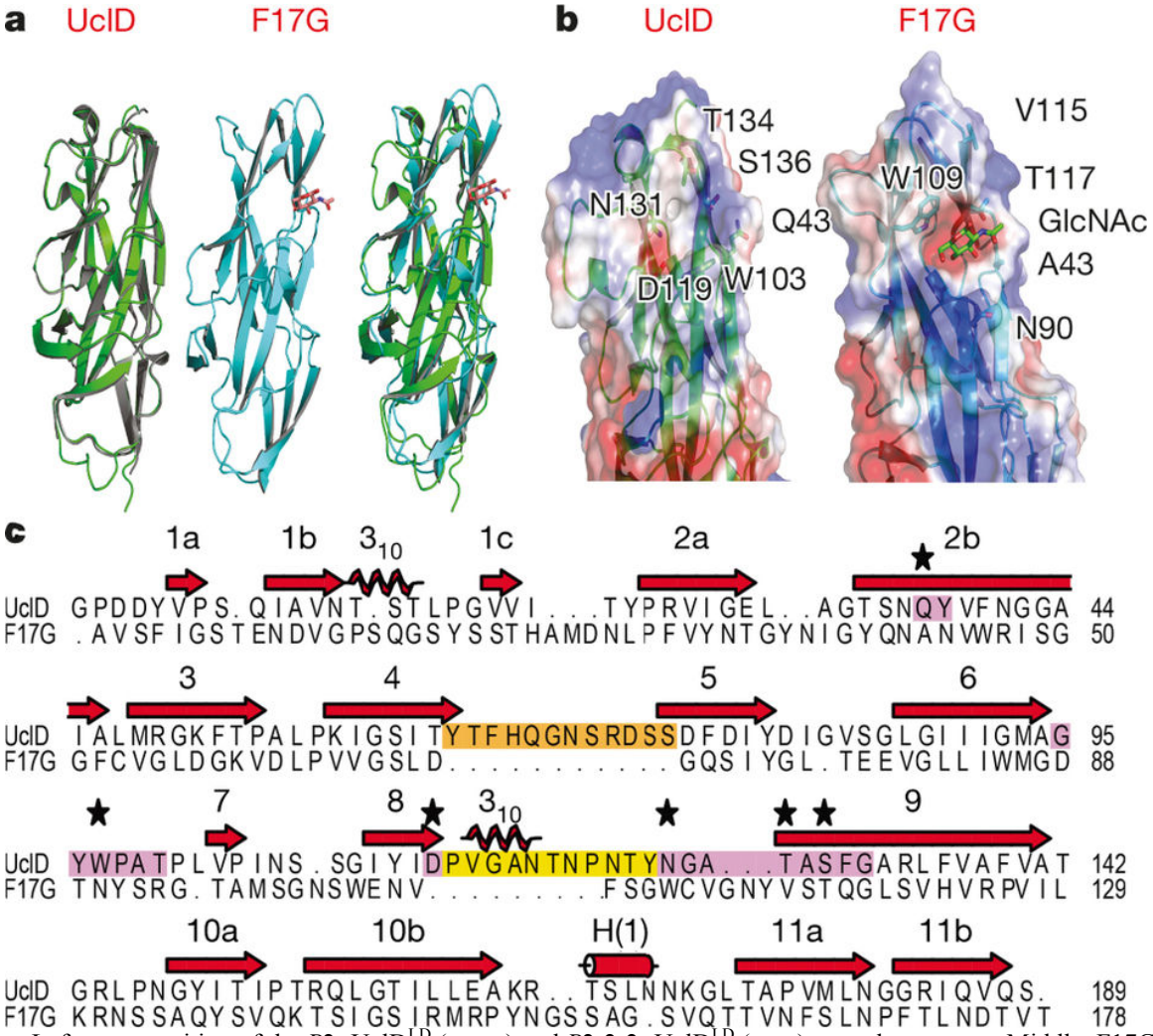
4.6 Figures

4.6.1 Figure 1. Type 1 and F17-like pili promote UPEC intestinal colonization.



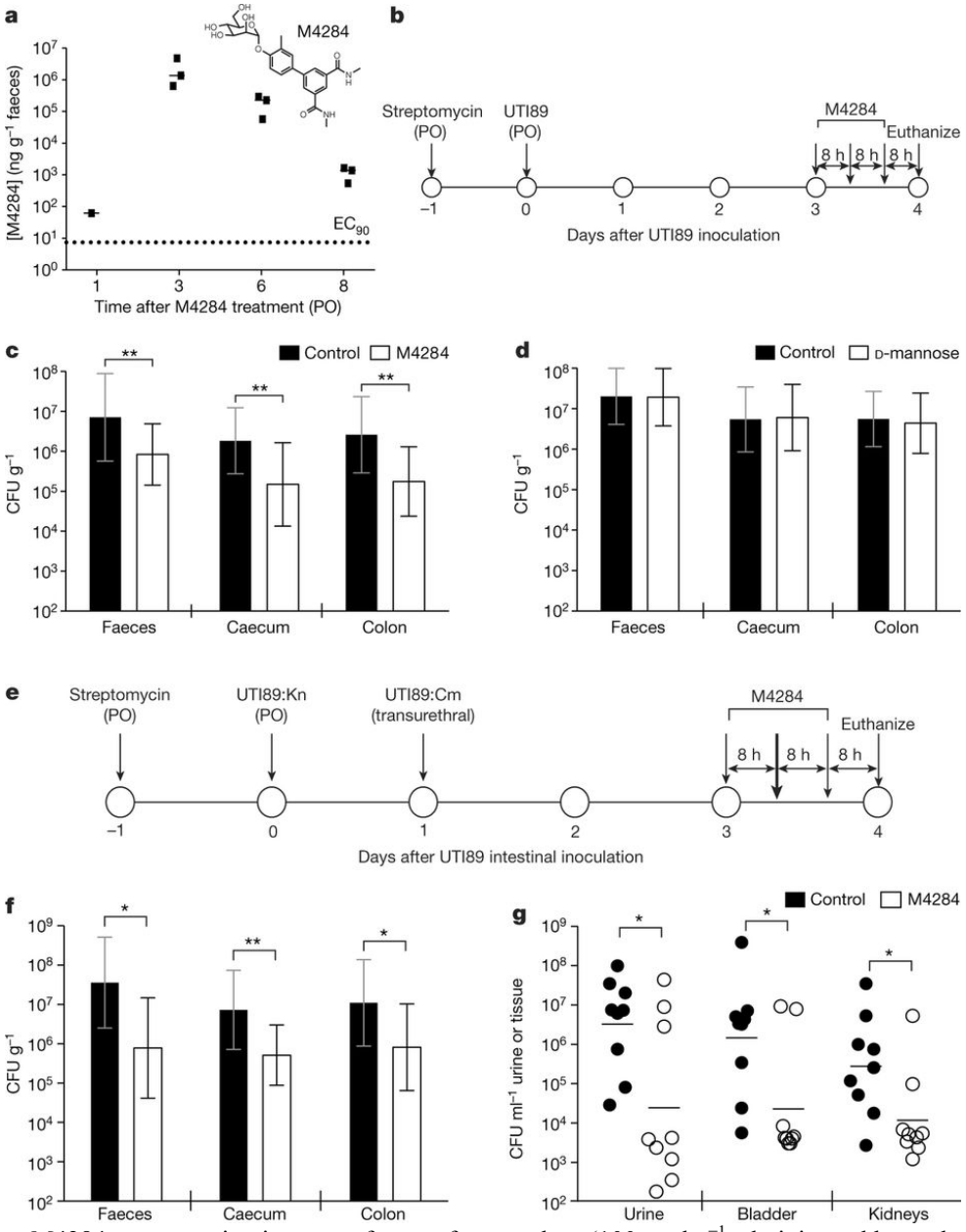
a–**k**, C3H/HeN mice pretreated with streptomycin were concurrently (**a**–**j**) or singly (**k**) colonized with wild-type UTI89 and/or UTI89 lacking one or more CUP operons (*yfc* (**a**), *yeh* (**b**), *yad* (**c**), *pap* (**d**), *sfa* (**e**), *yqi* (**f**), *mat* (**g**), *fim* (**h**), *ucl* (**i**) and both *fim* and *ucl* (**j**)). **l**, **m**, Purified adhesin lectin domains $FimH^{LD}$ (type 1 pili) and $UclD^{LD}$ (F17-like pili) were tested for binding to mouse colonic sections. Sections were stained with Hoechst (blue) and antibodies to the mucus-associated glycoprotein Muc2 (green). Binding of $FimH^{LD}$ and $UclD^{LD}$ was lost by pretreating tissue sections with PNGase F and *O*-glycosidase, respectively. Arrowheads highlight binding by $FimH^{LD}$ or $UclD^{LD}$. **n**, $UclD^{LD}$ does not bind five common monosaccharides. Cae, caecum; CI, competitive index; Col, colon; GalNAc, *N*-acetylgalactosamine; GlcNAc, N-acetylglucosamine; T_m, melting temperature. Data are mean ± se.m. (**a**–**j**, **n**), and geometric mean ± s.d. (**k**). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by Wilcoxon signed-ranked (**a**–**j**) or Mann–Whitney *U* test (**k**, **n**). $n = 5$ mice, 1 replicate (**a**, **d**–**g**); $n = 10$ mice, 2 replicates (**b**, **h**); $n = 6$ mice, 1 replicate (**c**); $n = 14$ mice, 3 replicates (**i**); $n = 8$ mice, 2 replicates (**j**); $n = 12$ mice, 3 replicates (UTI89); $n = 9$ mice, 2 replicates (UTI89Δ*fim*); $n = 15$ mice, 3 replicates (UTI89Δ*ucl*); $n = 10$ mice, 2 replicates (UTI89Δ*fim*Δ*ucl*) (**k**). Replicates are biological (**a**–**k**). $n = 3$ tissue sections, 4 representative images per section (**l**, **m**). $n = 3$ wells, 3 technical replicates (**n**).

4.6.2 Figure 2. Structural analysis of UclD^{LD}.



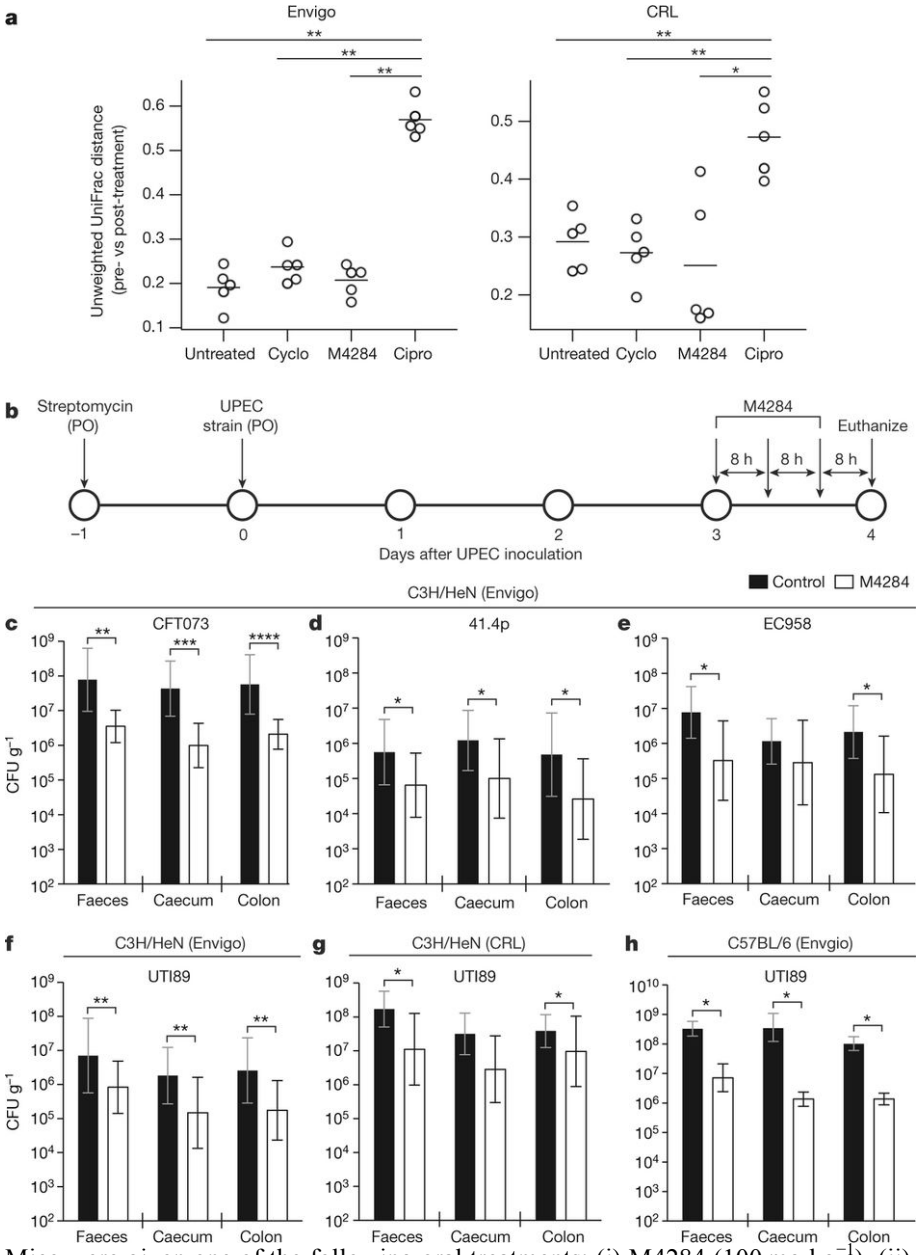
a, Left, superposition of the $P2_1$ UclD^{LD} (green) and $P2_12_12_1$ UclD^{LD} (grey) crystal structures. Middle, F17G adhesin crystal structure (PDB accession 1OIO). Right, superposition of $P2_1$ UclD^{LD} (green) and F17G^{LD} (cyan) structures. **b**, Comparison of residue positioning and electrostatic surface potential of the putative binding site between the UclD^{LD} structures and the known binding site of F17G^{LD}. **c**, Structural alignment of UclD^{LD} and F17G^{LD} amino acid sequences. Residues in the putative UclD^{LD} binding site are highlighted in purple. Insertions in the UclD^{LD} are highlighted in orange and yellow. Starred residues are proposed to mediate UclD ligand binding. β-strands (red arrows), 3_{10} helices (coils) and an α-helix (cylinder) are shown.

4.6.3 Figure 3. Mannoside simultaneously reduces the UPEC intestinal reservoir and treats UTI.



a, M4284 concentration in mouse faeces after one dose (100 mg kg^{-1} administered by oral gavage; PO, *per os*). EC_{90}, 90% effective concentration to inhibit haemagglutination *in vitro*. **b**, C3H/HeN mice were intestinally colonized with UTI89 and given three oral doses of M4284 (100 mg kg^{-1}), vehicle alone (10% cyclodextrin, control), or D-mannose (100 mg kg^{-1}). **c**, **d**, UTI89 levels in the faeces and intestinal segments. **e**, UTI89 was introduced into the gut of C3H/HeN mice by oral gavage and into the bladder by transurethral inoculation before receiving three doses of M4284. Cm, chloramphenicol; Kn, kanamycin. **f**, **g**, UTI89 levels in the gut and urinary tract were assessed. Error bars represent median (**a**), geometric mean ± s.d. (**c**, **d**, **f**); and geometric mean (**g**). *$P < 0.05$, **$P < 0.01$ by Mann–Whitney *U* test. $n = 3$ mice, 1 replicate (**a**); $n = 14$ mice (control); $n = 15$ mice (M4284); 3 replicates (**c**); $n = 10$ mice, 2 replicates (**d**); $n = 9$ mice, 2 replicates (**f**, **g**). All replicates are biological (**a**–**g**).

4.6.4 Figure 4. Mannoside treatment minimally effects the faecal microbiota configuration and targets human UPEC isolates in mice with different genetic backgrounds.

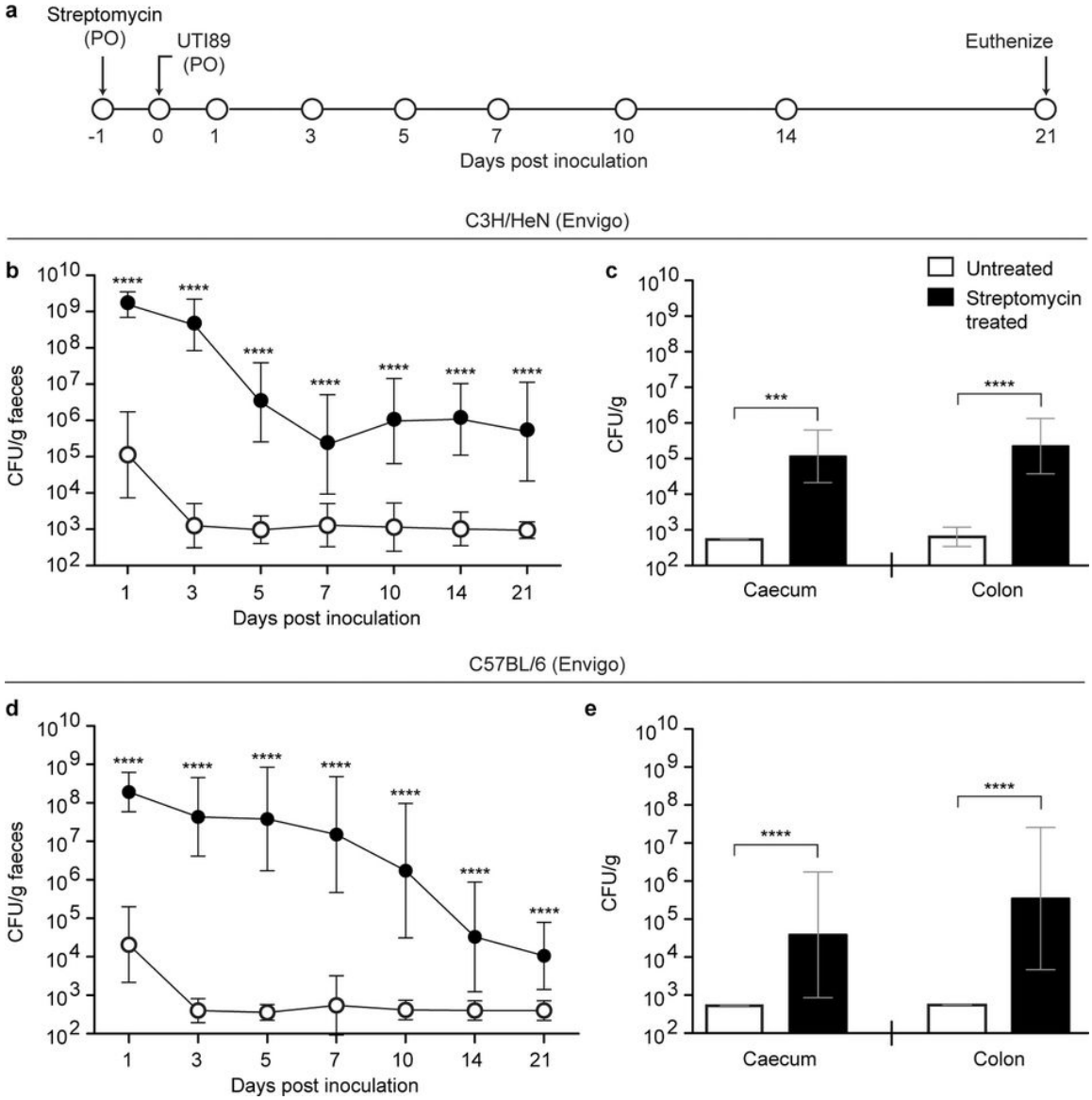


Mice were given one of the following oral treatments: (i) M4284 (100 mg kg^{-1}), (ii) cyclodextrin (cyclo; 10%), (iii) ciprofloxacin (cipro; 15 mg kg^{-1}), or (iv) none (untreated). Faecal community structure was defined by sequencing bacterial 16S rRNA gene amplicons. **a**, For each treatment performed on C3H/HeN mice from Envigo or Charles River Laboratories (CRL), the change in microbiota configuration was determined by measuring the unweighted UniFrac distance between samples obtained from each animal before treatment and 24 h after the last dose (larger UniFrac distance equates to a larger shift in community structure). **b**–**f**, Mice were colonized by oral gavage (PO) of one of four different UPEC strains and given three doses of M4284. **g**, **h**, The ability of M4284 to target UTI89 in C3H/HeN mice from CRL (**g**) and C57BL/6 mice from Envigo (**h**) was also assessed. Error bars represent median (**a**) or geometric mean ± s.d. (**c**–**h**). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$ by Mann–Whitney *U* test. $n = 5$ mice per vendor, 1 replicate (**a**); $n = 10$ mice, 2 replicates (**c**–**e**, **g**); $n = 14$ (control, 10%

cyclodextrin); $n = 15$ (M4284); 3 replicates (**f**). $n = 10$ mice (control, 10% cyclodextrin); $n = 9$ mice (M4284); 2 replicates (**h**). All replicates are biological (**a**–**h**).

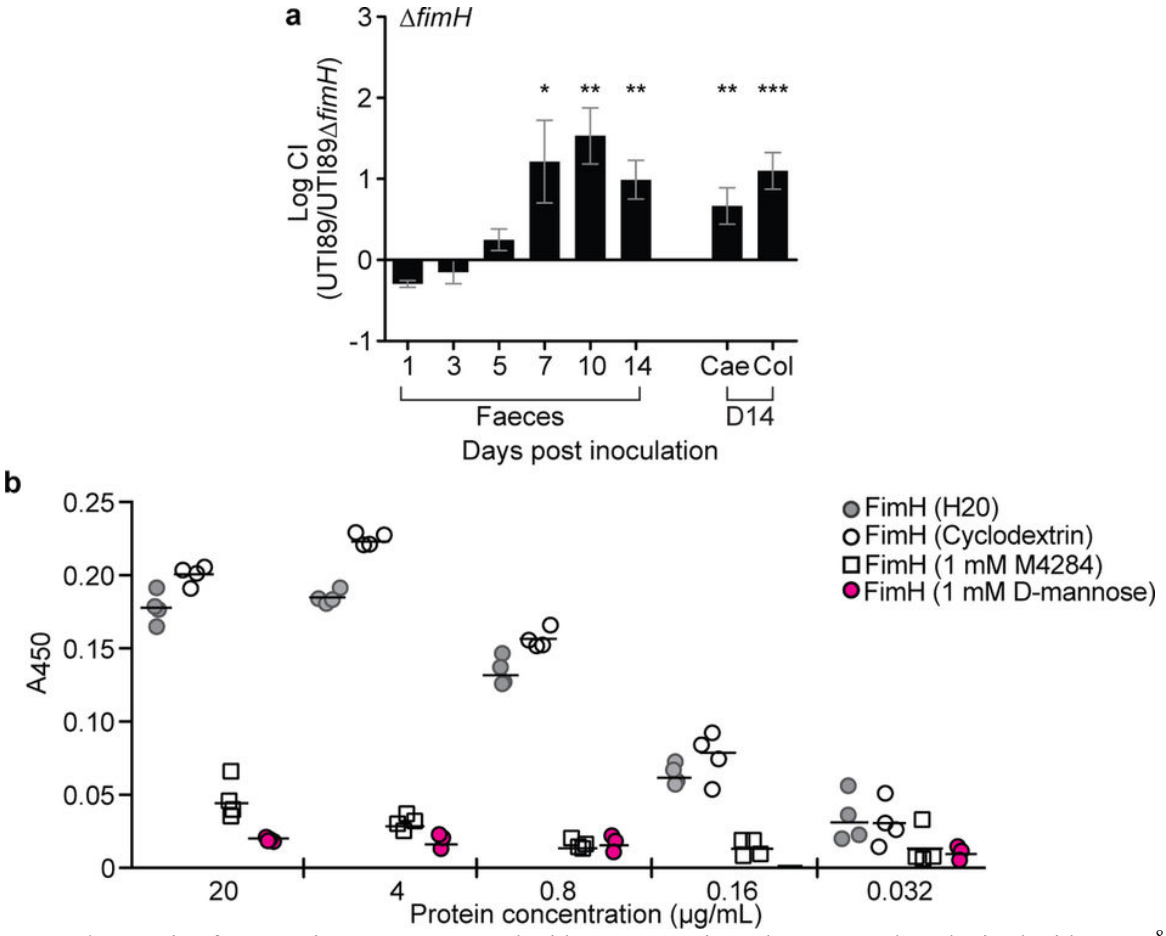
4.7 Supplementary Figures

4.7.1 Figure S1. Streptomycin treatment allows for persistent UTI89 colonization of the caecum and colon in female C3H/HeN and C57BL/6 mice.



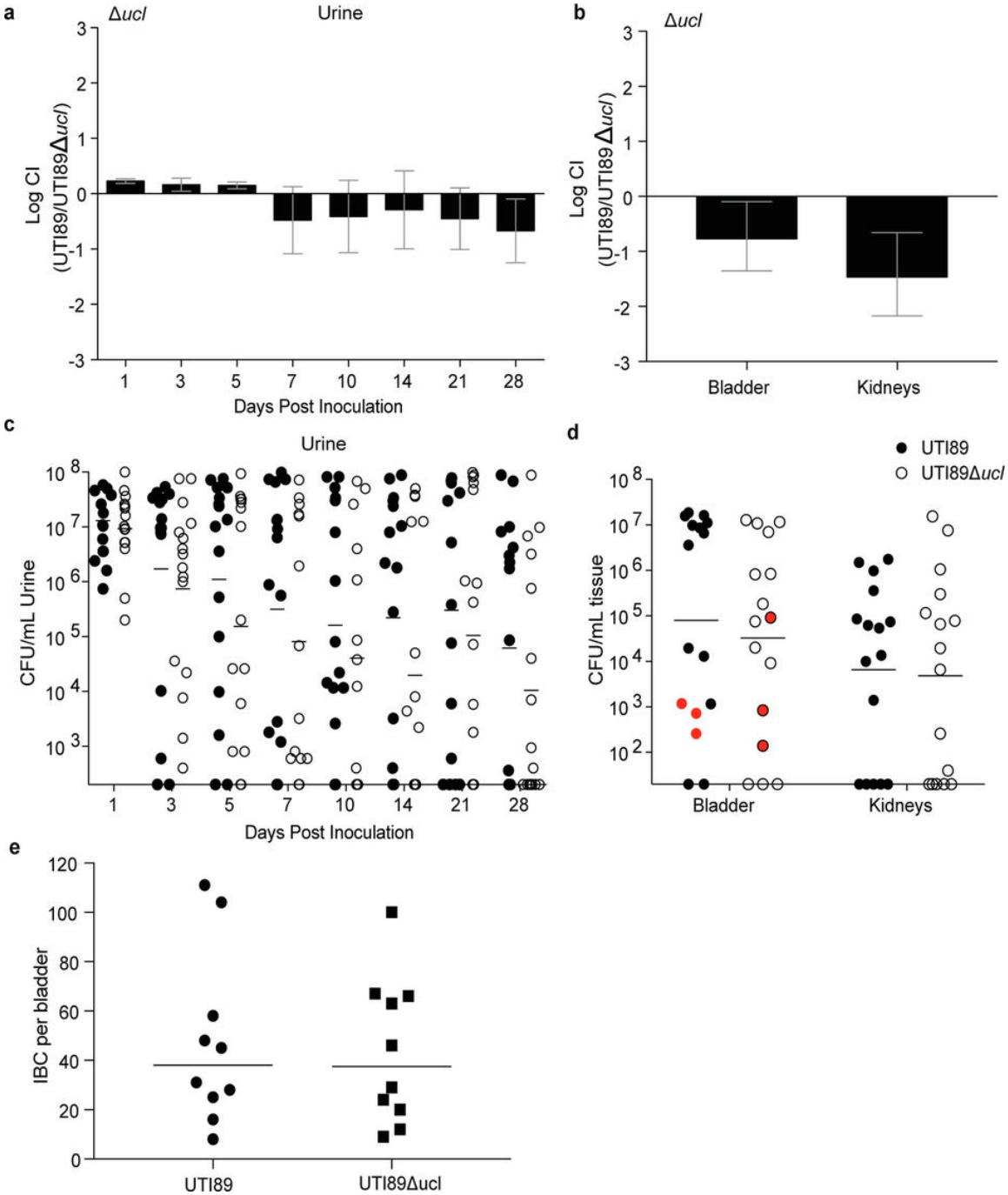
a, Mice were pretreated with streptomycin and subsequently colonized via oral gavage (PO) with UTI89, a prototypical human UPEC cystitis isolate. **b**–**e**, Colonization of UTI89 in C3H/HeN (**b**, **c**) or C57BL/6 (**d**, **e**) mice from Envigo was assessed by quantifying CFU in faecal samples collected over the course of 21 days from mice who did not receive streptomycin (white circles) or mice pretreated with the antibiotic (black circles). CFU analysis of levels of colonization in the caecum and colon were defined by analysing tissue homogenates prepared 21 days after colonization. Symbols represent geometric mean ± s.d. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ by Mann–Whitney U test. $n = 15$ mice, 3 biological replicates (**b**–**e**).

4.7.2 Figure S2. The FimH adhesin is required for type 1 pilus-dependent colonization of the mouse gut and for binding to human intestinal epithelial cells.



a, C3H/HeN mice from Envigo were pretreated with streptomycin and concurrently colonized with 1×10^8 CFU of wild-type UTI89 and UTI89Δ*fimH*. The wild-type strain outcompetes the strain lacking the FimH adhesin. **b**, The ability of $FimH^{LD}$ to bind to Caco-2 cells was assessed by a FimH ELISA. Pre-incubation of $FimH^{LD}$ with D-mannose (1 mM) or M4284 (1 mM) results in significant reductions in FimH binding to Caco-2 cells while 10% cyclodextrin (M4284 vehicle) had no significant effect. All data shown are normalized to wells that were not exposed to the purified adhesin. Data in **a** are mean ± s.e.m, and bars in **b**represent the median. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by Wilcoxon signed-rank test (**a**). $n = 14$ mice, 3 biological replicates (**a**); $n = 4$ wells examining FimH binding to Caco-2 cells per protein concentration, 4 technical replicates (**b**).

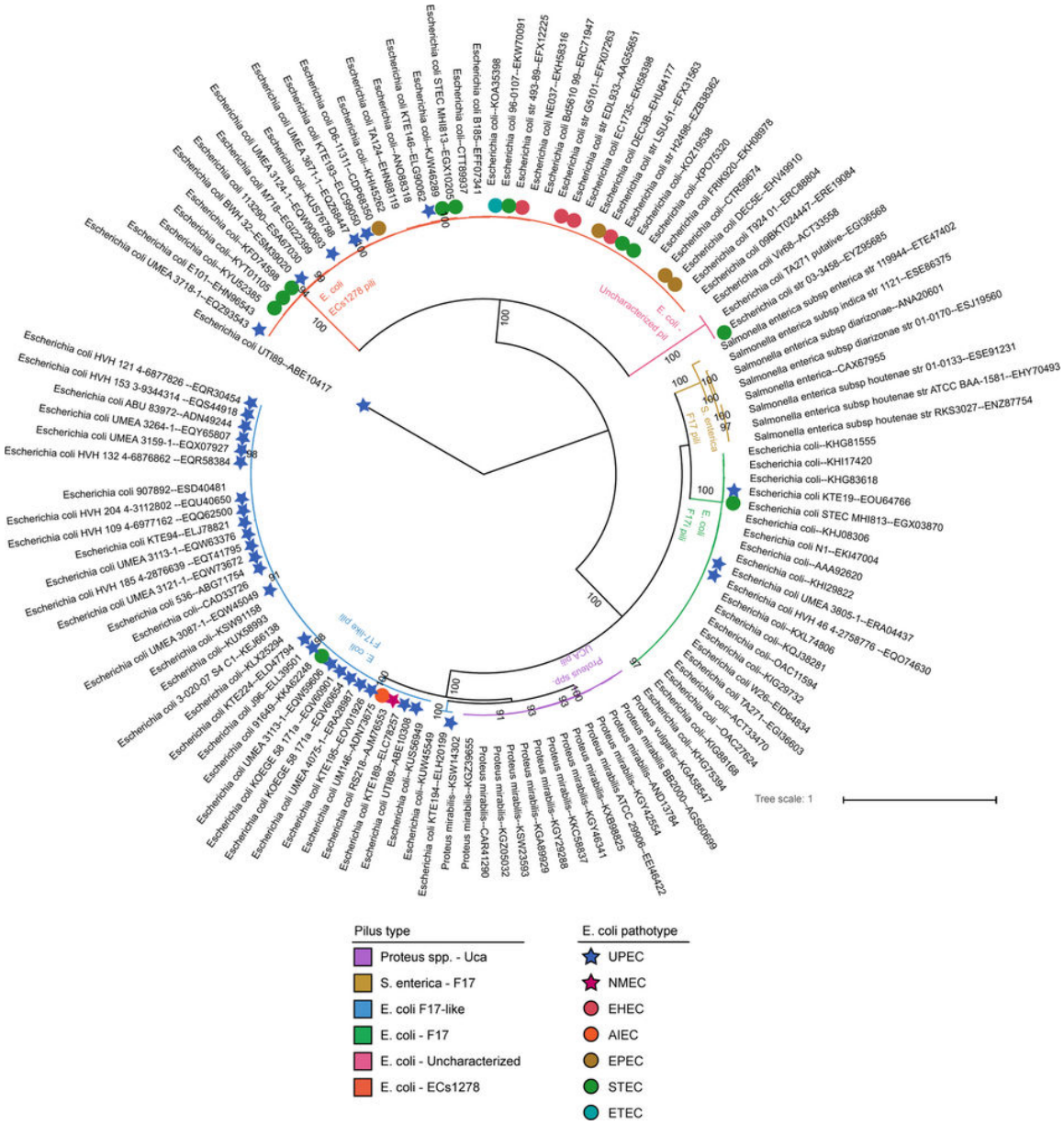
4.7.3 Figure S3. F17-like pili are not required for UTI in mice.



C3H/HeN mice received a transurethral inoculation of wild-type UTI89 and UTI89Δ*ucl*, concurrently (**a**, **b**), or individually (**c**–**e**). **a**, UTI89Δ*ucl* and wild-type strains persist at similar levels in the urine over 28 days in competitive infections. **b**, The two strains are also present at equal levels in the bladder and kidney at the time of euthanization (28 days after infection). **c**, Single infection with the wild-type strain (black circles) or the F17-like mutant strain (white circles) produces similar levels of bacteruria over 28 days. **d**, Single strain infection also produces similar levels of viable cells in homogenates of whole bladder or kidneys collected at the time of euthanization (28 days after infection). There was no statistically significant difference in the number of mice that resolved bacteriuria while maintaining bladder-associated CFUs after transurethral infection with either wild-type UTI89 or UTI89Δ*ucl* (highlighted in red in **d**), suggesting that both strains are capable of forming similar numbers of quiescent intracellular reservoirs. **e**, Mice infected transurethrally with wild-type or Δ*ucl*strains of UTI89 exhibit

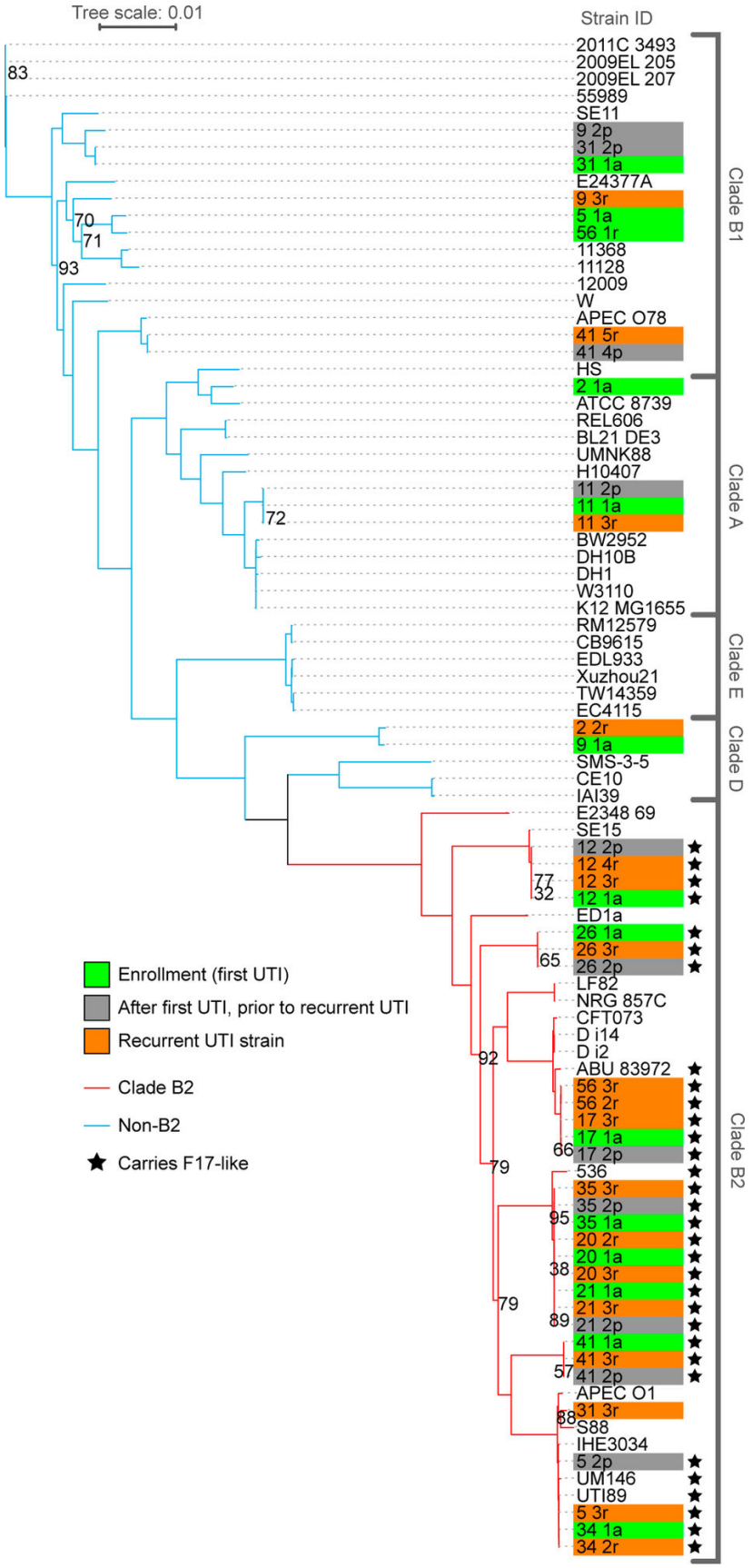
a similar number of IBCs at 6 h in the bladder, indicating that loss of the *ucl* operon does not alter the ability of UTI89 to form IBCs. Error bars represent mean ± s.e.m. (**a**, **b**), geometric mean (**c**, **d**) or median (**e**). No significant difference was detected between any samples by Wilcoxon signed-rank test (**a**, **b**) or Mann–Whitney *U* test (**c**–**e**). $n = 10$ mice, 2 biological replicates (**a**, **b**, **e**); $n = 16$ mice, 3 biological replicates (**c**, **d**).

4.7.4 Figure S4. Distribution of F17 usher homologues in members of Enterobacteriaceae.



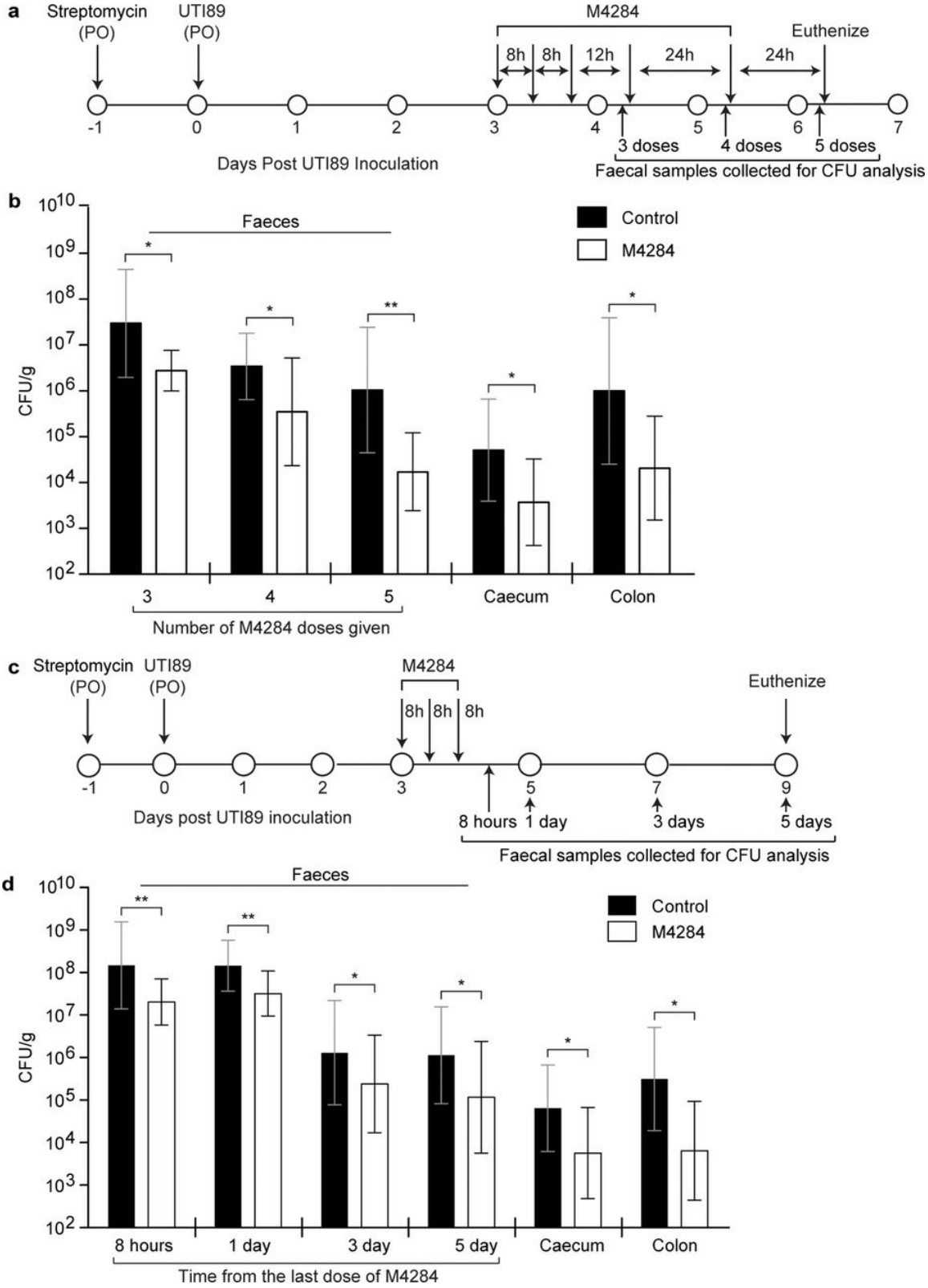
The phylogenetic relationships between F17 homologues were estimated using the sequence of the usher genes. Branch colours indicate host strain and pilus identity, and coloured symbols indicate the annotated pathotype of the *E. coli* strain for each sequence as determined by publically available annotations. Stars indicate extraintestinal pathogenic *E. coli* (ExPEC) strains, and circles indicate intestinal pathogenic *E. coli* strains. Carriage of F17-like pili is enriched in UPEC strains, whereas F17 and ECs1278 pili are more common in intestinal pathogens such as ETEC and EHEC, respectively. The strain names for each sequence and ENA accessions are given. Numbers beneath the branches indicate the percentage of support from 1,000 bootstrap replicates (numbers greater than 80% are shown).

4.7.5 Figure S5. Phylogenetic distribution of F17-like carriage in UPEC from patients with recurrent UTI.


Tree scale: 0.01
Strain ID
2011C 3493
2009EL 205
2009EL 207
55989
SE11
9 2p
31 2p
31 1a
E24377A
9 3r
5 1a
56 1r
11368
11128
12009
W
APEC O78
41 5r
41 4p
HS
2 1a
ATCC 8739
REL606
BL21 DE3
UMNK88
H10407
11 2p
11 1a
11 3r
BW2952
DH10B
DH1
W3110
K12 MG1655
RM12579
CB9615
EDL933
Xuzhou21
TW14359
EC4115
2 2r
9 1a
SMS-3-5
CE10
IAI39
E2348 69
SE15
12 2p
12 4r
12 3r
12 1a
ED1a
26 1a
26 3r
26 2p
LF82
NRG 857C
CFT073
D i14
D i2
ABU 83972
56 3r
56 2r
17 3r
17 1a
17 2p
536
35 3r
35 2p
35 1a
20 2r
20 1a
20 3r
21 1a
21 3r
21 2p
41 1a
41 3r
41 2p
APEC O1
31 3r
S88
IHE3034
5 2p
UM146
UTI89
5 3r
34 1a
34 2r
Clade B1
Clade A
Clade E
Clade D
Clade B2
83
70
71
93
72
77
32
65
92
66
79
95
38
79
89
57
88
Enrollment (first UTI)
After first UTI, prior to recurrent UTI
Recurrent UTI strain
Clade B2
Non-B2
Carries F17-like

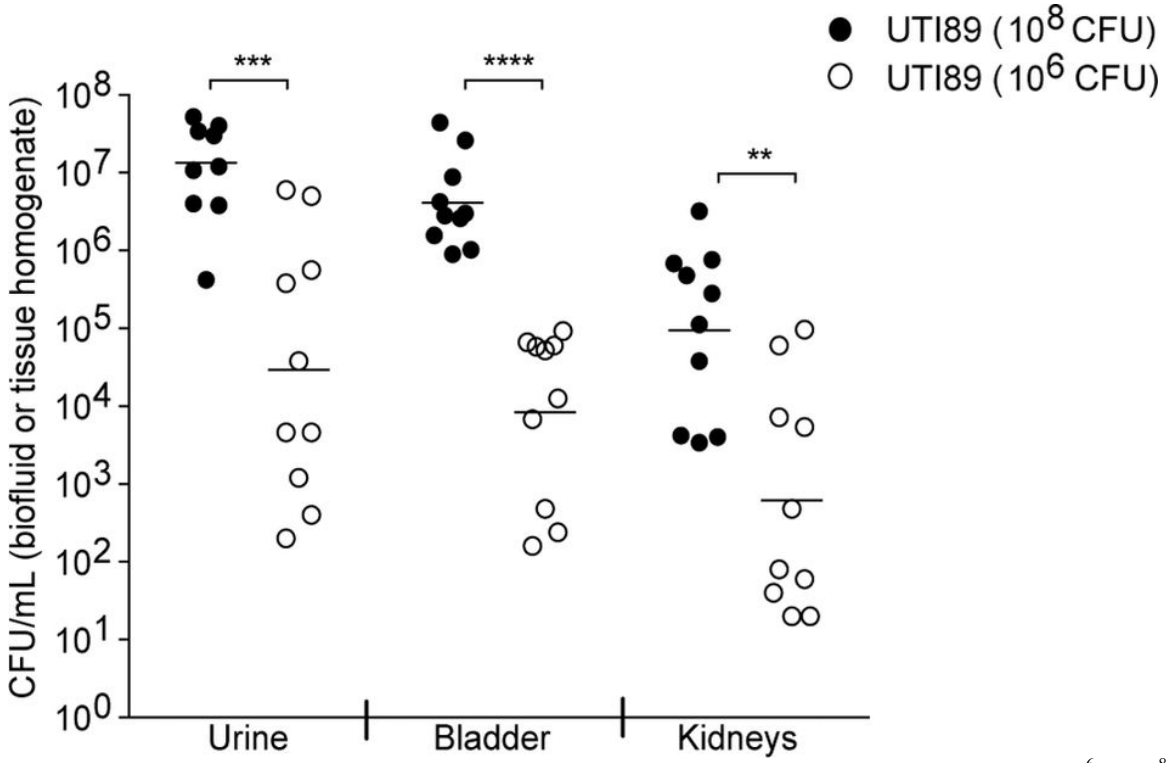
The phylogeny of a set of clinical UPEC strains ($n = 43$ with taxon labels highlighted in green, orange or grey) was contextualized with reference *E. coli* strains ($n = 46$, unhighlighted taxon labels) by comparing the concatenated single-copy, core genes of the strains using the RAxML algorithm and the GTRCAT model[46]. Highlighted taxon labels indicate UPEC isolates collected at enrolment (green) and during recurrent UTI (orange). In all cases, patients cleared each infection before recurrence, no patient exhibited signs of asymptomatic bacteriuria. The study design also allowed for the collection, from cohort participants, of *E. coli* isolates present in the urine in the days leading up to their clinical visit and recurrent UTI diagnosis (highlighted in grey)[17]. Branch lines indicate phylogenetic background for strains from clade B2 (red branch lines) and non-B2 clades (blue branch lines). Carriage of F17-like pili (black stars) was limited to the B2 clade and enriched within recurrent UTI UPEC isolates. Bootstrap supports are indicated at internal nodes. Bootstrap values greater than 95 have been removed. The clade to which each strain belongs is indicated in brackets to the right.

4.7.6 Figure S6. Testing the effects of more prolonged dosing of M4284 and analysis of the duration of its effects.



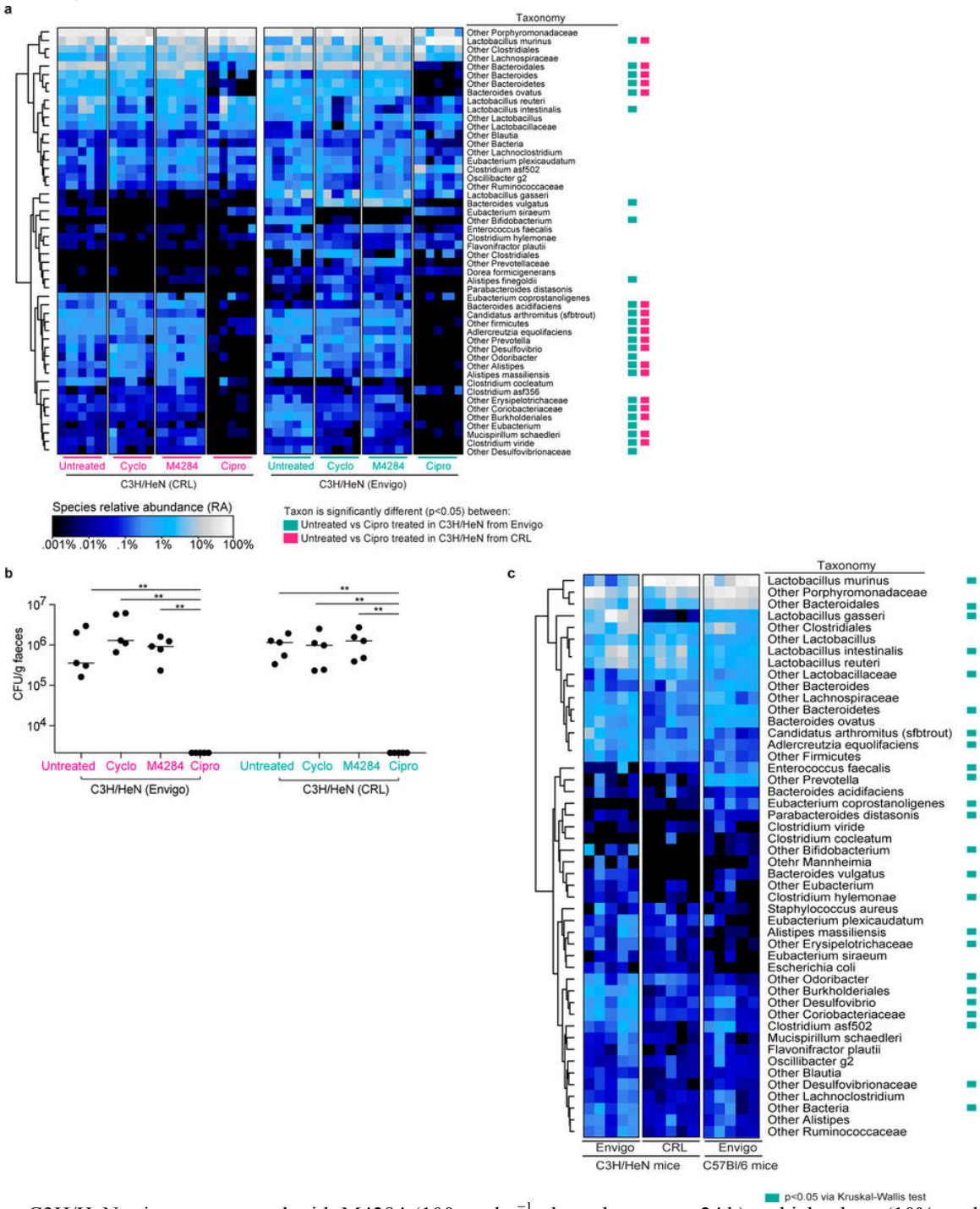
a, Experimental design. **b**, Animals treated as in **a** show a continued decrease in UTI89 levels in their faeces (samples were processed after 3, 4 and 5 doses of M4284), and at the time of euthanization in the caecum and colon, compared to control mice treated with vehicle alone (control, 10% cyclodextrin). **c**, **d**, The effects of mannoside treatment persist 5 days after M4284 exposure. Data are geometric mean ± s.d. $*P < 0.05$, $**P < 0.01$ by Mann–Whitney *U* test. $n = 9$ mice (control); $n = 10$ mice (M4284), 2 biological replicates (**b**); $n = 16$ mice, 3 biological replicates (**d**).

4.7.7 Figure S7. The severity of UTI outcome is directly linked to the dose of UTI89 inoculated into the urinary tract.



C3H/HeN mice (Envigo) were given an experimental UTI via transurethral inoculation of either 10^6 or 10^8 CFU of UTI89. The doses were chosen to represent the reduction observed in intestinal UTI89 titres before and after treatment with the M4284 mannoside. Mice were euthanized 24 h after inoculation, and UTI89 titres in urine, bladder and kidneys were defined by quantifying CFU. Mice receiving the 10^6 dose of UTI89 had significantly fewer bacteria in all three biospecimen types, indicating an important relationship between the number of bacteria introduced into the urinary tract and the severity of UTI outcome. Bars represent geometric means. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by Mann–Whitney *U* test. $n = 10$ mice, 2 biological replicates.

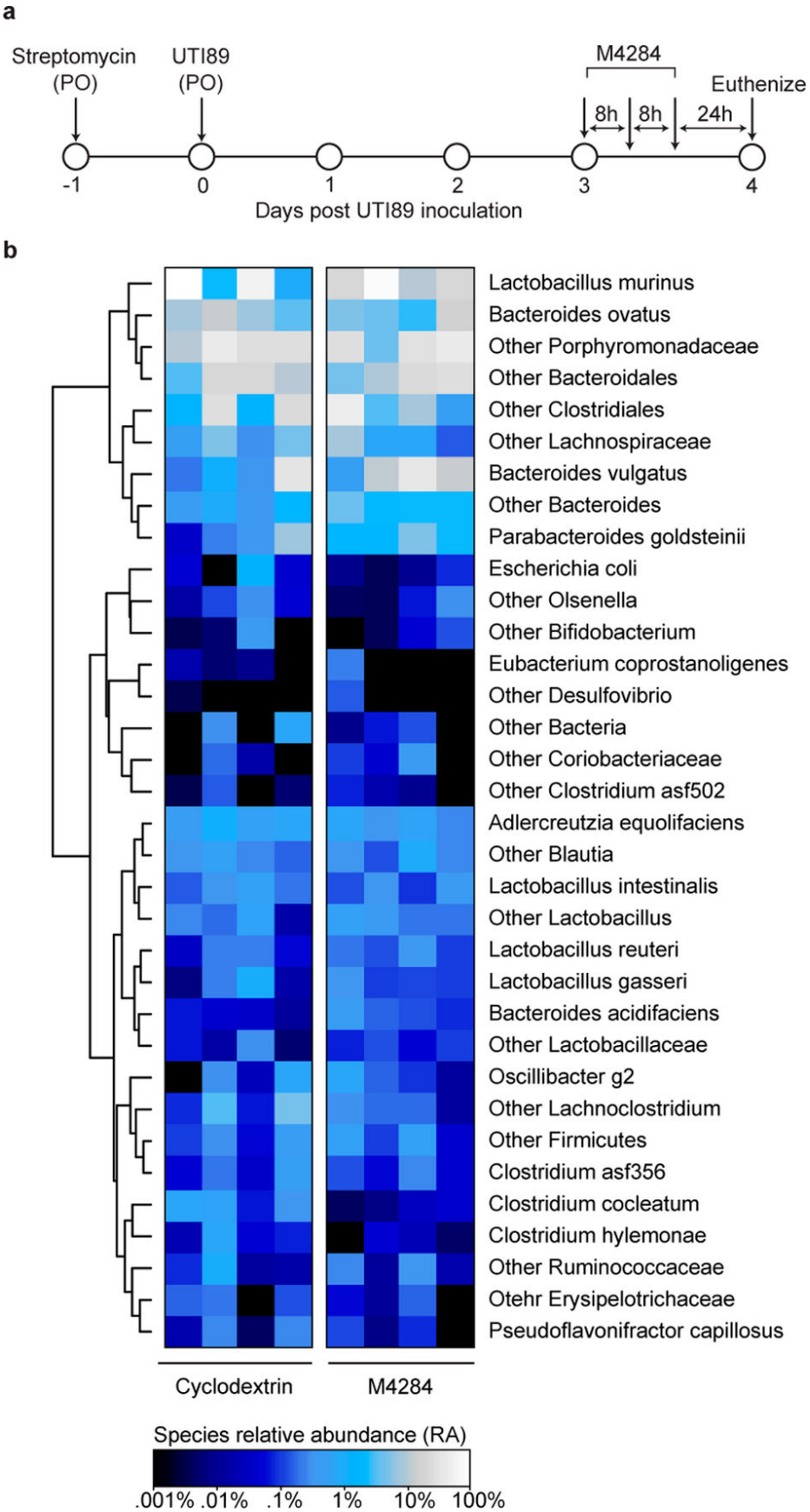
4.7.8 Figure S8. 16S rRNA-based comparison of faecal bacterial communities in mice obtained from Envigo and CRL and mice of different genetic backgrounds from a common vendor.



a, C3H/HeN mice were treated with M4284 (100 mg kg^{-1}, three doses over 24 h), vehicle alone (10% cyclodextrin, three doses over 24 h), or ciprofloxacin (15 mg kg^{-1}, two doses over 24 h). Untreated mice served as reference controls. Heat maps show the effect of each of the treatments on animals from CRL and Envigo. Each row represents a species-level bacterial taxon, while each column represents a mouse sampled 24 h after the termination of the indicated treatment. Coloured boxes next to the taxon names indicate species whose relative abundance was

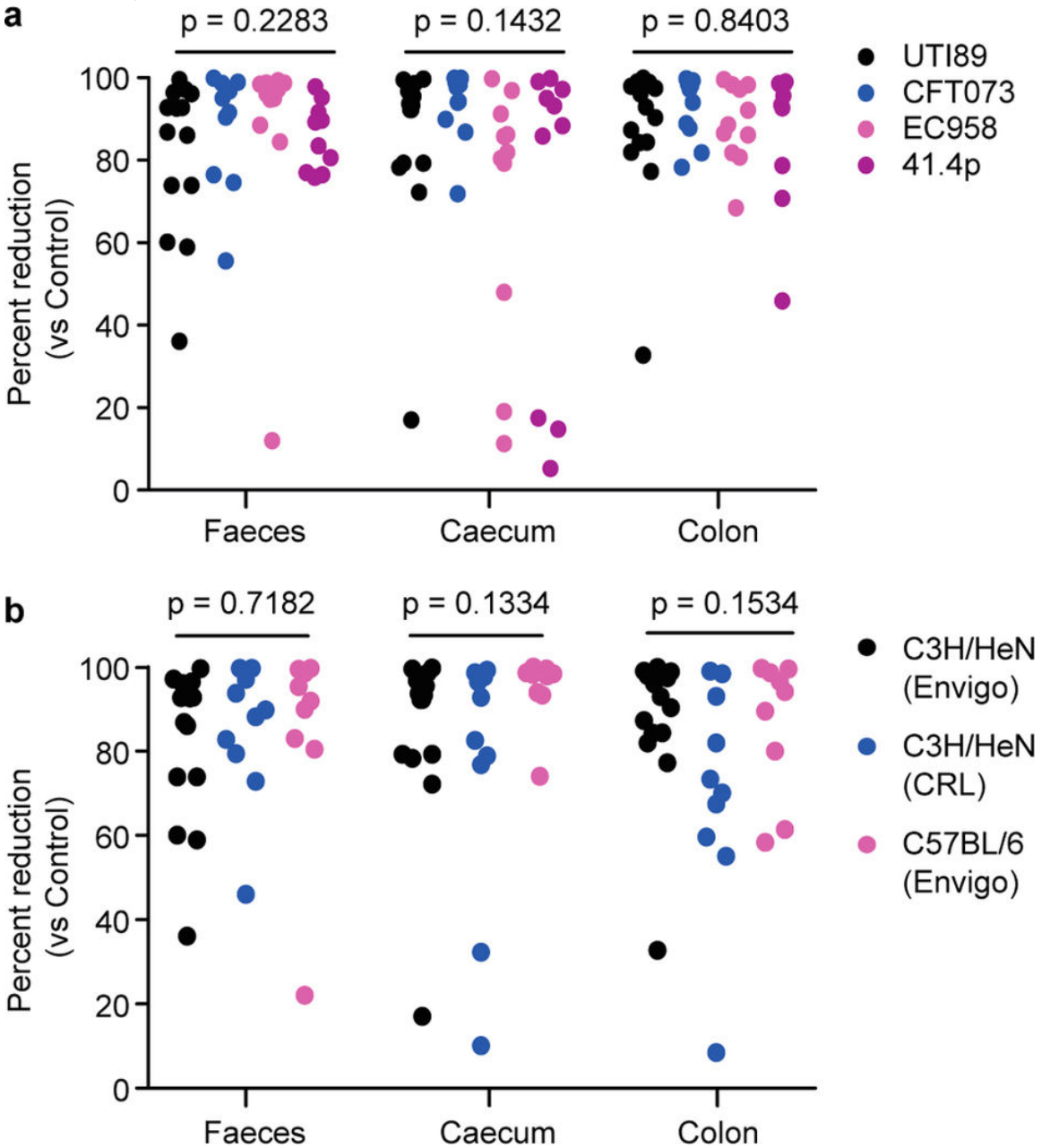
significantly changed by ciprofloxacin treatment ($P < 0.05$; Wilcoxon signed-rank test with false discovery rate (FDR) correction). Individual comparisons between untreated and other treatment types did not disclose changes that were statistically significant by Wilcoxon signed-rank test with FDR correction. **b**, Corresponding faecal samples collected 24 h after treatments (as shown in Extended Data Fig. 8a) were homogenized, diluted serially, and plated on MacConkey medium. The abundance of bacteria capable of growing on the selective medium was similar between faecal samples taken from untreated mice and those collected 24 h after treatment with cyclodextrin and M4284. No colonies were detected from faecal samples collected 24 h after ciprofloxacin treatment. **c**, Comparison of the representation of bacterial taxa in the faecal microbiota of untreated mice obtained from different vendors or representing different genetic backgrounds. Each row in the heat map represents a species-level taxon, while each column represents a mouse of the indicated genetic background from the indicated vendor. Coloured boxes indicate species whose relative abundances were significantly different ($P < 0.05$) between all three groups of animals (Kruskal–Wallis test with FDR correction). Rows of each heat map were hierarchically clustered according to pair-wise distances using Pearson correlation. $n = 5$ mice per treatment type, 1 biological replicate (**a**); $n = 5$ mice, 1 biological replicate (**b**); $n = 5$ mice per vendor/mouse strain, 1 biological replicate (**c**). Bars denote median. $^{**}P < 0.001$, Mann–Whitney U test (**b**).

4.7.9 Figure S9. The configuration of the faecal microbiota of C3H/HeN mice pretreated with streptomycin and colonized with UTI89 is minimally altered by M4284 treatment.


a
Streptomycin
(PO)
UTI89
(PO)
M4284
Euthenize
8h
8h
24h
-1
0
1
2
3
4
Days post UTI89 inoculation
b
Lactobacillus murinus
Bacteroides ovatus
Other Porphyromonadaceae
Other Bacteroidales
Other Clostridiales
Other Lachnospiraceae
Bacteroides vulgatus
Other Bacteroides
Parabacteroides goldsteinii
Escherichia coli
Other Olsenella
Other Bifidobacterium
Eubacterium coprostanoligenes
Other Desulfovibrio
Other Bacteria
Other Coriobacteriaceae
Other Clostridium asf502
Adlercreutzia equolifaciens
Other Blautia
Lactobacillus intestinalis
Other Lactobacillus
Lactobacillus reuteri
Lactobacillus gasseri
Bacteroides acidifaciens
Other Lactobacillaceae
Oscillibacter g2
Other Lachnoclostridium
Other Firmicutes
Clostridium asf356
Clostridium cocleatum
Clostridium hylemonae
Other Ruminococcaceae
Otehr Erysipelotrichaceae
Pseudoflavonifractor capillosus
Cyclodextrin
M4284
Species relative abundance (RA)
.001% .01% .1% 1% 10% 100%

a, C3H/HeN mice from Envigo were pretreated with streptomycin and 24 h later colonized with UTI89 by oral gavage. Three days after inoculation, animals were treated with three doses of M4284 (100 mg kg^{-1}, three doses over 24 h) or vehicle alone (10% cyclodextrin, 3 doses over 24 h). Faecal samples were collected 24 h after the last dose of M4284 or vehicle. **b**, Heat map showing the effect of each treatment type. Each row represents a bacterial species-level taxon, while each column represents a mouse 24 h after the indicated treatment. Rows of the heat map were hierarchically clustered according to pair-wise distances using Pearson correlation. No treatments produced changes that were statistically significant, as judged by Wilcoxon signed-rank test with FDR correction. $n = 4$ mice per treatment type, 1 biological replicate.

4.7.10 Figure S10. The configuration of the faecal microbiota of C3H/HeN mice pretreated with streptomycin and colonized with UTI89 is minimally altered by M4284 treatment.



a, The percentage reduction in CFU for the indicated UPEC strains from M4284-treated versus untreated control C3H/HeN mice obtained from Envigo (based on data in Fig. 4c–f). **b**, CFU data obtained from C3H/HeN mice from Envigo and CRL and C57BL/6 mice from Envigo (based on data in Fig. 4f–h). *P* values calculated using Kruskal–Wallis test. $n = 14$ mice, 3 biological replicates (UTI89); $n = 10$ mice, 2 biological replicates (CFT073, EC958 and 41.4p) (**a**); $n = 14$ mice, 3 biological replicates (C3H/HeN from Envigo); $n = 10$ mice, 2 biological replicates (C3H/HeN from CRL); and $n = 9$ mice, 2 biological replicates (C57BL/6 from Envigo) (**b**).

4.8 Supplementary Tables

4.8.1 Table S1. The phylogenetic relationships between F17 homologs as determined by comparing relatedness of bacterial usher gene sequences.

Available in Appendix A

4.8.2 Table S2. The carriage of F17-like pili in rUTI UPEC strains.

Strain	Pathotype	Clade	BioProject	F17-like Carriage	Sample Time Point**
536	UPEC	B2	PRJNA16235	Yes	
11128	EHEC	B1	PRJDA32513		
11368	EHEC	B1	PRJDA32509		
12009	EHEC	B1	PRJDA32511		
55989	EAEC	B1	PRJNA33413		
2009EL-2050	EHEC	B1	PRJNA81097		
2009EL-2071	EHEC	B1	PRJNA81099		
2011C-3493	EHEC	B1	PRJNA81095		
ABU 83972	ABU	B2	PRJNA38725	Yes	
APEC O1	APEC	B2	PRJNA16718		
APEC O78	APEC	B1	PRJNA184588		
ATCC 8739	Commensal	A	PRJNA18083		
BL21(DE3)	Lab Strain	A	PRJNA20713		
BW2952	Lab Strain	A	PRJNA33775		
CB9615	EPEC	E	PRJNA42729		
CE10	NMEC	F	PRJNA63597		
CFT073	UPEC	B2	PRJNA313		
clone D i14	UPEC	B2	PRJNA52023		
clone D i2	UPEC	B2	PRJNA52021		
DH1	Lab Strain	A	PRJDA52077		
DH10B	Lab Strain	A	PRJNA20079		
E2348/69	EPEC	B2	PRJEA32571		
E24377A	ETEC	B1	PRJNA13960		
EC4115	EHEC	E	PRJNA27739		
ED1a	Commensal	B2	PRJNA33409		
EDL933	EHEC	E	PRJNA259		
H10407	ETEC	A	PRJEA42749		
HS	Commensal	A	PRJNA13959		
IAI39	UPEC	F	PRJNA33411		
IHE3034	NMEC	B2	PRJNA43693		
K12 MG1655	Commensal	A	PRJNA40075		
LF82	AIEC	B2	PRJNA33825		
NRG 857C	AIEC	B2	PRJNA41221		

REL606	Lab Strain	A	PRJNA18281		
RM12579	EPEC	E	PRJNA68245		
S88	Commensal	B2	PRJNA33375		
SE11	Commensal	B1	PRJNA18057		
SE15	Commensal	B2	PRJDA19053		
SMS-3-5	Environmental	F	PRJNA19469		
TW14359	EHEC	E	PRJNA30045		
UM146	AIEC	B2	PRJNA50883	Yes	
UMNK88	ETEC	A	PRJNA42137		
UTI89	UPEC	B2	PRJNA16259	Yes	
W	Lab Strain	B1	PRJNA48011		
W3110	Lab Strain	A	PRJNA16351		
Xuzhou21	EHEC	E	PRJNA45823		
2.1a*	UPEC	A	PRJNA269984		Enrollment UTI
2.2r*	UPEC	D	PRJNA269984		Different-strain rUTI
5.1a*	UPEC	B1	PRJNA269984		Enrollment UTI
5.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
5.3r*	UPEC	B2	PRJNA269984	Yes	Different-strain rUTI
9.1a*	UPEC	D	PRJNA269984		Enrollment UTI
9.2p*	UPEC	B1	PRJNA269984		Prior to Recurrence
9.3r*	UPEC	B1	PRJNA269984		Different-strain rUTI
11.1a*	UPEC	A	PRJNA269984		Enrollment UTI
11.2p*	UPEC	A	PRJNA269984		Prior to Recurrence
11.3r*	UPEC	A	PRJNA269984		Same-strain rUTI
12.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
12.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
12.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
12.4r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
17.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
17.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
17.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
20.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
20.2r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
20.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
21.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
21.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
21.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
26.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
26.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
26.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
31.1a*	UPEC	B1	PRJNA269984		Enrollment UTI

31.2p*	UPEC	B1	PRJNA269984		Prior to Recurrence
31.3r*	UPEC	B2	PRJNA269984		Different-strain rUTI
34.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
34.2r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
35.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
35.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
35.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
41.1a*	UPEC	B2	PRJNA269984	Yes	Enrollment UTI
41.2p*	UPEC	B2	PRJNA269984	Yes	Prior to Recurrence
41.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI
41.4p*	UPEC	B1	PRJNA269984		Prior to Recurrence
41.5r*	UPEC	B1	PRJNA269984		Different-strain rUTI
56.1a*	UPEC	B1	PRJNA269984		Enrollment UTI
56.2r*	UPEC	B2	PRJNA269984	Yes	Different-strain rUTI
56.3r*	UPEC	B2	PRJNA269984	Yes	Same-strain rUTI

* Collected during a previous longitudinal study on recurrent UTI
** Sample time point as defined in [REF]

4.9 Acknowledgements

We thank A. Earl and A. Manson for assistance in acquiring genomic data, J. Nix and Molecular Biology Consortium (Beamline 4.2.2) for collection/processing crystallographic data, L. Vereecke and A. Goncalves at VIB Bio Imaging Core for support with tissue sectioning/microscopy, Z. Han for synthesizing M4284, K. Tamadonfar for assistance with tissue binding studies, and M. Lukaszczyk for help with protein production.

4.9.1 Funding

This work was supported by grants from the NIH (K08AI113184 (A.L.K.), RO1DK051406 (S.J.H) R01AI048689 (S.J.H.), P50DK064540 (S.J.H.), RC1DK086378 (S.J.H.), DK30292 (J.I.G.) and 1F31DK107057 (C.N.S.)), FWO-Flanders (G030411N), Hercules Foundation (UABR/09/005) and VIB PRJ9.

4.9.2 Conflicts of interest

J.W.J. and S.J.H. are inventors on US patent US8937167 B2, which covers the use of mannoside-based FimH ligand antagonists for the treatment of disease. J.W.J. and S.J.H. have ownership interest in Fimbrion Therapeutics, and may benefit if the company is successful in marketing mannosides.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2012). PHENIX: a comprehensive Python-based system for macromolecular structure solution (Chester, England: International Union of Crystallography).

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403–410.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences *108 Suppl 1*, 4516–4522.

Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I., Kim, S.G., Li, H., Gao, Z., Mahana, D., et al. (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell *158*, 705–721.

Cusumano, C.K., Pinkner, J.S., Han, Z., Greene, S.E., Ford, B.A., Crowley, J.R., Henderson, J.P., Janetka, J.W., and Hultgren, S.J. (2011). Treatment and prevention of urinary tract infection with orally active FimH inhibitors. Science Translational Medicine *3*, 109ra115–109ra115.

Czaja, C.A., Stamm, W.E., Stapleton, A.E., Roberts, P.L., Hawn, T.R., Scholes, D., Samadpour, M., Hultgren, S.J., and Hooton, T.M. (2009). Prospective Cohort Study of Microbial and Inflammatory Events Immediately Preceding *Escherichia coli* Recurrent Urinary Tract Infection in Women. J. Infect. Dis. *200*, 528–536.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U.S.a. *97*, 6640–6645.

Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences *108 Suppl 1*, 4554–4561.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Flores-Mireles, A.L., Walker, J.N., Caparon, M., and Hultgren, S.J. (2015). Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nature Reviews Microbiology *13*, 269–284.

Foxman, B. (2014). Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect. Dis. Clin. North Am. *28*, 1–13.

Galtier, M., De Sordi, L., Maura, D., Arachchi, H., Volant, S., Dillies, M.A., and Debarbieux, L. (2016). Bacteriophages to reduce gut carriage of antibiotic resistant uropathogens with low impact on microbiota composition. Environ Microbiol *18*, 2237–2245.

Han, Z., Pinkner, J.S., Ford, B., Chorell, E., Crowley, J.M., Cusumano, C.K., Campbell, S., Henderson, J.P., Hultgren, S.J., and Janetka, J.W. (2012). Lead optimization studies on FimH antagonists: discovery of potent and orally bioavailable ortho-substituted biphenyl mannosides. J. Med. Chem. *55*, 3945–3959.

Holm, L., and ½m, P.R. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Research.

Hung, C.-S., Dodson, K.W., and Hultgren, S.J. (2009). A murine model of urinary tract infection. Nat Protoc *4*, 1230–1243.

Jarvis, C., Han, Z., Kalas, V., Klein, R., Pinkner, J.S., Ford, B., Binkley, J., Cusumano, C.K., Cusumano, Z., Mydock McGrane, L., et al. (2016). Antivirulence Isoquinolone Mannosides: Optimization of the Biaryl Aglycone for FimH Lectin Binding Affinity and Efficacy in the Treatment of Chronic UTI. ChemMedChem *11*, 367–373.

Johansson, M.E.V., and Hansson, G.C. (2012). Preservation of mucus in histological sections, immunostaining of mucins in fixed tissue, and localization of bacteria with FISH. Methods Mol. Biol. *842*, 229–235.

Jones, C.H., Pinkner, J.S., Roth, R., Heuser, J., Nicholes, A.V., Abraham, S.N., and Hultgren, S.J. (1995). FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. Proc. Natl. Acad. Sci. U.S.a. *92*, 2081–2085.

Justice, S.S., Lauer, S.R., Hultgren, S.J., and Hunstad, D.A. (2006). Maturation of intracellular *Escherichia coli* communities requires SurA. Infection and Immunity *74*, 4793–4800.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kaiser, P., Diard, M., Stecher, B., and Hardt, W.-D. (2012). The streptomycin mouse model for Salmonella diarrhea: functional analysis of the microbiota, the pathogen“s virulence factors, and the host”s mucosal immune response. Immunol. Rev. *245*, 56–83.

Kalas, V., Pinkner, J.S., Hannan, T.J., Hibbing, M.E., Dodson, K.W., Holehouse, A.S., Zhang, H., Tolia, N.H., Gross, M.L., Pappu, R.V., et al. (2017). Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. Sci Adv *3*, e1601944.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research *33*, 511–518.

Kau, A.L., Planer, J.D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., Manary, M.J., Liu, T.-C., Stappenbeck, T.S., Maleta, K.M., et al. (2015). Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. Science Translational Medicine *7*, 276ra24–276ra24.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and

extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research *44*, W242–W245.

Low, A.S., Holden, N., Rosser, T., Roe, A.J., Constantinidou, C., Hobman, J.L., Smith, D.G.E., Low, J.C., and Gally, D.L. (2006). Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. Environ Microbiol *8*, 1033–1047.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Mediavilla, J.R., Patrawalla, A., Chen, L., Chavda, K.D., Mathema, B., Vinnard, C., Dever, L.L., and Kreiswirth, B.N. (2016). Colistin- and Carbapenem-Resistant *Escherichia coli* Harboring mcr-1 and blaNDM-5, Causing a Complicated Urinary Tract Infection in a Patient from the United States. mBio *7*, e01191–16.

Merckel, M.C., Tanskanen, J., Edelman, S., Westerlund-Wikström, B., Korhonen, T.K., and Goldman, A. (2003). The Structural Basis of Receptor-binding by *Escherichia coli* Associated with Diarrhea and Septicemia. Journal of Molecular Biology *331*, 897–905.

Nuccio, S.-P., and Bäumler, A.J. (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol. Mol. Biol. Rev. *71*, 551–575.

Richards, V.P., Palmer, S.R., Pavinski Bitar, P.D., Qin, X., Weinstock, G.M., Highlander, S.K., Town, C.D., Burne, R.A., and Stanhope, M.J. (2014). Phylogenomics and the Dynamic Genome Evolution of the Genus Streptococcus. Genome Biology and Evolution *6*, 741–753.

Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M., Chase, J., McDonald, D., Gonzalez, A., Robbins-Pianka, A., et al. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ *2*, e545.

Rosen, D.A., Hung, C.-S., Kline, K.A., and Hultgren, S.J. (2008). Streptozocin-induced diabetic mouse model of urinary tract infection. Infection and Immunity *76*, 4290–4298.

Schreiber, H.L., Conover, M.S., Chou, W.-C., Hibbing, M.E., Manson, A.L., Dodson, K.W., Hannan, T.J., Roberts, P.L., Stapleton, A.E., Hooton, T.M., et al. (2017a). Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. Science Translational Medicine *9*.

Schreiber, H.L., IV, Conover, M.S., Chou, W.-C., Hibbing, M.E., Manson, A.L., Dodson, K.W., Hannan, T.J., Roberts, P.L., Stapleton, A.E., Hooton, T.M., et al. (2017b). Bacterial virulence phenotypes of *Escherichia coli*and host susceptibility determine risk for urinary tract infections. Science Translational Medicine *9*, eaaf1283.

Spaulding, C.N., and Hultgren, S.J. (2016). Adhesive Pili in UTI Pathogenesis and Drug Development. Pathogens *5*, 30.

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics *21*, 456–463.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

Totsika, M., Beatson, S.A., Sarkar, S., Phan, M.-D., Petty, N.K., Bachmann, N., Szubert, M., Sidjabat, H.E., Paterson, D.L., Upton, M., et al. (2011). Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. PLoS ONE *6*, e26578.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.a. *99*, 17020–17024.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli*depends on type 1 pili. Cell. Microbiol. *9*, 2230–2241.

Wurpel, D.J., Beatson, S.A., Totsika, M., Petty, N.K., and Schembri, M.A. (2013). Chaperone-usher fimbriae of *Escherichia coli*. PLoS ONE *8*, e52835.

Wurpel, D.J., Totsika, M., Allsopp, L.P., Webb, R.I., Moriel, D.G., and Schembri, M.A. (2016). Comparative proteomics of uropathogenic *Escherichia coli* during growth in human urine identify UCA-like (UCL) fimbriae as an adherence factor involved in biofilm formation and binding to uroepithelial cells. J Proteomics *131*, 177–189.

Yamamoto, S., Tsukamoto, T., Terai, A., Kurazono, H., Takeda, Y., and Yoshida, O. (1997). Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. J. Urol. *157*, 1127–1129.

Yao, J., Carter, R.A., Vuagniaux, G., Barbier, M., Rosch, J.W., and Rock, C.O. (2016). A Pathogen-Selective Antibiotic Minimizes Disturbance to the Microbiome. Antimicrob. Agents Chemother. *60*, 4264–4273.

Zhang, L., Foxman, B., and Marrs, C. (2002). Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2. Journal of Clinical Microbiology *40*, 3951–3955.

Zowawi, H.M., Harris, P.N.A., Roberts, M.J., Tambyah, P.A., Schembri, M.A., Pezzani, M.D., Williamson, D.A., and Paterson, D.L. (2015). The emerging threat of multidrug-resistant Gram-negative bacteria in urology. Nature Publishing Group *12*, 570–584.

Chapter 5:

Summary, Perspective, and Future Directions

Excerpted from:

One size doesn't fit all: unraveling the diversity of factors and interactions that drive *E. coli* urovirulence

By

Henry L. Schreiber IV, Caitlin N. Spaulding, Jonthan. Livny, Scott J. Hultgren

Annals of Translational Medicine

2017 January. 5(2):28 doi: 10.21037/atm.2016.12.73. PMID 28217693

5.1 Summary

The clinical manifestations and symptomatology of urinary tract infections (UTIs) caused by uropathogenic *Escherichia coli* (UPEC) are strikingly complex and result from the interactions between the bacterium and host urothelial tissues. The definition of UPEC itself is variable, as it is currently defined as any strain that is recovered from the urine of symptomatic UTI patients (Picard et al., 1999), a classification that fails to account for differences in host susceptibility or the possibility for multiple evolutionary and mechanistic paths to urovirulence. Further, despite widespread acceptance that the gut is a reservoir of UPEC that lead to UTI (Chen et al., 2013; Moreno et al., 2008), we know little about the host and bacterial factors that promote the establishment and maintenance of UPEC in this habitat. Towards the goal of identifying conserved bacterial features enabling urovirulence, my thesis work has integrated different animal models of bladder and gut colonization with comparative multi-omics and evolutionary analyses. This work has led us to a new way of thinking about UTI risk - a "Key and Lock" model. This model posits that the outcome of an encounter (i.e., the risk of UTI) between a specific host and potential UPEC isolate is dependent upon the combination of the particular fitness state of the UPEC isolate (the "key") matched with the particular host environment and susceptibility (the "lock"). Below, I will describe the reasoning, and the work supporting it, that has formed the basis for this model of UTI susceptibility with a focus on the UPEC "key".

First, in Chapter 2, I found that the ability to colonize the mouse bladder was a common feature amongst diverse *E. coli* and that transcriptional expression, rather than gene carriage, was a better predictor of bladder colonization in a mouse model. Here, I highlighted the genetic and phylogenetic diversity of clinical urine-associated *E. coli* (UAEC) strains isolated from women with a history of frequent rUTIs. I found that two UAEC can share as little as 60% of their genes

and that UACE can be found throughout the *E. coli* phylogenetic tree, although most clinical strains reside in clade B2, including the model UPEC strain UTI89. Using a standard mouse model of cystitis, a transurethral inoculation of juvenile C3H/HeN mice, I found that many, but not all, UAEC were robust colonizers of the mouse bladders at levels that were similar to a model, B2 UPEC strain. I further found that these robust colonizer UAEC displayed all the histopathological hallmarks of infection by model, B2 UPEC strains, such as development of intracellular bacterial communities (IBCs), quiescent intracellular reservoirs (QIRs), and the progression to chronic cystitis. Strikingly, we found that no single set of genes, including a set of commonly studied PUF genes, was capable of clearly delineating those UAEC strains that were capable of colonizing the mouse bladder robustly from those that were deficient in colonization. Instead, I found that the transcriptional expression of core genes in defined media conditions were better predictors of mouse bladder colonization than gene carriage or phylogeny. Importantly, these core genes control common bacterial behaviors, such as chemotaxis, nutrient uptake, and flagellar motility. In addition, through *in vitro* phenotype measurements, we found that function of the ubiquitous type 1 pili, part of the core genome in UAEC and critical in host colonization, also correlated with mouse bladder colonization. Taken together, these results indicate that the UPEC "key" to uropathogenesis is common to many *E. coli* strains and that its shape is influenced by the transcriptional state of the core genome in the bacterium.

Second, I found that evolution has encoded dynamic mechanical properties of the critically important type 1 pilus rod into the *E. coli* genome and that these mechanical properties are critical for UPEC colonization of the bladder and the gut. In Chapter 3, I presented an integrative study incorporating the fields structural biology, biophysics, and cellular pathology, and microbial pathogenesis that has revealed that the major subunit of the type 1 pilus rod, FimA,

is under extreme selective pressure to maintain the correct folding of the subunit to enable the pilus rod to act as a "molecular spring". The conservation of the FimA subunit was localized to the subunit-subunit interfaces within the quarternary structure of the rod, as determined through cryo-electron microscopy, and mutation in some these conserved residues resulted in a significant reduction in the force needed to unwind the pilus. Importantly, a reduction in the tension of the pilus rod through point mutations in the *fimA* gene resulted in defects in gut colonization and bladder pathogenesis in mouse models by the model UPEC strain UTI89. In addition to tension within the pilus rod, I found that residues that mediated proper assembly of the type 1 pilus, as measured by hemagglutination assays, were under similarly strong purifying selection to conserve their identity. Despite the overall conservation of the FimA protein, I did identify specific codons in *fimA* that are under strong adaptive selection, which has resulted in a diversification of amino acid residues at those positions within the subunit. These residues are located on the external surface of the type 1 pilus rod, which is exposed to the host immune system, which is indicative of antigenic diversification and an attempt to evade immune recognition. Overall, while the FimA subunit was strongly conserved, this diversity in surface-exposed subunits was enough to make the FimA protein the most divergent component of the type 1 pilus rod. Thus, overall, I found that the FimA protein is under strong purifying selection to conserve mechanical properties that are important in bladder and gut colonization while also experience episodic bouts of diversifying selection that increase the diversity of surface exposed residues in a pattern reminiscent of antigenic diversification. This shows that an parts of the UPEC "key" to uropathogenesis are under strong evolutionary conservation across all of the *E. coli* phylogeny.

Third, I found clade B2 UPEC have obtained genetic machinery from other bacteria through horizontal gene transfer that enhances gut colonization, which can seed recurrent UTIs. In Chapter 5, I describe my work in identifying the evolutionary history of the so-called F17-like pilus, which is a member of the chaperone-usher pathway (CUP) pilus family along with type 1 pilus. Here, I showed that the *ucl* operon, which encodes the F17-like pilus, is most closely related to the *uca* operon in *Proteus mirabilis* and that the *ucl* operon is significantly enriched in B2 strains of *E. coli*. I also presented evidence that F17-like pili, in addition to type 1 pili, enable gut colonization by a model B2 UPEC strain, UTI89. Importantly, the localization of the F17-like ligand, whose identity remains unknown, is located deep within the intestinal crypts and is separate from the type 1 pili ligand, which is located closer to the lumen of the gut. These data strongly suggest that the F17-like machinery were obtained by B2 *E. coli* from other bacteria to enable colonization of the gut in niches that are different than those that are occupied by other *E. coli* strains which express type 1 pili. I further investigated the role that this gut colonization may have on recurrent UTIs and found that the *ucl* was significantly enriched in UPEC isolates causing recurrent infections relative to other clade B2 strains of *E coli*. Importantly, this coincides with my findings described in Chapter 2 that clade B2 strains are significantly enriched in UPEC causing same-strain rUTI. Further, it is known that some PUF genes, especially those involved in iron uptake, are also important in colonization of the gut by *E. coli* strains (Deriu et al., 2013; Winter et al., 2013) and are also enriched in B2 strains (Picard et al., 1999; Schreiber et al., 2017). Taken together, these results indicate that evolution has selected for features that enable bladder colonization, such as structurally dynamic type 1 pili and specific transcriptional states, in genetically and phylogenetically diverse strains of *E. coli*. However, we also see that certain clades of *E. coli* are enriched in genetic factors that enable gut colonization, which, in

turn, are also enriched in UPEC strains that recurrent UTI. Thus, while gene carriage does not appear to differentiate strains that robustly colonize the bladder from those that are poor colonizers, it appears likely the gene carriage can influence UPEC pathogenicity overall by enabling persistence in a different body site. Thus, we find that the "keys" to uropathogenesis in UPEC do not have to be directly mediating bladder colonization, but, instead, could have acitivity in other body sites.

Finally, I have shown that the ability of an *E. coli* strain to colonize one host bladder does not guarantee that *E. coli* will be able to colonize all host environments, or host "locks". In Chapter 2, I have shown that some UAEC strains that are robust colonizers of C3H/HeN mice, which are generally considered to be more susceptible to UTI, are not be able to colonize C57BL/6 mice, which are generally considered more resistant to UTI. Importantly, we found that this was not a general phenomenon; most UAEC strains were able to colonize the C57BL/6 mice at similar levels. Instead, it appears that specific host-pathogen interactions resulted in different ability to colonize the mouse bladders as the UAEC strains that failed to colonize the C57BL/6 mice were genetically distinct from each other. This is an important consideration because much of what we know about UPEC pathogenicity has been based on understanding of a single bacterium's interaction with a single host environment. Given the complexity of UTI susceptibility factors resulting from environmental, behavioral, and genetic differences (Foxman et al., 2000; Foxman, 2010; Hooton et al., 1996) and the diversity in clinical UPEC (Schreiber et al., 2017), the use of a single host and a single bacterium are likely to fail to capture the variability in host-pathogen interactions seen in the clinic.

5.2 Perspective

Thus, my research has lead us to a new perspective on UPEC pathogenesis in which UTI risk is determined by diverse bacterial virulence phenotypes, variable and changing host susceptibilities, and the interactions of these phenotypes and susceptibilities in specific host-pathogen combinations. Thus, we propose a multi-dimensional and dynamic "Lock and Key" model of UTI risk in which the "Lock" of host susceptibility is matched against the "Key" of an individual *E. coli* isolate's virulence potential, which has been specially tuned by evolution (Schreiber et al., 2017). Given the multi-dimensionality of this concept, it is best visualized as landscape of interactions between hosts and UPEC (**Figure 1**). In this model, some UPEC "keys" may act as "master keys" being able to open a variety of different "locks" and some keys may be host specific, only functioning on a narrow spectrum of host "locks." Similarly, some "locks" may differentially accommodate the fit of a broad spectrum of different "keys," while some may be specific to only a few. Importantly, we must understand that the "keys" to bladder colonization are broadly distributed amongst genetically diverse *E. coli* strains and that evolution has worked to conserve the function of at least critical component in bladder colonization, the type 1 pilus rod. Further, we posit that not all UPEC "keys" are limited in their activity of colonization and persistence within the urinary tract during an acute infection. Instead, additional attention must be paid to factors that contribute to the ability of UPEC strains to persist within host reservoirs outside of the bladder, such as in the gut microbiota. By incorporating previously unaccounted for phenotypic and transcriptional variation among *E. coli* strains and their differential responses to dynamic host environments, this conceptual model helps to explain UPEC diversity in the clinic, including the lack of a clear genetic signature of uropathogenesis. However, picturing the "locks and keys" in this analogy as immutable objects that are in incapable of change oversimplifies the state of UTI risk. Instead we suggest that this is better

envisioned with both the host susceptibility "lock" and the bacterial pathogen "key" being malleable in nature and changing over time in response to differing conditions due to the dynamic nature of both host and bacterial phenotypes. This is exemplified by recent work showing that UPEC alter their transcriptional expression of attachment organelles from type 1 pili to FML pili during the transition from acute cystitis to chronic cystitis in mouse bladders (Conover et al., 2016). Notably, this change in expression coincides with inflammation in the mouse bladder epithelium that results in the expression of new ligands that are recognized by newly expressed FML pili, thus enabling the persistence of UPEC that would otherwise be removed during bladder remodeling.

Numerous clinical infections are caused by genetically diverse pathogens that infect hosts with varying levels of susceptibility. Thus, the dynamic "Lock and Key" model of UTI risk is likely translatable to other infectious diseases that lack clear bacterial genetic signatures for pathogenicity. Thus, integrated research methods that combine clinical research with experimental model systems are needed to probe the effects of: (1) host and pathogen genetic diversity, (2) host and bacterial responses to changing infection conditions, (3) infection dynamics that lead to differential susceptibilities and outcomes of bacterial infections in humans, and (4) pathogen transmission between varied reservoir and infection habitats. This research paradigm promises to yield new insight into the conserved and targetable mechanisms of virulence, critical for development of novel therapeutic strategies that are increasingly needed to face the rising tide of antibiotic resistance.

5.3 Future directions

5.3.1 Identify transcriptional profiles UPEC that enable UPEC success in distinct bladder niches

As shown in Chapter 2, genetically distinct *E. coli* capable of colonizing multiple environments of the urinary tract (e.g., the bladder, and kidney) and are capable of existing extracellularly and intracellularly during both acute and chronic phases of UTI. Importantly, I found that transcriptional regulation was a key feature in predicting *E. coli* success in bladder colonization. Thus, a better understanding of UPEC pathogenesis requires a detailed understanding of the regulation of particular genes and pathways during the dynamic cross-talk that occurs at the host pathogen interface throughout the varied niches during progression of UTI in a wide variety of UPEC that better reflect the diversity seen in the clinic. Several new technologies enable the measurement of transcriptional regulation of bacteria at unprecedented levels of sensitivity and resolution while decreasing the economic costs of multiplex library construction for high-throughput RNA sequencing (RNA-seq)(Shishkin et al., 2015). In addition, protocols developed in the Hultgren lab, enable the isolation of UPEC from their disparate niches during UTI progression. For example, it is possible to separate the extracellular UPEC bacteria away from the host bladder for easy isolation of UPEC that are in the process of developing the intracellular bacterial communities (IBCs) that are critical steps in UTI pathogenesis (Schwartz et al., 2011). Thus, while previous studies have investigated the transcriptional expression of UPEC during UTI (Conover et al., 2015; Reigstad et al., 2007), the integration of these new technologies and protocols would enable an unparalleled examination of the pathogenic processes that were inaccessible with previous technology. Importantly, given the reduced cost of these new RNA-seq and the knowledge of the pathogenic cascade, future experiments would allow for the thorough interrogation of each step of the UTI progression. Understanding the temporal dynamics of UPEC interactions with host cells would enable us to answer open questions that have existed in the field. For example, it is currently known that UPEC require

type 1 pili to invade host cells and form IBCs (Wright et al., 2007); however, it is unknown how UPEC sense that they have successfully invaded a host cell or how UPEC begin the process of escaping into the host cytoplasm to begin IBC formation. Further, while UPEC filamentation out of host urothelial cells is an important step in continuing UTI pathogenesis (Justice et al., 2006), it is unknown what signal UPEC receive or what transcriptional programs are activated to switch from growth in a biofilm-like community in the IBC to a drastic morphological change into long filaments. To answer these questions, I would collect a panel of genetically diverse UPEC strains that are all capable of colonizing the mouse bladder at high levels and then interrogating their transcriptional expression at early, middle, and late time points in the IBC development process. With this information in hand, it would be possible to identify a few genetic pathways of interest that could be further investigated using novel technologies, such as RNAscope (Wang et al., 2012), which is a tool that enables the quantitative measurement and visualization of low copy mRNA molecules at the single-cell level across diverse bacterial strains. This technology would provide a spatial understanding of UPEC transcriptional regulation within the IBC that has never before been accomplished. Taken together, these experiments would enable the dissection of spatial-temporal processes governing IBC development and UTI progression in a set of UPEC that accurately recapitulates clinical realities.

5.3.2 Define the influence of the gut microbiome on UTI susceptibility and recurrence

Over the last decade, new technological innovations have enabled detailed analysis of the gut microbiota as a critical mediator of immune status (Hooper et al., 2012) and education (Mazmanian et al., 2005), gastrointestinal tract development (Reinhardt et al., 2012), brain function (Sampson et al., 2016), and nutrition (Ley, 2005). Critically, the gut microbiota also mediates a phenotype of “colonization resistance” whereby bacterial pathogens are excluded

from the gut microbiota through occupation of host niches that are required by the pathogens for persistence (as reviewed in (Buffie and Pamer, 2013)). Here, I have described the importance of the UPEC reservoir, particularly in recurrent UTI (rUTI) and detailed features that enable UPEC to colonize the gut of mouse models. However, many questions remain about the role the gut microbiota plays in UTI, particularly in women who experience frequent rUTI. For example, we still do not know if there are any changes to the population structure of a woman's gut microbiota before, during or after UTI and antimicrobial therapy or if there are significant differences between women who suffer from frequent rUTIs and those who do not. Several lines of evidence suggest a link between the gut microbiota and UTIs as we know: (1) that population dynamics of *E. coli* the gut and bladder are linked, such that the dominant *E. coli* strain in one niche is the dominant strain in the other, (2) that inflammatory diseases, such as Crohn's disease, is associated with dysbiosis (i.e., pathogenic changes in the gut microbiota) (Gevers et al., 2014)and that UTIs are associated with increased systemic inflammation (Hannan et al., 2010; 2014; O'Brien et al., 2016), and (3) the most common treatment for UTIs are antimicrobial therapies (Hooton, 2012), which are known to disrupt the gut microbiota and enable blooms of *Enterobacteriaceae*, including *E. coli* (Spees et al., 2013). Thus, to better understand how the gut microbiota influences UTI susceptibility and recurrence, I have established a longitudinal study of a cohort of women with and without a history of frequent rUTIs that will use a combination of multi-omics technology and analysis of host immune status with mouse models of both gut and bladder colonization. Specifically, we aim to understand if population dynamics in the gut microbiota, such as blooms in *Enterobacteriaceae* or differences in colonization resistance, and/or changes in the host immune system influence the onset and severity of UTI. This project, a collaboration between the lab of Dr. Scott Hultgren, at the center for Women's Infectious

Disease Research at Washington University School of Medicine, Dr. Henry Lai's office, in the Department of Urological Surgery at Barnes Jewish Hospital, and Dr. Ashlee Earl's lab at the Broad Institute of MIT and Harvard, has recently finished the sample collection phase and has begun data generation and analysis using cutting edge technologies.

Briefly, we collected fecal/rectal swab, urine, and blood samples from women with and without frequent rUTIs over the course of one calendar year at both healthy time points, through monthly sample collection kits, as well as during and directly after any UTI episodes. At each collection, participants complete questionnaires detailing symptoms of UTI and gastrointestinal illness, as well as a diet and general health questionnaire. More than 400 fecal/rectal swab samples were collected from 14 women with frequent rUTI and an equal number of age-matched controls and we captured a total of 19 UTI events within the study window. Currently, we are in the process of measuring the gut microbiota structure (i.e., the prevalence and abundance between the groups) as well as their functional differences at both healthy and UTI time points through a combination of metagenomic and metatranscriptomic sequencing. While the majority of samples have yet to be analyzed, we performed whole genome sequencing and assembly on a set of 96 fecal samples collected from eight "cases" (i.e., women with frequent rUTI) and eight "controls" (i.e., age-matched participants without frequent rUTI). Eight fecal samples from each participant in the case group (n=64 total) were chosen to include samples before, during, and after UTI episodes, while four samples were collected from each of the participants in the case group (n=32 total). Care was taken to standardize the time between sample collection for all the samples included in this analysis, as much as possible. These metagenomic datasets were then analyzed to measure: (1) the distribution of alpha diversity (i.e., the number of species present in a single sample) as measured by the Chao1 richness indicator, in the gut microbiota between

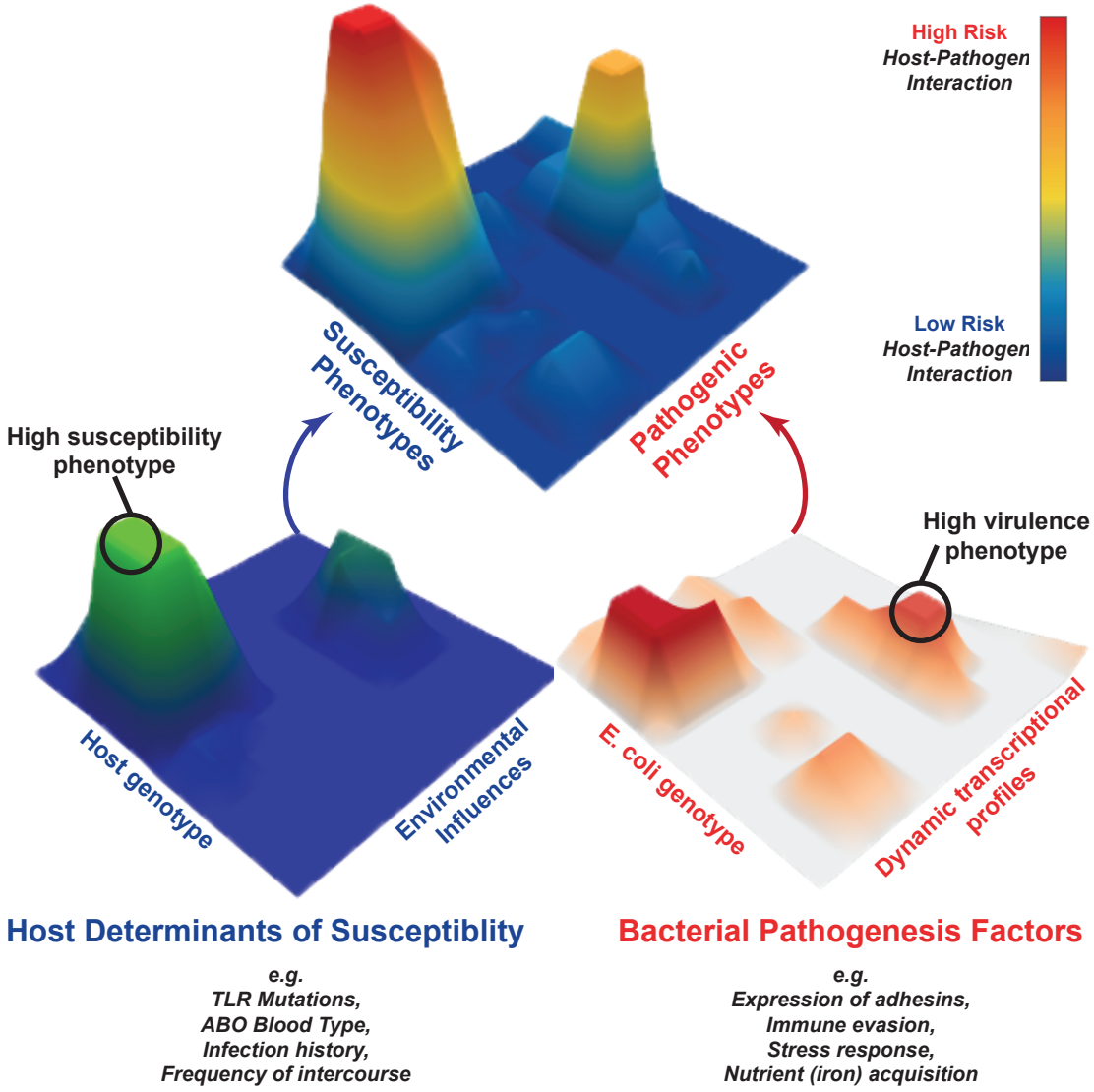
groups of participants, (2) the rates of turnover in taxa (i.e., the immigration and emigration of bacterial species) between time points in the analysis using the Jaccard diversity index, and (3) the enrichment of particular bacterial taxa in either the case or control group using the linear discriminant analysis effect size (LEfSe) tool. Comparisons between the diversity and turnover of the case and control groups was performed using ANOVA after correction for fixed effects arising from longitudinal sampling of the same participant.

Preliminary results from this whole metagenomic sequencing has revealed that, in general, the gut microbiota of women with frequent rUTI are less diverse and contain less bacteria normally associated with healthy guts than the gut microbiota of healthy women (**Figure 2**). Here, we found that control gut microbiotas were significantly more diverse, as measured by species richness, than the gut microbiota of women with frequent rUTI (**Figure 2A**). This means that, at any one time point, it is more likely that women with frequent rUTI have fewer bacterial species, and thus less genetic diversity in their gut microbiome, than healthy women. Surprisingly, we do not see a statistically significant difference in the rate of turnover between time points between the groups of participants (**Figure 2B**). In general, we saw more variability in the turnover of gut microbiota in the control group than in the case group; however, this could be due to the smaller number of samples sequenced in the control group, which resulted in longer gaps of time between sample collection and greater opportunity for natural fluctuation in the composition of the gut microbiota. Finally, we saw significant differences in the carriage of different taxa between case and control participants (**Figure 2C,D**). Importantly, we saw differences in the carriage of taxa that have been identified as biomarkers of healthy gut microbiota, namely *Akkermansia muciniphila* (Png et al., 2010) and *Faecalibacterium prausnitzii* (Sokol et al., 2008), both of which have been shown to be significantly reduced in

people suffering from inflammatory diseases, such as Crohn's disease. The fact that the case group show significantly reduced carriage of these bacteria indicate that their gut microbiota may be responding to levels of intestinal inflammation that are affecting the composition of the bacterial communities. Future analyses will continue confirm these preliminary findings and new analysis will compare the immune states and systemic inflammation of case and control groups to correlate these findings with the host response. Finally, we will use these data as a foundation for investigation into the population dynamics and colonization capability in gnotobiotic mice that have been humanized with fecal transplantations from these women. Together, these analyses will define the link between the gut microbiota, the host immune system, and subsequent UTI outcome.

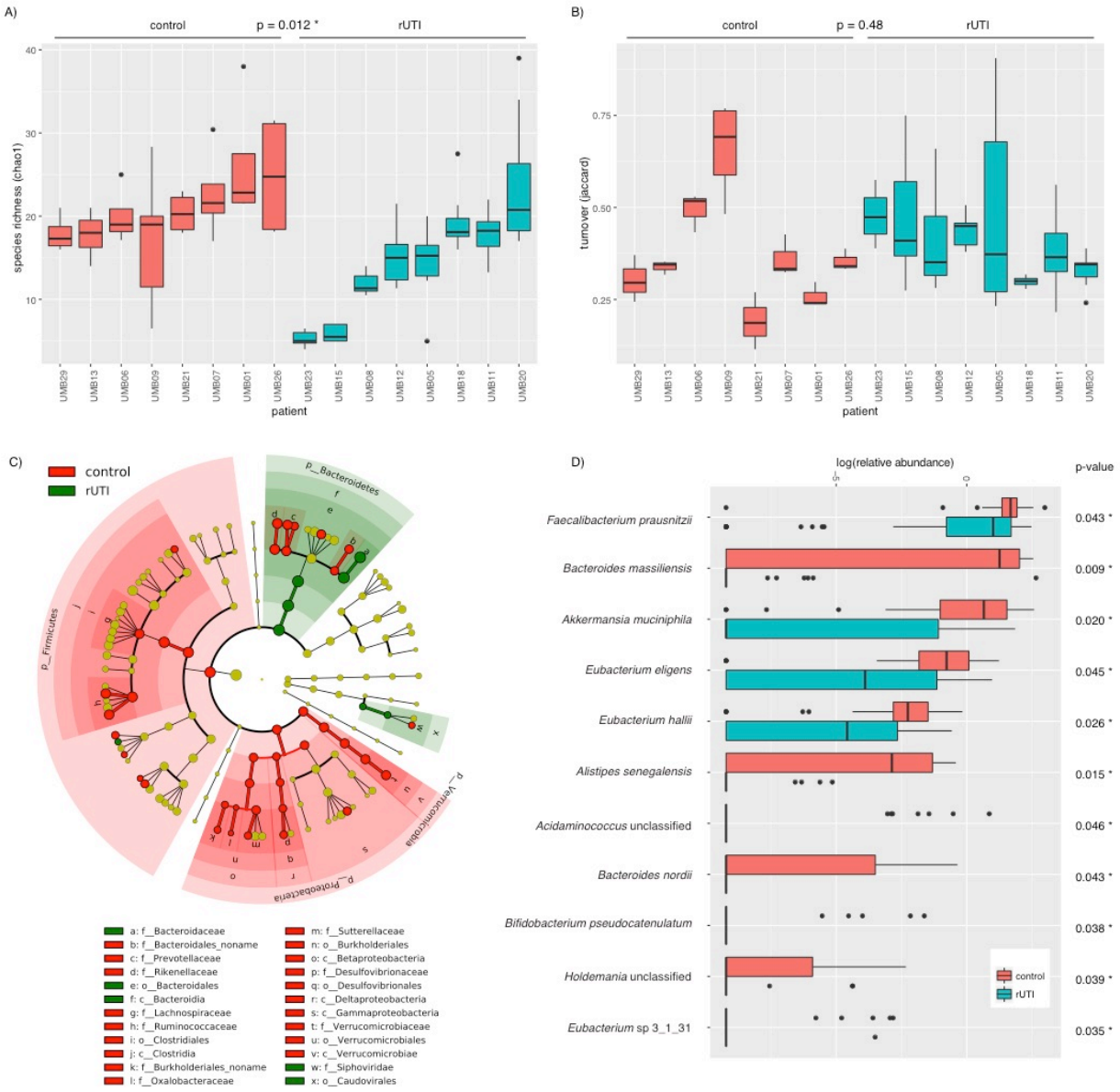
5.4 Figures

5.4.1 Figure 1. The "Key and Lock" model of UTI susceptibility.



For a given host-pathogen interaction (shown at the top), the risk of a UTI (shown in color scale from blue being low risk and red being high risk) depends on the susceptibility profile of the host (on one axis) and the pathogenicity of the UPEC (on the other axis) resulting in peaks of high risk and valleys of low risk. Previous research has shown that host susceptibility (shown on the bottom left) is mediated by the interaction of genetic and environmental influences (listed below) resulting in variable susceptibility profiles. More recent research has shown that the genotype of an *E. coli* strain alone is insufficient to predict its virulence across diverse hosts. Instead the transcriptional profile, in addition to the genotype, is better able to predict high urovirulence phenotypes (shown in the bottom right).

5.4.2 Figure 2. Patients with frequent recurrent urinary tract infections have less diverse microbiomes than control patients.



(A) Participants with recurrent urinary tract infections have significantly fewer species in their gut microbiotas than control patients using the Chao1 richness indicator. Linear mixed models were built with a patient intercept term with treatment as a variable and compared using ANOVA ($p = 0.012$). (B) Patients with rUTIs did not have significantly more turnover of microbiome species (replacement of species between sequential time points) as measured by the Jaccard diversity index. Linear mixed models were built with a patient intercept term with treatment as a variable and compared using ANOVA ($p = 0.27$). C) Control patients were enriched for many higher taxonomic groups, including a wide range of Firmicutes, Proteobacteria, and Verrucomicrobia, while patients with rUTI were enriched for Bacteroides, based on linear discriminant analysis effect size (LEfSe). D) Ten species passed the significance in both LEfSe and a linear mixed model (build as described above, p-values shown to the right of species). All ten species were enriched in control patient microbiotas, including *Akkermansia muciniphila* and *Faecalibacterium prausnitzii*, the absence of which has been associated with gastrointestinal inflammatory disorders.

5.5 References

Buffie, C.G., and Pamer, E.G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol. *13*, 790–801.

Chen, S.L., Wu, M., Henderson, J.P., Hooton, T.M., Hibbing, M.E., Hultgren, S.J., and Gordon, J.I. (2013). Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Science Translational Medicine *5*, 184ra60.

Conover, M.S., Ruer, S., Taganna, J., Kalas, V., De Greve, H., Pinkner, J.S., Dodson, K.W., Remaut, H., and Hultgren, S.J. (2016). Inflammation-Induced Adhesin-Receptor Interaction Provides a Fitness Advantage to Uropathogenic *E. coli* during Chronic Infection. Cell Host and Microbe *20*, 482–492.

Deriu, E., Liu, J.Z., Pezeshki, M., Edwards, R.A., Ochoa, R.J., Contreras, H., Libby, S.J., Fang, F.C., and Raffatellu, M. (2013). Probiotic bacteria reduce salmonella typhimurium intestinal colonization by competing for iron. Cell Host and Microbe *14*, 26–37.

Foxman, B., Gillespie, B., Koopman, J., Zhang, L., Palin, K., Tallman, P., Marsh, J.V., Spear, S., Sobel, J.D., Marty, M.J., et al. (2000). Risk factors for second urinary tract infection among college women. American Journal of Epidemiology *151*, 1194–1205.

Foxman, B. (2010). The epidemiology of urinary tract infection. Nature Publishing Group *7*, 653–660.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. Cell Host and Microbe *15*, 382–392.

Hannan, T.J., Mysorekar, I.U., Hung, C.S., Isaacson-Schmid, M.L., and Hultgren, S.J. (2010). Early Severe Inflammatory Responses to Uropathogenic *E. coli* Predispose to Chronic and Recurrent Urinary Tract Infection. PLoS Pathog *6*, e1001042.

Hannan, T.J., Roberts, P.L., Riehl, T.E., van der Post, S., Binkley, J.M., Schwartz, D.J., Miyoshi, H., Mack, M., Schwendener, R.A., Hooton, T.M., et al. (2014). Inhibition of Cyclooxygenase-2 Prevents Chronic and Recurrent Cystitis. EBioMedicine *1*, 46–57.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions Between the Microbiota and the Immune System. Science *336*, 1268–1273.

Hooton, T.M., Scholes, D., Hughes, J.P., Winter, C., Roberts, P.L., Stapleton, A.E., Stergachis, A., and Stamm, W.E. (1996). A prospective study of risk factors for symptomatic urinary tract infection in young women. N. Engl. J. Med. *335*, 468–474.

Hooton, T.M. (2012). Clinical practice. Uncomplicated urinary tract infection. N. Engl. J. Med. *366*, 1028–1037.

Justice, S.S., Hunstad, D.A., Seed, P.C., and Hultgren, S.J. (2006). Filamentation by *Escherichia coli* subverts innate defenses during urinary tract infection. Proc. Natl. Acad. Sci. U.S.a. *103*, 19884–19889.

Ley, R.E. (2005). Obesity alters gut microbial ecology. Proceedings of the National Academy of Sciences *102*, 11070–11075.

Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., and Kasper, D.L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell *122*, 107–118.

Moreno, E., Andreu, A., Pigrau, C., Kuskowski, M.A., Johnson, J.R., and Prats, G. (2008). 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.

O'Brien, V.P., Hannan, T.J., Yu, L., Livny, J., Roberson, E.D.O., Schwartz, D.J., Souza, S., Mendelsohn, C.L., Colonna, M., Lewis, A.L., et al. (2016). A mucosal imprint left by prior *Escherichia coli* bladder infection sensitizes to recurrent disease. Nat Microbiol *2*, 16196.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infection and Immunity *67*, 546–553.

Png, C.W., Lindén, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., McGuckin, M.A., and Florin, T.H.J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am. J. Gastroenterol. *105*, 2420–2428.

Reigstad, C.S., Hultgren, S.J., and Gordon, J.I. (2007). Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled. J. Biol. Chem. *282*, 21259–21267.

Reinhardt, C., Bergentall, M., Greiner, T.U., Schaffner, F., Ostergren-Lundén, G., Petersen, L.C., Ruf, W., and Bäckhed, F. (2012). Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling. Nature *483*, 627–631.

Sampson, T.R., Debelius, J.W., Thron, T., Janssen, S., Shastri, G.G., Ilhan, Z.E., Challis, C., Schretter, C.E., Rocha, S., Gradinaru, V., et al. (2016). Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. Cell *167*, 1469–1480.e12.

Schreiber, H.L., Conover, M.S., Chou, W.-C., Hibbing, M.E., Manson, A.L., Dodson, K.W., Hannan, T.J., Roberts, P.L., Stapleton, A.E., Hooton, T.M., et al. (2017). Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. Science Translational Medicine *9*.

Schwartz, D.J., Chen, S.L., Hultgren, S.J., and Seed, P.C. (2011). Population Dynamics and Niche Distribution of Uropathogenic *Escherichia coli* during Acute and Chronic Urinary Tract Infection. Infection and Immunity *79*, 4250–4259.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nature Methods *12*, 323–325.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences *105*, 16731–16736.

Spees, A.M., Wangdi, T., Lopez, C.A., Kingsbury, D.D., Xavier, M.N., Winter, S.E., Tsolis, R.M., and Bäumler, A.J. (2013). Streptomycin-induced inflammation enhances *Escherichia coli* gut colonization through nitrate respiration. mBio *4*.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. Journal of Molecular Diagnostics *14*, 22–29.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., Laughlin, R.C., Gomez, G., Wu, J., Lawhon, S.D., et al. (2013). Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. Science *339*, 708–711.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli*depends on type 1 pili. Cellular Microbiology *9*, 2230–2241.