Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection

Swaine L. Chen[1,2,3,*,+], **Meng Wu**[1,*], **Jeffrey P. Henderson**[2,4], **Thomas M. Hooton**[5], **Michael E. Hibbing**[3], **Scott J. Hultgren**[2,3], and **Jeffrey I. Gordon**[1]

[1]Center for Genome Sciences and Systems Biology, Washington University, St. Louis, MO 63110

[2]Center for Women's Infectious Disease Research, Washington University, St. Louis, MO 63110

[3]Department of Molecular Microbiology, Washington University, St. Louis, MO 63110

[4]Department of Medicine, Washington University, St. Louis, MO 63110

[5]Department of Medicine, University of Miami Miller School of Medicine, Miami, FL 33136

Abstract

Urinary tract infections (UTIs) are common in women and recurrence is a major clinical problem. Most UTIs are caused by uropathogenic *Escherichia coli* (UPEC). UPEC are generally thought to migrate from the gut to the bladder to cause UTI. UPEC strains form specialized intracellular bacterial communities (IBCs) in the bladder urothelium as part of a pathogenic mechanism to establish a foothold during acute stages of infection. Evolutionarily, such a specific adaptation to the bladder environment would be predicted to result in decreased fitness in other habitats, such as the gut. To examine this concept, we characterized 45 *E. coli* strains isolated from the feces and urine of four otherwise healthy women with recurrent UTIs. Multi-locus sequence typing revealed that two of the patients maintained a clonal population in both of these body habitats throughout their recurrent UTIs, whereas the other two manifested a wholesale shift in the dominant UPEC strain colonizing their urinary tract and gut between UTIs. These results were confirmed when we subjected 26 isolates from two patients, one representing the persistent clonal pattern and the other representing the dynamic population shift, to whole genome sequencing. *In vivo* competition studies conducted in mouse models of bladder and gut colonization, using isolates taken from one of the patients with a wholesale population shift, and a newly developed SNP-based method for quantifying strains, revealed that the strain that dominated in her last UTI episode had increased fitness in both body habitats relative to the one that dominated in the preceding episodes. Furthermore, increased fitness was correlated with differences in the strains' gene repertoires and their *in vitro* carbohydrate and amino acid utilization profiles. Thus, UPEC appear capable of persisting in both the gut and urinary tract without a fitness tradeoff. Determination of all of the

potential reservoirs for UPEC strains that cause recurrent UTI will require additional longitudinal studies of the type described in this report, with sampling of multiple body habitats during and between episodes.

Introduction

More than half of all women develop at least one episode of urinary tract infection (UTI) during their lifetimes. Up to 25% of women have recurrent UTI, which is defined as two or more episodes within a 6-month period (1). The majority of community-acquired UTIs are caused by uropathogenic *Escherichia coli* (UPEC) (2). A generally accepted model for infection is that UPEC migrate from the gastrointestinal tract to the periurethral area, and eventually up the urethra into the bladder (3).

The gut and urinary tract are very distinct habitats from the perspective of their metabolic, immunologic, and microbial features. The gut is home to our largest population of microbes (4–6), while the bladder is considered a normally sterile environment, guarded by physical and biological barriers to microbial invasion (7–9). Studies of the molecular pathogenesis of UTI in a mouse model (10–12) have identified numerous virulence factors, including adhesins, toxins, iron acquisition systems, capsular structures, flagellae, pathogenicity islands, and factors important for biofilm formation (13). Among adhesins, UPEC strains typically encode a multitude of chaperone/usher pathway (CUP) pilus gene clusters. CUP pili contain adhesins at their tips that play critical roles in host-pathogen interactions, recognizing specific receptors with stereochemical specificity (14). For example, FimH, the type 1 pilus tip adhesin, binds mannosylated glycoproteins, as well as N-linked oligosaccharides of β1- and α3- integrins that are expressed on the luminal surface of the bladder epithelium (urothelium) in humans and mice (15, 16). Type 1 pilus-mediated binding can lead to invasion of UPEC into mouse and human bladder epithelial cells (17–19). Invading UPEC can be expelled from the host cell (20) or they can 'escape' into the cell's cytoplasm where they replicate rapidly and form a biofilm-like structure, composed of 10^4–10^5 organisms, known as an intracellular bacterial community (IBC) (21, 22). Bacteria in the IBC are protected from antibiotics (23, 24) and from immune responses (11, 25). IBCs are transient; after maturation, UPEC can disperse from the IBC, exit their host cells, enter the lumen of the bladder, and subsequently invade other urothelial cells (21). One primary host defense that eliminates IBCs is exfoliation, where urothelial cells undergo an apoptotic-like cell death, detach from the underlying transitional epithelium, and are eliminated in the urine (25, 26). Exfoliated bladder epithelial cells containing IBCs have been observed in urine collected from women with recurrent UTI but not in healthy controls or in cases of UTI caused by Gram-positive pathogens (26). However, exfoliation exposes underlying cell layers of the urothelium. Subsequent UPEC invasion of these underlying cells in mice results in formation of additional intracellular structures termed quiescent intracellular reservoirs (QIRs) (27, 28). Bacteria in the QIR are dormant, are resistant to antibiotic treatment, and elude recognition by host immune defenses. Mouse models have been used to demonstrate that bacteria in QIRs can contribute to recurrent infection after antibiotic treatment has rendered the urine sterile (23, 28).

Consistent with these adaptations for colonizing the bladder habitat, UPEC have been classified as a subset of extra-intestinal pathogenic *E. coli* strains (ExPEC). ExPEC are distinguished from other gut-associated mutualistic and pathogenic *E. coli* strains based on the spectrum of diseases they cause and their genomic features. ExPEC are typically members of the B2 and D subtypes of *E. coli* and often carry pathogenicity islands (PAIs) encoding virulence-associated genes (29). Mutations in *fimH* are postulated to be an important component in the evolution of UPEC strains, with secondary contributions from mutations affecting other loci in gut-associated *E. coli* populations (30–32). In this

conceptualization, mutations conferring a fitness advantage within the urinary tract are selected in this body habitat (33). The ability of multiple UPEC strains to form specialized intracellular structures such as the IBC (34) and QIR suggests a very specific adaptation to the bladder environment. Increased fitness in the urinary tract has been hypothesized to confer decreased fitness in the gut habitat of origin, so that strains successfully colonizing the urinary tract encounter a "dead-end" evolutionary path; this has been cited as an example of "source-sink" evolutionary dynamics (31, 35).

In the present study, we have used comparative genomics, *in vitro* assays of growth in the presence of a broad range of potential nutrients, and *in vivo* fitness tests of representative *E. coli* strains obtained from both the urine and feces of women with recurrent UTIs to address several questions. How dynamic are the *E. coli* populations in the gut and urinary tract of a given individual sampled over a period when recurrent UTIs are experienced? Can genome-wide comparisons of gut and urinary tract isolates provide insights into whether recurrences arise from re-inoculation of gut-derived bacterial strains into the urinary tract or from intracellular reservoirs within the bladder? Are there fitness tradeoffs due to adaptation to either the gut or urinary tract environment? We observed two very different patterns: recurrent UTI caused repeatedly by the same strain, or rapid and apparently complete replacement of one strain with another in both body habitats between UTI episodes (arguing against a fitness tradeoff model). We were able to correlate replacement of one strain by another with their genomic and metabolic features. The significance of these results is discussed in the context of understanding disease pathogenesis and designing clinical translational studies focused on new approaches for pathogen surveillance and treatment.

Results

E. coli strains cluster by host instead of habitat

One hundred fourteen women were enrolled in a now completed study of recurrent UTI (36); each presented with symptoms of acute cystitis. Eight of the 114 individuals had a negative enrollment urine culture, while two were lost to follow-up after the initial visit. Of the 104 remaining participants, nine had three episodes of recurrent UTI (i.e., the greatest number among enrollees). However, fecal and urine samples were collected at each episode in only four of these nine individuals. All four patients received a similar cycle of antimicrobial therapy: trimethoprim-sulfamethoxazole (TMP-SMZ) for their enrollment UTI, nitrofurantoin for their second UTI, and ciprofloxacin for their third UTI (see ref. 36 and Table S1 for further details about clinical characteristics and treatment; note that patient 56 was initially treated with TMP-SMZ during episode 2 but then switched to nitrofurantoin when her urine isolates were found to be TMP-SMZ resistant). Sequencing near full-length amplicons generated from the 16S rRNA genes in the 45 strains recovered from fecal and urine samples collected during the three UTI episodes experienced by each of these four individuals confirmed that all were *E. coli*. These 45 *E.coli* strains (Table S2), which included one urine isolate and on average 3 fecal isolates from each of the four patients at the time of each episode of UTI, were selected for the current study to address our questions about source-sink dynamics (where do strains arise and how do they distribute themselves among different body habitats) and about the relative fitness of *E. coli* strains from a patient or patients that experienced a wholesale population shift between episodes.

To investigate the relatedness of the 45 *E. coli* strains and the relationships between strain characteristics and the body habitat from which they had been recovered, we first conducted multi-locus sequence typing (MLST) using seven well-conserved housekeeping genes: *adk* (adenylate kinase), *fumC* (fumerase isozyme C), *gyrB* (DNA gyrase subunit B), *icd* (isocitrate hydrogenase), *mdh* (malate dehydrogenase), *purA* (adenylosuccinate synthetase), and *recA* (recombinase A) (see Table S3 for primers used for MSLT). Two patterns emerged

from the MLST analysis: (i) a "stable clonal" pattern where isolates from the same patient were nearly indistinguishable at all time points surveyed (Patients 12 and 13); and (ii) a "dynamic" pattern where isolates from the same patient included several different MLST groups during the study (Patients 56 and 72). Even among the two patients with the dynamic pattern, strains isolated at a given clinic visit tended to have identical or very similar MLST profiles, regardless of the body site from which they had been recovered (Fig. 1). In the case of patient 72, sequence typing revealed two MLST groups among her fecal and urine isolates during episode 1. During episode 2, all her fecal and urine isolates had a single MLST group assignment that was the same as one of the groups in episode 1. In episode 3, fecal and urine isolates had a single MLST group but it differed from all the MLST groups of strains recovered during episodes 1 and 2, leading us to conclude that she harbored nearly clonal *E. coli* populations in both body habitats, but that the population had changed between episodes 2 and 3. In contrast, only one fecal isolate, from the second UTI episode from patient 13, differed in its MLST group from the group assigned to all her other fecal and urine isolates recovered at the time of UTI episodes 1, 2 and 3, leading us to conclude from the MLST analysis that she possessed a largely clonal population across episodes.

Our initial assessment of clonality was based on MLST sequencing of a limited number of strains. While no additional strains were available to extend sampling depth to search for minor *E. coli* populations, we proceeded to collect more data on the available strains to overcome potential limitations of MLST in assessing strain relatedness (37). We chose to examine the three urine and 11 fecal strains isolated from Patient 13 (stable 'clonal pattern') and the three urine and nine fecal strains recovered from Patient 72 ('dynamic pattern') as two contrasting individual examples of colonization patterns. These isolates represent each of the three episodes of UTI experienced by each of these two women (Table S4). We sequenced PCR amplicons generated from the *fimH* gene. A maximum likelihood *fimH* gene tree is presented in Fig. S1, with the position of *fimH* alleles in the various urine and fecal isolates recovered from Patients 13 and 72 shown. All three of the urine isolates from Patient 13, representing each of her three UTI episodes, had *fimH* sequences that were identical to the *fimH* sequences in all 11 of her fecal isolates. For Patient 72, this sequence identity was true for the majority of isolates from the first two UTI episodes but not for the third UTI where the urine strain had a *fimH* sequence that was different from the strains representing the two previous UTIs. However, the *fimH* sequence of this UTI episode 3-associated urine isolate was identical to the *fimH* sequences in the contemporaneously collected four fecal isolates (Fig. S1). Compared to Patient 72's first and second episode isolates, there were seven amino acid substitutions in the FimH allele of fecal and urine isolates from her third UTI episode: A10V, N70S, S78N, V119A, T234A, and A273G and N6T. The latter residue is in the signal sequence that is not part of the mature FimH protein assembled on the tips of pili. Positions 10, 70, 78, and 119 are in the lectin domain of the protein, while residues occupying positions 234 and 273 are in the pilin domain; all are solvent exposed with the exception of position 119 in a crystal structure of the FimC-FimH complex (15). None are in the mannose-binding pocket. However, residue 273 is positioned near the hydrophobic groove where donor strand complementation and donor strand exchange occur - processes essential for pilus biogenesis (14).

Unsupervised hierarchical clustering of isolates based on their gene and SNP content

To obtain more definitive data on strain relatedness and clonality, we performed whole genome shotgun sequencing of all six urine and all 20 fecal strains recovered from Patients 13 and 72 (60–140X coverage; see *Methods*). Using the Velvet (38) and AMOScmp (39) assemblers, 25/26 of the isolates yielded genome assemblies that averaged 4.98 Mbp with an average N50 contig length of 72,747 bp (Table S4).

To identify and compare the gene content of the isolates, we first compiled a database of all annotated genes from the genomes of 54 *E. coli* strains deposited in the NCBI RefSeq database that were classified as 'complete', as well as 324 other draft *E. coli* genomes present in RefSeq and the PATRIC database (Table S5). These genes, together with the genes predicted by Glimmer3 (41) in the assembled genomes of the newly sequenced strains from this study, were clustered using the program CD-HIT with default parameters (95% similarity) (42) to generate "OGUs" (Operational Gene Units). All raw reads for each of the 26 newly sequenced *E. coli* genomes were mapped to this *E. coli* pan-genome using BLAT with default parameters (43). The total number of raw reads mapping to a given OGU was then used as the score for that OGU. A null cutoff score was calculated by dividing the total number of reads by the total length of OGU representatives (as determined by CD-HIT); this cutoff represents the expected number of reads per OGU normalized by length if reads were randomly selected from all OGU representatives. OGUs with scores less than this cutoff were called 'absent'; those above were called 'present' (44). By mapping the raw reads from each isolate and the *in silico* fragmented sequences from the finished UTI89 genome onto this *E. coli* pan-genome dataset, we identified a total of 11,151 OGUs that were present in at least one of the 26 clinical isolates from the two patients or in the finished genome of UTI89; 3,488 of the 11,151 OGUs (31.3%) were conserved in all 26 strains.

We next identified a total 295,099 SNPs among the 25 isolates at positions present in the finished UTI89 genome (see *Methods*, one fecal sample Ec72_E2F2 was excluded from this analysis given the problems encountered assembling its genome; see Table S4). The SNP and OGU data were then used to more rigorously examine clonality. *First*, based on the SNP rate (SNPs/aligned bp), we computed a matrix of pairwise distance measurements between isolates from our clinical study of recurrent UTI and the 378 strains we used in the OGU analysis. Unsupervised hierarchical clustering based on these distance measurements (Fig. S2A) showed that the clinical isolates that we deemed clonally related to one another by MLST also clustered together in the SNP-based tree of 403 strains. With the exception of one fecal strain, the SNP rate between any pair of isolates from Patient 13 was in the noise range (< 100 total, SNP rates <0.005/bp) regardless of their date of isolation or whether they were recovered from urine or feces. In the case of Patient 72, we identified distinct groups of strains among the urine and fecal isolates, representing distinct branches on the tree. Fecal and urine strains isolated from her first and second UTI episodes clustered separately from those of the third UTI episode. *Second*, the OGU-based, unsupervised hierarchical clustering of the 404 strains produced the same patterns for the 26 isolates as those obtained from SNP rates (Fig. S2B). Thus, based on their SNP and OGU content, we considered UTI episodes 1–3 in Patient 13 to have all been caused by the same UPEC strain (which in turn was very similar to the strain recovered from feces and urine during UTI episode 3 in Patient 72).

Finally, we quantified the growth of each isolate from Patient 72 under 190 different culture conditions using Biolog phenotype microarrays (see *Methods*). Phenotype-based hierarchical clustering of her urine and fecal isolates yielded results that were virtually the same as those obtained from the MLST-, OGU-, and SNP-based comparisons (see Fig. 4A and below for additional information about the relationship between growth properties and gene content).

A representative urinary isolate from patient 72's third UTI has higher fitness compared to a representative urinary isolate from her first UTI both in bladder and gut

Our SNP- and OGU-based hierarchical clustering segregated the 26 isolates from the current study and other isolates, including UPEC and ExPEC strains, into two major clades at the top level of the tree. This first tree division is consistent with previous phylogenetic characterizations of various other *E. coli* strains (e.g., groups A, B1, and D in clade 1 and B2 in clade 2) and we refer to this top tree division using the terms clade 1 and clade 2 for

convenience (Figs. 2, 3). Comparison of the content of known virulence factors in the 26 sequenced isolates from our study and the 54 complete *E. coli* genomes from RefSeq revealed that strains located in clade 1 had significantly fewer virulence genes compared to strains in clade 2 (see Table S6 for p-values, Chi-square tests).

We selected urine isolate Ec72_E1U1 obtained from Patient 72 during her first UTI episode and located in clade 1 as a proxy for fecal and urine samples from her first two UTI episodes, and urine isolate Ec72_E3U1 recovered from her last UTI episode and located in clade 2 as a proxy for all urine and fecal isolates from this episode and a very close strain to all urinary and fecal isolates from all three UTI episodes in Patient 13. (Figs. 2 and 3). Since her urine strain from episode 1 was replaced by the urine strain in episode 3, we asked whether the episode 3 strain had higher relative fitness compared to the episode 1 strain in both the bladder and gut. If this were true, it would provide a counter-example to the notion that there is a fitness tradeoff between the urinary tract and gut, and would suggest that the urinary tract is not necessarily a "sink" or evolutionary "dead-end" habitat for UPEC strains.

To address this question, we first turned to a well-established mouse model of UTI in conventionally-raised C3H/HeN mice (45). The UTI episode 1 strain was marked with low copy plasmids containing genes conveying resistance to kanamycin (pACYC177; Kan^R) or chloramphenicol (pACYC184; $Chlor^R$). The UTI episode 3 strain was marked with pACYC184 ($Chlor^R$). The p15A origin driving replication and segregation of these plasmids confers stable inheritance in *E. coli* and many other *Enterobacteriaceae* (46). *In vitro* competition experiments revealed that these plasmids were indeed stable in both strains, and that there was no growth (fitness) defect between the marked strains and their unmarked counterpart, as judged by quantifying colony forming units (CFUs) over the course of 8 h of growth under shaking conditions or during 48 h of growth under static conditions in LB medium. Furthermore, there was no significant difference in growth of the episode 1 strain (Ec72_E1U1) when it contained pACYC177 (Kan^R) versus pACYC184 ($Chlor^R$).

We subsequently compared the fitness of Ec72_E1U1 with Ec72_E3U1 by co-inoculating a 1:1 mixture of the Ec72_E1U1/pACYC177 (Kan^R) and Ec72_E3U1/pACYC184 ($Chlor^R$) strains into the bladders of female C3H/HeN mice ($1–3 \times 10^7$ CFU of each strain; n=5 mice). Mice were sacrificed 24h after inoculation, and the number of bladder CFU of each strain was defined by plating bladder homogenates on selective media. The marked episode 3 strain (Ec72_E3U1/pACYC184) was the only strain detectable in bladders 24h post inoculation, with the exception of one mouse where the episode 1 strain was present at a low level (10^3 CFU; Fig. 5A). In follow-up single strain infections using unmarked strains without any antibiotic resistance plasmids, the episode 1 strain was undetectable in the bladder tissue of mice 24h post-inoculation while the episode 3 strain achieved a median colonization density of 5.2×10^4 CFU/bladder (range $4.4 \times 10^2 – 1.12 \times 10^8$) (Fig. 5A). Confocal microscopy of bladder whole mounts, prepared 6, 12, 16, and 24 h after mice were mono-infected with the same strains containing pANT4 (encoding GFP; 47) revealed that the episode 3 but not the episode 1 strain was able to form small IBCs (Fig. S3), consistent with other reports that intracellular infection contributes to fitness during UTI. Additionally, we analyzed urine samples as well as bladder and kidney homogenates prepared from mice 24h after transurethral inoculation of ten times more CFUs (10^8) of one or the other of these strains (or the reference control UPEC strain UTI89). The results revealed barely detectable levels of the episode 1 strain, Ec72_E1U1, in bladder homogenates, although it was present in kidney and urine. In contrast, the episode 3 strain (and the control UTI89 isolate), was present in all three sample types and at significantly higher levels than the episode 1 strain ($p<0.05$; 2-way ANOVA and Mann Whitney U test; Fig. S4).

To test the stability of these plasmids during a much longer period of colonization in the gut, we introduced a 1:1 mixture of the episode 1 strain marked with pACYC177 (Kan[R]) or pACYC184 (Chlor[R]) into adult germ-free male C57Bl/6J mice using a single oral gavage (n=5 mice). Fecal samples were collected daily during the first four days after gavage, and then every 2 days for 2 weeks. Samples were plated on LB agar without and with antibiotics. Total fecal levels of *E. coli* ranged from $0.9–1.3 \times 10^7$ CFU/mg (wet weight) throughout the experiment (non-selective medium). However, levels of the Chlor[R] marked strain fell throughout the experiment, while levels of the Kan[R] marked strain remained constant (Fig. 5B), indicating that in contrast to short term (24h) colonization of the bladder, pACYC184 (Chlor[R]) conferred a fitness disadvantage, or that this plasmid was being lost from the episode 1 strain during the 2 week period of monitoring gut colonization.

To circumvent the problem of having to mark strains with plasmids, we developed a method we named FitSeq that differentiates sequenced strains based on their SNP content and provides a digital output of their abundance (see *Methods*). To validate FitSeq, we began with an *in silico* simulation using reads from the whole genome sequencing datasets obtained from strains Ec72_E1U1 and Ec72_E3U1. Mixtures of reads were created with the fractional representation of Ec72_E1U1 set at 0.4, 0.5 and 0.6, and the observed ratios of the two strains calculated based on SNP content over a 5 order of magnitude range of input reads (10–1,000,000/strain). Fig. S5A demonstrates that 100,000 reads are more than sufficient to determine the ratio of the two strains. Next, the two strains were each grown in monoculture, genomic DNA was extracted, and the two purified DNAs mixed in a manner such that the fractional representation of strain Ec72_E1U1 was systemically varied from 0 to 1 in 0.05 increments. An Illumina sequencer was used to generate 36 nt reads from these defined mixtures. Employing 500,000 reads per sample, *in silico* simulations of the type described above and direct analysis of the defined mixtures showed excellent correlation between expected and detected representation (R^2=0.999, Fig. S5B).

With these results in hand, we gavaged germ-free, adult male and female C57BL/6J mice (n=5/group) with a 1:1 mixture of the two urine strains recovered from Patient 72 during her UTI episodes 1 and 3. Fecal samples were collected as described above. FitSeq disclosed that the strain from episode 1 was rapidly outcompeted by the episode 3 strain in the guts of both male and female animals (Fig. 5C), similar to the temporal profile seen in the original human patient.

Mechanisms that could underlie the increased fitness of episode 3 strain from patient 72 in both the bladder and gut environments

Our assembly of Ec72_E1U1 and Ec72_E3U1 indicated that they share 4,714 OGUs and have 1,432 and 1,969 unique OGUs, respectively. To better understand the differences in fitness of the episode 1 and 3 strains, we undertook a more in depth genomic and phenotypic analysis. We generated a more complete assembly of their genomes after re-sequencing [150 nt x 2 (paired end) Illumina MiSeq reads; 39–42 fold coverage of each genome; N50 contig length, 108,524 bp (Ec72_E1U1) and 126,534 bp (Ec72_E3U1); Table S4]. There was a high correlation between gene coverage with the initial short read assembly and gene coverage with the longer MiSeq reads (r^2=0.99). In addition, BLAST searches of the new genome assemblies confirmed the presence or absence of OGUs (as defined from the earlier analysis), at both the nucleotide and predicted protein levels.

The episode 3 strain, Ec72_E3U1, contains the complete *fim* operon encoding type 1 pili. While the episode 1 strain, Ec72_E1U1, has a full *fimH* gene, it is missing the other structural genes required for assembling a functional type 1 pilus. Indeed, under laboratory growth conditions, we were unable to induce expression of functional type 1 pili in Ec72_E1U1 as measured by hemagglutination of guinea pig red blood cells. Consequently,

this strain was unable to form type 1 pilus-dependent biofilms after growth in LB broth in polyvinylchloride wells. The episode 1 strain was also deficient in forming pellicle biofilms during growth in YESCA broth (note that pellicle biofilm formation is not dependent on type 1 pili; 48). In contrast, the episode 3 strain was similar to the prototypic human cystitis isolate UTI89 in assays for type 1 pilus expression and function (Fig. S6A,B), but unlike UTI89, it was not capable of pellicle biofilm formation (Table S7).

The episode 1 strain (Ec72_E1U1) was significantly depleted in genes involved in flagellar assembly function (p<0.05, χ^2 test). Ten core flagellar assembly genes are present in the episode 3 strain and absent in the episode 1 strain: they include genes essential for formation of the MS ring (*fliF*), the C ring (*fliG, fliM* and *fliN*), and the export apparatus (*fliH, fliI, fliO, fliQ, fliP* and *fliR*) (Fig. S7). The lack of these essential components of the basal body would severely impact the ability of the flagellum to rotate, thus affecting motility. Indeed, the episode 1 strain was non-motile, while the episode 3 strain and UTI89 were motile as measured in swimming and swarming assays (Fig. S6C, Table S7). The other four strains recovered from the urine and feces of patient 72 that clustered together with Ec72_E1U1 in the MLST-, OGU- and SNP-based trees (Ec72_E1F1, Ec72_E2U1, Ec72_E2F1, Ec72_E2F2 in Figs. 1,2,3, and **S2**) also lacked these ten core flagellar assembly genes (Table S8A).

The episode 3 strain possessed most of the canonical UPEC virulence-associated pathogenicity island (PAI) elements (PAI-II, PAI-III, PAI-IV), 8 chaperone-usher pilus systems, and several additional toxins and iron acquisition systems (α-hemolysin and four major siderophore systems). In contrast, the episode 1 strain was missing most of these PAIs, had only one intact chaperone-usher pilus system, and lacked all of the siderophore systems and toxins associated with UPEC (Table S6 and Fig. S8). *In vitro* assays confirmed the absence of α-hemolysin activity in this strain (see Ec72_E1U1 in Table S6 and Fig. S8).

The episode 3 strain was also enriched in phosphotransferase systems (PTS) relative to the episode 1 strain (p<0.05, χ^2 test). The genes encoding PTS-Sor-EIIA, PTS-Sor-EIIB, PTS-Sor-EIIC, PTS-Sor-EIID, which comprise L-sorbose-specific enzymes II in the phosphenolpyruvate (PEP)-dependent PTS system (Fig. S9), were absent in the less fit episode 1 and present in the episode 3 strain. Three components are required in the PTS system: the two common PTS proteins, enzyme I (EI) and HPr, which transfer a phosphoryl group from PEP to the substrate-specific Enzymes II (EII) complex. Enzymes II are involved in the first step of sorbose utilization, transport of L-sorbose into the cell and phosphorylation to L-sorbose-1-phosphate. The absence of this particular sorbose PTS EII suggested that the episode 1 strain cannot use L-sorbose, a fact confirmed in the phenotypic microarray assay (Fig. 4B). L-Sorbose derived from dietary vegetables exists in both the human gut and urinary tract (49). L-Sorbose utilization is a distinctive feature of virulent *E. coli*, including ETEC, EIEC, STEC, EPEC, and other UPEC strains (50). The lack of the L-sorbose PTS-system was also observed in (i) the genomes of the four other strains judged to be clonal with Ec72_E1U1 based on MLST-, OGU-, SNP analyses (Ec72_E1F1, Ec72_E2U1, Ec72_E2F1, Ec72_E2F2) and (ii) the 13 reference strains in phylogenetic group A that clustered with this clonal population in the tree shown in Figs. 2 and 3 (Table S8B).

The differential representation of other genes in the genomes of these two strains suggest a genetic basis for the observed differences in their *in vitro* growth phenotypes (Fig. 4B) and fitness in the gut and bladder (Fig. 5A,C). For example, genes involved in galactose utilization were more prominent in the episode 3 strain and correlated with its higher growth rates on substrates requiring galactose metabolism (Fig. 4B). Defects in galactose utilization are known to affect colonization of the intestine. In fact, *E. coli* uses multiple sugars for growth in the intestine and multiple mutations affecting different sugar utilization pathways

have an additive effect on the colonization levels of the enterohemorrhagic *E. coli* strain EDL933 in CD-1 mice (51). Peptides or amino acids are the primary carbon source for *E. coli* during UTI (52), and peptide transport, gluconeogenesis and the TCA cycle are required for UTI caused by the UPEC strain CFT073 (49,52). The phenotype microarray analysis disclosed that the more fit episode 3 strain had higher growth rates on all four dipeptides and 10 of 22 amino acids tested (Fig. 4B). Nine of these ten amino acids are glucogenic in the TCA cycle. Genes involved in gluconeogenesis and TCA cycle were enriched in the variable component of the episode 3 isolate's genome, compared to the episode 1 isolate's genome (Table S9).

Discussion

We have conducted a study analyzing the genomic features of *E. coli* strains isolated from the feces and urine of four women during recurrent bouts of UTI and assessed the relative fitness of representative strains in mouse models. We found two very different colonization patterns, each represented by two patients, with respect to the dominant *E. coli* population in the gut and bladder. One pattern, exemplified by Patient 13, can have stable and seemingly clonal *E. coli* populations in the gut and urinary tract for several months over the course of multiple recurrent urinary tract infections. In contrast, the other pattern, illustrated by Patient 72, can be very dynamic with a wholesale shift in the major population colonizing both the intestinal and urinary tract occurring over the 1 month period between her second and third episodes of UTI. Unsupervised hierarchical clustering of both genetic and phenotypic data (MLST, whole genome gene and SNP content, and *in vitro* growth on 190 substrates) supported the clonal relationship of strains representing these dominant *E. coli* populations.

We tested the *in vivo* fitness of two isolates: (1) Ec72_E1U1, a representative of the strains present in Patient 72's gut and bladder during her first two episodes, and (2) Ec72_E3U1, a representative of the gut and urine isolates from the last UTI episode of Patient 72. The results revealed that the latter strain had higher fitness in both the mouse bladder and gut, consistent with the population shift documented in both the gut and bladder of patient 72. These fitness differences correlate with a number of genomic and metabolic features that provide insights about the requirements for survival in these body habitats. Nonpathogenic *E. coli* strains generally contain fewer chaperone-usher pilus systems than pathogenic strains (53). We found that the less competitive episode 1 strain carries fewer pilus systems (only one of the 13 known CUP systems in *E. coli*) than the episode 3 strain. Moreover, there were seven predicted amino acid differences between the FimH proteins encoded by the two isolates. FimH residues 70 and 78 define two major groups of FimH sequences (54, 55); the first UTI episode isolate (Ec72_E1U1) has 70N and 78S, which are associated with fecal and non-UPEC strains, while the third UTI isolate (Ec72_E3U1) has 70S and 78N, which are associated with UPEC strains.

Flagella are thought to be important for UTI pathogenesis. The flagellum consists of a basal body, hook and filament. Flagellar synthesis is a highly ordered and regulated process involving three classes of genes. Class I genes include *flhDC*, which encode the FlhD/FlhC complex that functions as a transcriptional activator of flagellar class II operons. Class II genes encode the basal body and hook, as well as FliA and FlgM, which are the sigma factor and anti-sigma factor that regulate transcription of class III genes. Class III genes encode the hook-associated proteins and the filament of the flagellum (FliC), as well as proteins necessary for motility (e.g., MotA, MotB) (56). Studies of isogenic wild-type and Δ*fliC* strains have shown that loss of the flagellar protein FliC results in reduced persistence in the urinary tract of mice, while IBC formation and dispersal are not affected (57). *E. coli* strains are classically grouped by serotyping based on their LPS O-antigens and flagellar H-antigens. Nearly all *E. coli* have an H-typeable flagellar antigen, including nonmotile strains.

H-typeable but nonmotile strains include sorbitol-fermenting O157 strains isolated from hemolytic uremic syndrome in which there is a 12bp deletion in *flhC* (58). The urine isolate and one fecal isolate from the first two UTI episodes in patient 72, which would be typed as an H30 strain based on sequence identity to *fliC* from strain HW32 (59), have deletions that eliminate a subset of flagellar class II genes. All these strains, which include the episode 1 strain, Ec72_E1U1, have an intact *flhDC* operon. Thus, deletion of genes encoding flagellar structural protein represents another (alternative) route to disruption of *E. coli* motility.

No common set of virulence determinants has been identified that is specific to UPEC strains and absent from *E. coli* strains that have a mutualistic relationship with their host (60, 61). The general lack of known virulence determinants in the episode 1 strain from Patient 72 raises the question of how this strain was able to cause a symptomatic UTI. The episode 1 strain, Ec72_E1U1, and the closely related episode 2 strain, Ec72_E2U1, were isolated from the urine of patient 72. They were cultured from midstream urine samples using previously well-defined and validated protocols (36) that make fecal contamination highly unlikely. In a recent study, 80% and 40% of urine isolates collected from women with symptomatic UTI were capable of expressing functional type 1 and P pili, respectively, after *in vitro* growth (60). In addition, studies of isolates from women with asymptomatic bacteriuria have found that they are enriched for UPEC strains that have lost the ability to make functional pili (62). Thus, alternate mechanisms for *E. coli* colonization of the urinary tract exist. The association of UTI symptoms with the episode 1 strain is even more puzzling given its limited number of chaperone-usher pili, its *fimH* sequence, its lack of PAIs, and its clustering with other non-pathogenic strains based on comparison of their sequenced genomes. Sexual intercourse has been shown to introduce bacteria into the female bladder (3). Patient 72 engaged in sexual intercourse just prior to bacteriuria and the development of symptoms (36). Therefore, it seems reasonable to propose that Ec72_E1U1, despite lacking functional P and type 1 pili, was inoculated into her bladder in sufficient quantities to maintain itself for a period of time to cause symptoms. Indeed, we found that an inoculation of 10^8 CFU of Ec72_E1UI into the bladders of mice was sufficient to maintain bacteriuria and kidney colonization even in the absence of significant invasion of the bladder urothelium (Fig. S4). Furthermore, studies in mouse models have shown that certain UPEC strains that are defective in IBC formation can be complemented to form IBCs when there is co-infection with other UPEC (34); thus, in the context of a mixed infection, this strain may be able to persist within the bladder. In contrast, Ec72_E3U1 colonized and invaded the bladder. Our finding that the episode 3 strain Ec72_E3U1 has intact coding sequences for functional type 1 pili and flagella and a greater flexibility in utilizing carbon sources available in the gut and urinary tract emphasizes the multigenic underpinnings of virulence, provides mechanistic understanding for its observed displacement of the less fit Ec72_E1U1 between Patient 72's UTI episodes 1 and 3, and underscores the need to examine the role of metabolic capabilities in determining the fitness of UPEC in two body habitats involved in disease pathogenesis. For example, we have shown that mutants with deletions in *sdhB* (required for conversion of succinate to fumarate in the TCA cycle) or *mdh* (catalyzes metabolism of malate to oxaloacetate; loss of *mdh* blocks the TCA cycle and glyoxylate shunt) are both attenuated in a mouse model of UTI, correlating with a decreased ability to form IBCs (63).

The observation of the same strain of *E. coli* in both urine and fecal isolates during a given UTI episode and across successive UTI episodes is somewhat surprising given the hypothesis that the urinary tract is an evolutionary dead end from which *E. coli* do not emerge to seed other habitats (31). Two possibilities that are not mutually exclusive could explain our results: (i) UPEC are very fit in both the gut and urinary tract, and UTI results from gut to bladder inoculation but the reverse never happens; (ii) UPEC transiently occupy, and in fact dominate, the gut *E. coli* population *en route* to the urinary tract, and because we

have sampled strains only during UTI episodes, we observe the same strain in both the feces and urine. These two possibilities differ in that (i) assumes high UPEC fitness in the gut, while (ii) does not. Given our gut colonization data in gnotobiotic mice, it seems that (i) is the more likely of these two possibilities. If the urinary tract is not an evolutionary dead end, then a third possibility needs to be considered: the urinary tract may be a source for gut colonization with UPEC. The third possibility is consistent with a scenario where there is dynamic fluxing between both body habitats so that a strain originating from the gut causes a UTI episode and UPEC from the bladder/urine subsequently re-seed the gut; this 'cycle' could lead to the 'homogeneity' that we see in our study where the one strain dominates in both habitats during a given UTI.

Overall, this is a complex issue, as gut colonization by UPEC can be consistent with the urinary tract as a dead-end; i.e., if UTI always arises from infection with gut bacteria, then human-human transmission and epidemics of UTI could be caused merely by fecal-oral transmission of bacteria from gut to gut without requiring intervening occupancy of the urinary tract. Two observations argue against this notion. *First*, in the case of Patient 72, we find that the relative fitness of gut and urinary isolates in mice follows their dynamics in humans, and that isolates with higher relative fitness co-exist simultaneously in both host habitats. *Second*, in an analysis of heterosexual couples, colonization of the gut of both partners by the same strain of *E. coli* was associated with cunnilingus (64). One explanation for this latter result involves transmission from the urinary tract of the female to the gut of the male. Foodborne transmission of extra-intestinal *E. coli* may also represent a possible route of dissemination (65, 66).

The concept in evolutionary theory that specialization to one environment is generally detrimental to fitness in another has been extensively explored in the context of microbial survival in the presence and absence of antibiotics. Fitness costs due to adaptation to antibiotic pressure are seen. However, compensatory mutations that restore fitness and no-cost adaptive mutations have been identified in numerous systems (67). Interestingly, in *C. jejuni*, fitness tradeoffs are very clear in the development of resistance to macrolide antibiotics, yet evolution of resistance to fluoroquinolones confers equal or higher overall fitness in the absence of antibiotic pressure in animal models of infection, demonstrating that fitness landscapes can be dynamic and complex (68, 69). Therefore, while some UPEC strains may suffer in their gut fitness, our data from Patient 72 indicate that a pathway to high fitness in both urinary tract and gut exists. This is interesting in light of the highly specialized intracellular infection pathway used by many UPEC in the bladder, the efficiency of which varies from strain to strain (21, 22, 34). IBC formation is multifactorial, and there is evidence that this is a selected process among UPEC strains (33). Type 1 pili mediate binding to uroplakins (16, 25) but play an additional intracellular role in IBC formation (33, 70) and thus are key features of both UTI and IBC formation. Selection for this specialization towards uroepithelial cells is expected to decrease fitness in other habitats such as the gut, especially in a source-sink model where the urinary tract is an evolutionary dead end. We have now demonstrated that fitness in the urinary tract and gut are not necessarily inversely related (e.g., the episode 3 but not the episode 1 strain from patient 72 forms IBCs). It may be useful to consider how selection for the biofilm-like IBC in uroepithelial cells could enhance survival in the gastrointestinal tract, which is lined with a mucus layer containing polysaccharides that can serve as a nutrient repository as well as a place for attachment and establishment of syntrophic relationships with other members of the microbiota. In other words, the IBC may be a focal point for patho-adaptive changes that also increase fitness of strains in the gut.

Our findings provide a rationale for additional studies of populations of women representing different ages, genotypes, and lifestyles (including different diets and nutritional status) in

order to address the question of the origins of recurrent UTI and the relative importance of dynamic and stable patterns of colonization. This is especially important since only four of 104 patients met our inclusion criteria (three recurrent UTIs with *E. coli* isolates available from both feces and urine sampled at the same time during each episode), and since in-depth genomic and phenotypic characterization of isolates was performed for just two of these patients. If UPEC are able to move freely between the gut and urinary tract, this complicates our inference of the ultimate reservoir for recurrent UTI. Therefore, further studies of the type described in this report are needed to determine the migration directions of these bacteria between different sites within an individual and between individuals. We envision these studies as a part of a translational medicine pipeline directed at developing more informed concepts about the pathogenesis of recurrent UTI, as well as more effective therapies. For example, whole genome sequencing of isolates obtained from time series studies of patients with recurrent UTI should help determine whether the majority of *E. coli* isolates obtained from the gut and urinary tract of an individual at a given time point are clonally related, and whether or not there are barriers to homogenization across different body habitats (oral, fecal/perianal, vaginal, periurethral and bladder). FitSeq and the type of animal models employed here can then be used to compare the fitness of UPEC strains that sweep to dominance in the gut and urinary tract of a human host in the setting of UTI. The results may influence standards of care in the future: e.g., whether long-term surveillance of the fecal microbiota coupled with sampling microbial communities from other body habitats can identify population shifts in patients with histories of UTI prior to the onset of UTI. This surveillance may also help define new approaches to chemoprophylaxis, either involving existing antibiotics or next generation compounds that target UPEC through novel mechanisms, such as mannosides that impede or block FimH-mediated binding of UPEC strains to mannosylated epithelial surface receptors (71).

Materials and Methods

Strains

The TOP trial was conducted using protocols approved by the Human Studies Committee of the University of Washington. Exclusion criteria included known anatomic or functional abnormalities of the urinary tract, symptoms or signs of acute pyelonephritis, chronic illness requiring medical supervision, pregnancy or planned pregnancy in the three month period following enrollment. Clinical information and strains were collected during the trial as described in Czaja et al (36).

Amplicon sequencing

Genomic DNA was prepared from each *E. coli* isolate using the Promega Wizard Genomic DNA kit (Promega, Madison, WI) according to the manufacturer's instructions. Near full length ampilcons from the 16S rRNA gene were generated by PCR and sequenced using the dideoxy chain termination method as described (72). Gene targets for MLST, including the *fimH* allele, were amplified with the primers listed in Table S3 in the following reaction mixture: 1x PCR buffer (Invitrogen, Grand Island, NY) supplemented with 2.5mM $MgCl_2$, 1.4M betaine, 1.3% DMSO, and 200μM dNTPs, 5ng of template, 12.5pmol of each primer, and 1 unit of Taq polymerase (Invitrogen, Grand Island, NY) in 50 μL total volume. Reactions were heated to 95°C in a thermocycler for 5 min, then cycled 35 times using the following conditions: 95°C for 1 min, 55°C for 1 min, and 72°C for 1–3 min (depending on the expected product size). Reactions were finished with a 10 min incubation at 72°C, cooled to room temperature, and subsequently purified using the Qiagen QiaQuick PCR purification kit according to manufacturer's instructions. Amplicons were sequenced using standard dye-terminator capillary sequencing. Base calling and assembly of multiple reads were performed using the programs Phred, Phrap, and Consed with default parameters.

Biolog phenotype microarrays

Phenotype microarrays were run according to the manufacturer's standard protocols. Briefly, a strain was grown on solid LB agar with no antibiotic selection at 37°C overnight. Cells were scraped from this plate, resuspended in 10 mL of buffer IF 0a GN/GP Base IF (Biolog), normalized to a transmittance of 85% using the Biolog Turbidimeter, diluted 100-fold in buffer IF 0a GN/GP Base IF with Biolog Dye mix, and then pipetted into Biolog PM1 or PM2 plates. Plates were subsequently incubated at 37°C in the Biolog machine for 48 h with colorimetric measurements made every 30 min. Data were exported from the Biolog software as total AUC (area under the curve) for the 48 h assay, giving 96 AUC values for each PM plate. These values were subjected to unsupervised hierarchical clustering to determine the relatedness of the phenotype microarray profiles for each strain.

Whole genome sequencing and assembly

Genomic DNA libraries for the Illumina GA-II sequencer were prepared according to the manufacturer's protocol. Each strain was sequenced (36 nt reads) using 2 lanes of the 8-lane flow cell. The output files were converted to FASTA format, ignoring quality scores, and assembled with the Velvet short read assembler (version 0.7) using custom Perl scripts to optimize k-mer length and minimum coverage parameters for both N50 length and total assembly length. AMOScmp (version 2.0.5) was used to further improve the assembly with default parameters.

Paired-end libraries with 500 bp inserts were prepared for strain Ec72_E1U1 and Ec72_E3U1 as described by the manufacturer of the Illumina MiSeq instrument. Sample-specific, 8 nt Hamming barcodes 73) were incorporated into the sequencing adapter for multiplex sequencing [8 barcodes per strain; two strains sequenced in a single MiSeq flowcell; 16 barcodes per MiSeq flow cell to get a balanced base composition during the first four cycles of the sequencing run (critical for accurate cluster calling as well as good phase and pre-phasing values)]. The output FASTQ files were assigned to each strain using the barcodes (0.68 million reads for Ec72_E1U1 and 0.78 million reads for Ec72_E3U1), and then assembled with MIRA (version 3.4.0) using default parameters (74).

SNP analysis

Analysis of assembled genomes for SNPs was done as described (44). UTI89 was used as the reference finished UPEC genome. All genomes were aligned against UTI89 using BLASTN with default parameters. Only the alignment with highest p-value reported by BLASTN was used for each assembled contig. The position of each SNP, based on UTI89 genome sequence, was recorded. SNP rates were calculated for each pair of assembled genomes by counting the number of sequence differences only at positions in the UTI89 reference genome where both genomes had contigs that aligned. The total number of sequence differences in overlapping regions of the genomes was divided by the total length of overlapping regions, yielding SNP rate per aligned base pair.

Competitive tests of fitness

All experiments involving mice were performed using protocols approved by the Washington University Animal Studies Committee. *In vivo* fitness tests in the urinary tract were performed as previously described (45). Briefly, bacteria were grown under type 1 pili-inducing conditions (two passages at 37°C in LB broth without shaking for 16–20 h, with a 1:1000 dilution between passages). These static cultures were briefly agitated to resuspend settled bacteria. Cells were collected by centrifugation (~3,000 ×g for 10 min at 4°C) and resuspended in PBS to an OD_{600} of 1.0. An equal mixture of two strains was inoculated transurethrally into the bladders of 7–8 week old female C3H/HeN mice. Twenty four hours

later, mice were sacrificed and their bladders were removed aseptically, placed in 1 mL of PBS, mechanically homogenized with a stainless steel electric tissue homogenizer (PRO Scientific, Oxford, CT) for 15–20 sec, and plated on LB/agar plates with and without antibiotics [kanamycin (50 μg/mL), chloramphenicol (100 μg/mL), or kanamycin plus chloramphenicol]. CFU counts were defined after a 12–18 h incubation at 37°C under aerobic conditions. Single infections were performed similarly, except that bacterial suspensions were mixed 1:1 with PBS, to achieve a final OD_{600} of 0.5 before inoculation.

FitSeq

To assess their relative fitness in the gut, *E. coli* strains Ec72_E1U1 and Ec72_E3U1 were grown under type 1 pili-inducing conditions. The two strains were mixed in equal proportions ($2–3 \times 10^8$ CFU/200μL/strain) and inoculated by oral gavage into germ-free 8–10 week old C57BL/6J mice. Mice were maintained in plastic flexible film gnotobiotic isolators and fed a standard autoclaved chow diet (B&K Universal, East Yorkshire, U.K; diet 7378000) *ad libitum*. Fecal samples were collected from each mouse 1, 2, 3, 4, 6, 8, 10, 12, and 14 days after gavage. Each fecal pellet was placed in 1 mL of PBS and homogenized by vortexing. A 10 μL aliquot of the homogenate was plated directly on LB agar (n=4 plates/sample). Twenty four hours later, colonies were collected by scraping (1 mL PBS). Genomic DNA was isolated using the phenol:chloroform method (72), and then fragmented by sonication to 300~500 bp (Bioruptor® sonicator for ultrasonic liquid processing; 20 cycles of 30 sec ON at high power/30 sec OFF).

Illumina sequencing libraries were prepared as described by the manufacturer. Sample-specific, 8 nt Hamming barcodes 73) were incorporated into the sequencing adapter for multiplex sequencing (n=96 samples/lane of an Illumina HiSeq 2000 flowcell; 80–100 million reads/lane; >500,000 42 nt reads/sample).

Raw-reads were assigned to each sample using the barcodes, and then mapped to the database containing the genomes of both strains using Eland (75). Only reads that could be uniquely assigned to one of the two genomes were used to score the relative representation of that strain in a given fecal sample. Counts for each strain were normalized by the informative genome size (IGS). The IGS of each strain was calculated by generating, *in silico*, a mock sample containing a 1:1 ratio of both strains, and then mapping the sample reads back to each genome: the IGS for each strain was calculated from the number of reads that mapped uniquely to that strain's genome. The ratio of the two strains was calculated as ρ(FitSeq) after normalization by IGS.

Since bacterial cells had been harvested from LB agar plates after a 24 h incubation, to calculate the ratio (ρ) of strains in fecal samples, we needed to transform ρ(FitSeq) using the relative growth rate. To measure the relative growth rate, a 1:1 mixture of strain Ec72_E1U1 and Ec72_E3U1 were grown on LB agar for 24 h at room temperature as above. Bacterial cells were collected and the detected ratio (ρ(1:1)) calculated by FitSeq. Knowing that the original ratio of the two strains was 1, the relative growth rate ratio (v1/v2) of the two strains could be calculated as $\frac{v1}{v2} = \rho(1:1)$. Five independent mixtures were prepared and measured, and the average results used as the ratio for growth rate of the two strains on LB agar. Thus, we were able to calculate the ratios of two strains in fecal samples as

$$\rho = \frac{\rho(\text{FitSeq})}{\frac{v1}{v2}} = \rho(\text{FitSeq})/\rho(1:1).$$

Characterization of virulence related phenotypes in vitro

Hemagglutination assays were performed on cells normalized to an OD600 of 1.0 as described previously (n=3 biological replicates performed on different days; 2 technical replicates/biological replicate) (76). For biofilm assays, shaken cultures were grown

overnight at 37°C in LB were subcultured at a 1:1000 dilution into LB and grown statically in untreated polyvinylchloride 96 well plates at room temperature for 48 h. Adherent biomass was stained with 0.05% crystal violet, rinsed, solubilized with 35% acetic acid and quantified by measuring absorbance at 595 nm of the solubilized crystal violet (assays performed in duplicate; 5 technical replicates/biological replicate). Pellicle biofilms were performed by subculturing a 1:1000 dilution of an overnight LB shaken culture into Yeast Extract/Casamino Acid (YESCA) medium and incubating statically at 30°C for 72 h (pellicle assays performed twice). For motility assays, cultures were incubated statically at 37°C for 24 h in 10 mL of LB, subcultured at a 1:1000 dilution into 10 mL of fresh LB, and incubated again statically at 37°C for 24 h. Swimming motility was measured in 0.25% LB agar and swarming motility was measured in 0.6% LB agar supplemented with 0.5% glucose (swimming and swarming data shown are representative of results obtained from two separate experiments) (76). Hemolysin production was examined by plating on blood agar (assays performed in duplicate).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Foxman B. Epidemiology of urinary tract infections: Incidence, morbidity, and economic costs. Am J Med. 2002; 113:5.

2. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol. 2010; 7:653. [PubMed: 21139641]

3. Hooton TM. Recurrent urinary tract infection in women. Int J Antimicrob Ag. 2001; 17:259.

4. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. Nature. 2009; 457:480. [PubMed: 19043404]

5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010; 464:59. [PubMed: 20203603]

6. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207. [PubMed: 22699609]

7. Schilling JD, Mulvey MA, Hultgren SJ. Dynamic interactions between host and pathogen during acute urinary tract infections. Urology. 2001; 57:56. [PubMed: 11378051]

8. Song J, Abraham SN. TLR-mediated immune responses in the urinary tract. Curr Opin Microbiol. 2008; 11:66. [PubMed: 18243043]

9. Zasloff M. Antimicrobial peptides, normally sterile urinary innate immunity, and the tract. J Am Soc Nephrol. 2007; 18:2810. [PubMed: 17942949]

10. Sivick KE, Mobley HL. Waging war against uropathogenic Escherichia coli: winning back the urinary tract. Infect Immun. 2010; 78:568. [PubMed: 19917708]

11. Hunstad DA, Justice SS. Intracellular lifestyles and immune evasion strategies of uropathogenic Escherichia coli. Annu Rev Microbiol. 2010; 64:203. [PubMed: 20825346]

12. Hannan TJ, Totsika M, Mansfield KJ, Moore KH, Schembri MA, Hultgren SJ. Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic Escherichia coli bladder infection. FEMS Microbiol Rev. 2012; 36:616. [PubMed: 22404313]

13. HM, Kostakioti M, Cusumano CK, Hannan TJ, Janetka JW, Hultgren SJ. Distinguishing the contribution of Type 1 pili from that of other QseB-misregulated factors when QseC Is absent during urinary tract infection. Infect Immun. 2012; 80:2826. [PubMed: 22665375]

14. Waksman G, Hultgren SJ. Structural biology of the chaperone-usher pathway of pilus biogenesis. Nature Rev Microbiol. 2009; 7:765. [PubMed: 19820722]

15. Hung CS, Bouckaert J, Hung D, Pinkner J, Widberg C, DeFusco A, Auguste CG, Strouse R, Langermann S, Waksman G, Hultgren SJ. Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. Mol Microbiol. 2002; 44:903. [PubMed: 12010488]

16. Xie B, Zhou G, Chan SY, Shapiro E, Kong XP, Wu XR, Sun TT, Costello CE. Distinct glycan structures of uroplakins Ia and Ib: structural basis for the selective binding of FimH adhesin to uroplakin Ia. J Biol Chem. 2006; 281:14644. [PubMed: 16567801]

17. Martinez JJ, Mulvey MA, Schilling JD, Pinkner JS, Hultgren SJ. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. Embo J. 2000; 19:2803. [PubMed: 10856226]

18. Duncan MJ, Li GJ, Shin JS, Carson JL, Abraham SN. Bacterial penetration of bladder epithelium through lipid rafts. J Biol Chem. 2004; 279:18944. [PubMed: 14976212]

19. Eto DS, Jones TA, Sundsbak JL, Mulvey MA. Integrin-mediated host cell invasion by type 1-piliated uropathogenic Escherichia coli. PLoS Pathog. 2007; 3:949.

20. Bishop BL, Duncan MJ, Song J, Li G, Zaas D, Abraham SN. Cyclic AMP-regulated exocytosis of Escherichia coli from infected bladder epithelial cells. Nat Med. 2007; 13:625. [PubMed: 17417648]

21. Justice SS, Hung C, Theriot JA, Fletcher DA, Anderson GG, Footer MJ, Hultgren SJ. Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. Proc Natl Acad Sci U S A. 2004; 101:1333. [PubMed: 14739341]

22. Anderson GG, Palermo JJ, Schilling JD, Roth R, Heuser J, Hultgren SJ. Intracellular bacterial biofilm-like pods in urinary tract infections. Science. 2003; 301:105. [PubMed: 12843396]

23. Schilling JD, Lorenz RG, Hultgren SJ. Effect of trimethoprim-sulfamethoxazole on recurrent bacteriuria and bacterial persistence in mice infected with uropathogenic Escherichia coli. Infect Immun. 2002; 70:7042. [PubMed: 12438384]

24. Blango MG, Mulvey MA. Persistence of Uropathogenic Escherichia coli in the Face of Multiple Antibiotics. Antimicrob Agents Chemother. 2010; 54:1855. [PubMed: 20231390]

25. Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, Heuser J, Hultgren SJ. Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. Science. 1998; 282:1494. [PubMed: 9822381]

26. Rosen DA, Hooton TM, Stamm WE, Humphrey PA, Hultgren SJ. Detection of intracellular bacterial communities in human urinary tract infection. PLoS Med. 2007; 4:1949.

27. Mulvey MA, Schilling JD, Hultgren SJ. Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. Infect Immun. 2001; 69:4572. [PubMed: 11402001]

28. Mysorekar IU, Hultgren SJ. Mechanisms of uropathogenic Escherichia coli persistence and eradication from the urinary tract. Proc Natl Acad Sci U S A. 2006; 103:14170. [PubMed: 16968784]

29. Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC. J Infectious Dis. 2000; 181:1753. [PubMed: 10823778]

30. Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL. Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. Proc Natl Acad Sci U S A. 1998; 95:8922. [PubMed: 9671780]

31. Chattopadhyay S, Feldgarden M, Weissman SJ, Dykhuizen DE, van Belle G, Sokurenko EV. Haplotype diversity in "source-sink" dynamics of Escherichia coli urovirulence. J Mol Evol. 2007; 64:204. [PubMed: 17177088]

32. Shaikh N, Holt NJ, Johnson JR, Tarr PI. Fim operon variation in the emergence of Enterohemorrhagic Escherichia coli: an evolutionary and functional analysis. FEMS Microbiol Lett. 2007; 273:58. [PubMed: 17559392]

33. Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li ZL, Bouckaert J, Gordon JI, Hultgren SJ. Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. Proc Natl Acad Sci U S A. 2009; 106:22439. [PubMed: 20018753]

34. Garofalo CK, Hooton TA, Martin SM, Stamm WE, Palermo JJ, Gordon JI, Hultgren SJ. Escherichia coli from urine of female patients with urinary tract infections is competent for intracellular bacterial community formation. Infect Immun. 2007; 75:52. [PubMed: 17074856]

35. Sokurenko EV, Gomulkiewicz R, Dykhuizen DE. Source-sink dynamics of virulence evolution. Nat Rev Microbiol. 2006; 4:548. [PubMed: 16778839]

36. Czaja CA, Stamm WE, Stapleton AE, Roberts PL, Hawn TR, Scholes D, Samadpour M, Hultgren SJ, Hooton TM. Prospective Cohort Study of Microbial and Inflammatory Events Immediately Preceding Escherichia coli Recurrent Urinary Tract Infection in Women. J Infect Dis. 2009; 200:528. [PubMed: 19586416]

37. Paul S, Linardopoulou EV, Billig M, Tchesnokova V, Price LB, Johnson JR, Chattopadhyay S, Sokurenko EV. Role of Homologous Recombination in Adaptive Diversification of Extraintestinal Escherichia coli. J Bacteriol. 2013; 195:231. [PubMed: 23123908]

38. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821. [PubMed: 18349386]

39. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics. 2007; 8:64. [PubMed: 17324286]

40. Chen SL, Hung CS, Xu JA, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, Armstrong JR, Fulton RS, Latreille JP, Spieth J, Hooton TM, Mardis ER, Hultgren SJ, Gordon JI. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: A comparative genomics approach. Proc Natl Acad Sci U S A. 2006; 103:5977. [PubMed: 16585510]

41. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007; 23:673. [PubMed: 17237039]

42. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22:1658. [PubMed: 16731699]

43. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002; 12:656. [PubMed: 11932250]

44. Giannakis M, Backhed HK, Chen SL, Faith JJ, Wu M, Guruge JL, Engstrand L, Gordon JI. Response of gastric epithelial progenitors to Helicobacter pylori Isolates obtained from Swedish patients with chronic atrophic gastritis. J Biol Chem. 2009; 284:30383. [PubMed: 19723631]

45. Hung CS, Dodson KW, Hultgren SJ. A murine model of urinary tract infection. Nat Protoc. 2009; 4:1230. [PubMed: 19644462]

46. Selzer G, Som T, Itoh T, Tomizawa J. The Origin of Replication of Plasmid-P15a and Comparative Studies on the Nucleotide-Sequences around the Origin of Related Plasmids. Cell. 1983; 32:119. [PubMed: 6186390]

47. Lee AK, Falkow S. Constitutive and inducible green fluorescent protein expression in Bartonella henselae. Infect Immun. 1998; 66:3964. [PubMed: 9673287]

48. Cegelski L, Pinkner JS, Hammer ND, Cusumano CK, Hung CS, Chorell E, Aberg V, Walker JN, Seed PC, Almqvist F, Chapman MR, Hultgren SJ. Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. Nat Chem Biol. 2009; 5:913. [PubMed: 19915538]

49. Bryant WA, Krabben P, Baganz F, Zhou Y, Ward JM. The Analysis of Multiple Genome Comparisons in Genus Escherichia and Its Application to the Discovery of Uncharacterised Metabolic Genes in Uropathogenic Escherichia coli CFT073. Comp Funct Genomics. 2009:782924. [PubMed: 20204180]

50. Lehmacher A, Bockemuehl J. L-Sorbose utilization by virulent Escherichia coli and Shigella: Different metabolic adaptation of pathotypes. Int J Med Microbiol. 2007; 297:245. [PubMed: 17382590]

51. Fabich AJ, Jones SA, Chowdhury FZ, Cernosek A, Anderson A, Smalley D, McHargue JW, Hightower GA, Smith JT, Autieri SM, Leatham MP, Lins JJ, Allen RL, Laux DC, Cohen PS, Conway T. Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. Infect Immun. 2008; 76:1143. [PubMed: 18180286]

52. Alteri CJ, Smith SN, Mobley HLT. Fitness of Escherichia coli during Urinary Tract Infection Requires Gluconeogenesis and the TCA Cycle. PLoS Pathog. 2009; 5:e1000448. [PubMed: 19478872]

53. Spurbeck RR, Stapleton AE, Johnson JR, Walk ST, Hooton TM, Mobley HL. Fimbrial profiles predict virulence of uropathogenic Escherichia coli strains: contribution of ygi and yad fimbriae. Infect Immun. 2011; 79:4753. [PubMed: 21911462]

54. Chen SL, Jackson SL, Boyko EJ. Diabetes mellitus and urinary tract infection: epidemiology, pathogenesis and proposed studies in animal models. J Urol. 2009; 182:S51. [PubMed: 19846134]

55. Sokurenko EV, Feldgarden M, Trintchina E, Weissman SJ, Avagyan S, Chattopadhyay S, Johnson JR, Dykhuizen DE. Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in Escherichia coli. Mol Biol Evol. 2004; 21:1373. [PubMed: 15044596]

56. Lane MC, Lockatell V, Monterosso G, Lamphier D, Weinert J, Hebel JR, Johnson DE, Mobley HLT. Role of Motility in the Colonization of Uropathogenic Escherichia coli in the Urinary Tract. Infect Immun. 2005; 73:7644. [PubMed: 16239569]

57. Wright KJ, Seed PC, Hultgren SJ. Uropathogenic Escherichia coli Flagella Aid in Efficient Urinary Tract Colonization. Infect Immun. 2005; 73:7657. [PubMed: 16239570]

58. Monday SR, Minnich SA, Feng PCH. A 12-base-pair deletion in the flagellar master control gene flhC causes nonmotility of the pathogenic German sorbitol-fermenting Escherichia coli O157: H-strains. J Bacteriol. 2004; 186:2319. [PubMed: 15060034]

59. Wang L, Rothemund D, Curd H, Reeves PR. Species-wide variation in the Escherichia coli flagellin (H-antigen) gene. J Bacteriol. 2003; 185:2936. [PubMed: 12700273]

60. Norinder BS, Koves B, Yadav M, Brauner A, Svanborg C. Do Escherichia coli strains causing acute cystitis have a distinct virulence repertoire? Microb Pathogenesis. 2012; 52:10.

61. Mabbett AN, Ulett GC, Watts RE, Tree JJ, Totsika M, Ong CLY, Wood JM, Monaghan W, Looke DF, Nimmo GR, Svanborg C, Schembri MA. Virulence properties of asymptomatic bacteriuria Escherichia coli. Int J Med Microbiol. 2009; 299:53. [PubMed: 18706859]

62. Zdziarski J, Svanborg C, Wullt B, Hacker J, Dobrindt U. Molecular basis of commensalism in the urinary tract: Low virulence or virulence attenuation? Infect Immun. 2008; 76:695. [PubMed: 18039831]

63. Hadjifrangiskou M, Kostakioti M, Chen SL, Henderson JP, Greene SE, Hultgren SJ. A central metabolic circuit controlled by QseC in pathogenic Escherichia coli. Mol Microbiol. 2011; 80:1516. [PubMed: 21542868]

64. Foxman B, Manning SD, Tallman P, Bauer R, Zhang LX, Koopman JS, Gillespie B, Sobel JD, Marrs CF. Uropathogenic Escherichia coli are more likely than commensal E-coli to be shared between heterosexual sex partners. Am J Epidemiol. 2002; 156:1133. [PubMed: 12480658]

65. Belanger L, Garenaux A, Harel J, Boulianne M, Nadeau E, Dozois CM. Escherichia coli from animal reservoirs as a potential source of human extraintestinal pathogenic E-coli. Fems Immunol Med Mic. 2011; 62:1.

66. Manges AR, Johnson JR. Food-Borne Origins of Escherichia coli Causing Extraintestinal Infections. Clin Infect Dis. 2012

67. Zhang QJ, Sahin O, McDermott PF, Payot S. Fitness of antimicrobial-resistant Campylobacter and Salmonella. Microbes Infect. 2006; 8:1972. [PubMed: 16714138]

68. Luangtongkum T, Shen ZQ, Seng VW, Sahin O, Jeon B, Liu P, Zhang QJ. Impaired Fitness and Transmission of Macrolide-Resistant Campylobacter jejuni in Its Natural Host. Antimicrob Agents Chemother (Bethesda). 2012; 56:1300.

69. Luo ND, Pereira S, Sahin O, Lin J, Huang SX, Michel L, Zhang QJ. Enhanced in vivo fitness of fluoroquinolone-resistant Campylobacter jejuni in the absence of antibiotic selection pressure. Proc Natl Acad Sci U S A. 2005; 102:541. [PubMed: 15634738]

70. Wright KJ, Seed PC, Hultgren SJ. Development of intracellular bacterial communities of uropathogenic Escherichia coli depends on type 1 pili. Cell Microbiol. 2007; 9:2230. [PubMed: 17490405]

71. Cusumano CK, Pinkner JS, Han Z, Greene SE, Ford BA, Crowley JR, Henderson JP, Janetka JW, Hultgren SJ. Treatment and prevention of urinary tract infection with orally active FimH inhibitors. Sci Transl Med. 2011; 3:109(ra115).

72. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005; 102:11070. [PubMed: 16033867]

73. Krishnan AR, Sweeney M, Vasic J, Galbraith DW, Vasic B. Barcodes for DNA sequencing with guaranteed error correction capability. Electron Lett. 2011; 47:236.

74. Chevreux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 1999; 99:45.

75. Cox, A. ELAND: Efficient large-scale alignment of nucleotide databases. Illumina, San Diego: 2007.

76. Kostakioti M, Hadjifrangiskou M, Pinkner JS, Hultgren SJ. QseC-mediated dephosphorylation of QseB is required for expression of genes associated with virulence in uropathogenic Escherichia coli. Mol Microbiol. 2009; 73:1020. [PubMed: 19703104]

Editor's Summary

UTI Bugs without Barriers

Urinary tract infections (UTIs) are among the most common infections in women, with uropathogenic *E. coli* (UPEC) being the major cause. Recurrent infections are troublesome and can persist for years. Studies in mice have led to the realization that UPEC strains can specialize so that once they enter the urinary tract they can invade bladder tissue, forming protected bacterial communities that contribute to recurrent UTIs. A prevailing view is that recurrent UTIs also represent repeated movement of UPEC strains from the gut to the bladder. This migration is thought to be unidirectional, reflecting a view that fitness (ability to succeed) in the bladder comes at a cost of loss of fitness in the gut. We re-examined this fitness "tradeoff" by characterizing the genomes of urine and fecal *E. coli* isolates obtained from four healthy women, enrolled in a large patient study of recurrent UTI, who each had three recurrent UTIs. In two women, the dominant UPEC strain in both their urine and feces was the same throughout all three UTIs. In the other two, the UPEC strain present in both urine and feces in the initial UTI episode was replaced by a different strain at the third recurrence. In mouse models of bladder infection and gut colonization, the strain that dominated in the later UTI episode had increased fitness in both habitats compared to the strain it replaced. Increased fitness correlated with genetic differences affecting nutrient utilization and virulence. Thus, recurrent UTI is complex and may involve strains moving freely, without fitness tradeoffs, between the bladder and gut in addition to invasion of bladder tissue. Whereas further human studies are needed to assess the role of gut bacteria and their genetic characteristics during recurrent UTI, this broader view could lead to new approaches for prevention, diagnosis, and treatment of this troublesome infection.
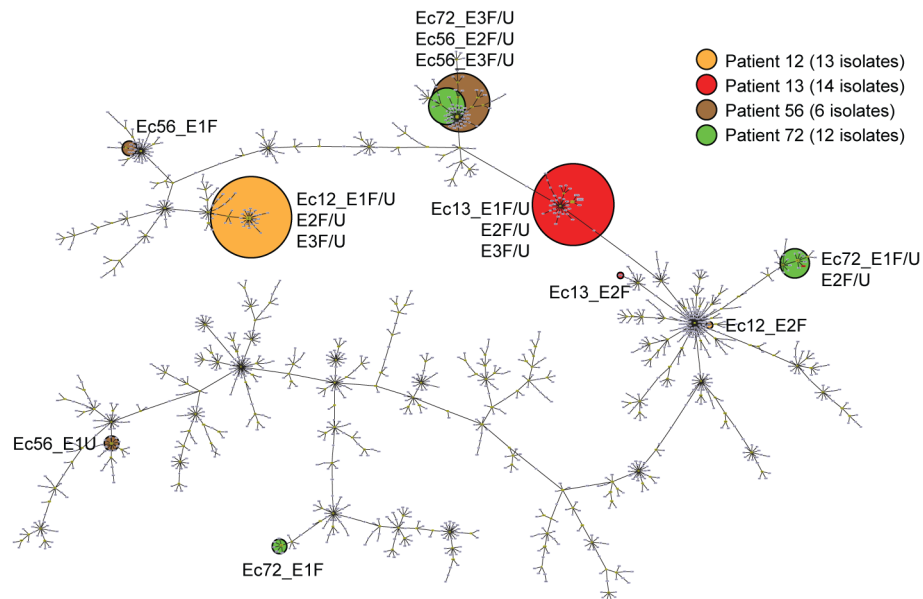
Fig. 1. MLST analysis of patient isolates

An eBURST diagram constructed based on publicly available *E. coli* MLST sequence types. MLST types from patient isolates from the present study are represented by colored circles; the color denotes the patient while the size (diameter) of the circle indicates the proportion of strains that share that same MLST type or an MLST type that differs by only one allele.
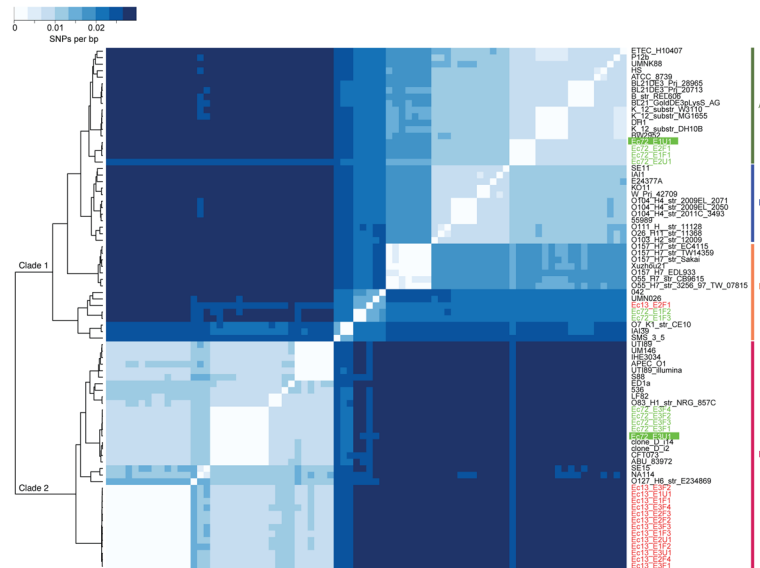
**Fig. 2. Heatmap of SNP differences between *E. coli* strains**

SNP rates (SNPs/aligned bp) between different sequenced strains were interpreted as a distance matrix. Hierarchical clustering was done on this symmetric matrix of SNP rates. Color in the central heatmap represents SNP rate as shown in the legend at the top left. A tree based on Euclidean distance is shown to left of the central heatmap. The colors of the strains indicate their patient of origin. UTI89_finished refers to the finished UTI89 genome (40), while UTI89_illumina is the re-sequenced and re-assembled genome from a 36 nt read dataset generated for this study. The comparison between the UTI89_finished and UTI89_illumina identified the noise range of the sequencing and assembly pipeline we used.
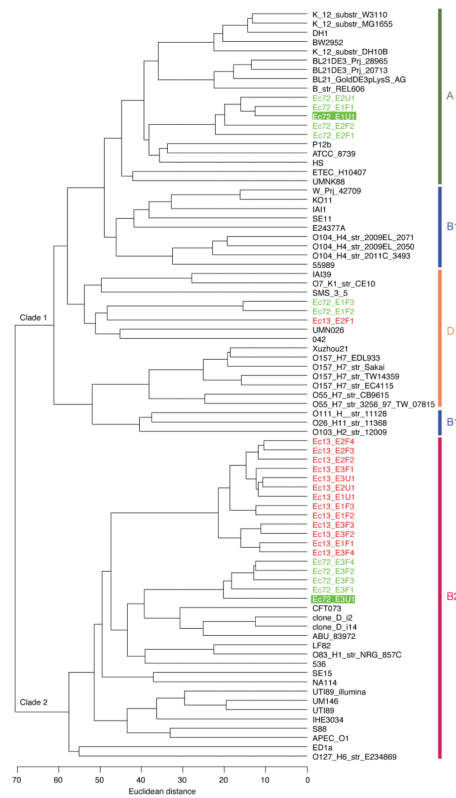
**Fig. 3. OGU-based clustering of sequenced *E. coli* genomes**

The presence or absence of OGUs in the newly sequenced UPEC genomes characterized in the present study and 54 other publicly available *E. coli* strains were used to construct an OGU matrix. Hierarchical clustering was performed on the matrix based on Euclidean distance. There are two clades suggested by the clustering analysis. Phylogenetic group membership of the strains is indicated with bars at the right of the figure. All isolates from Patient 13 are colored red while all isolates from Patient 72 are colored green. The two strains used for *in vivo* competition experiments and phenotype microarray analyses are highlighted.
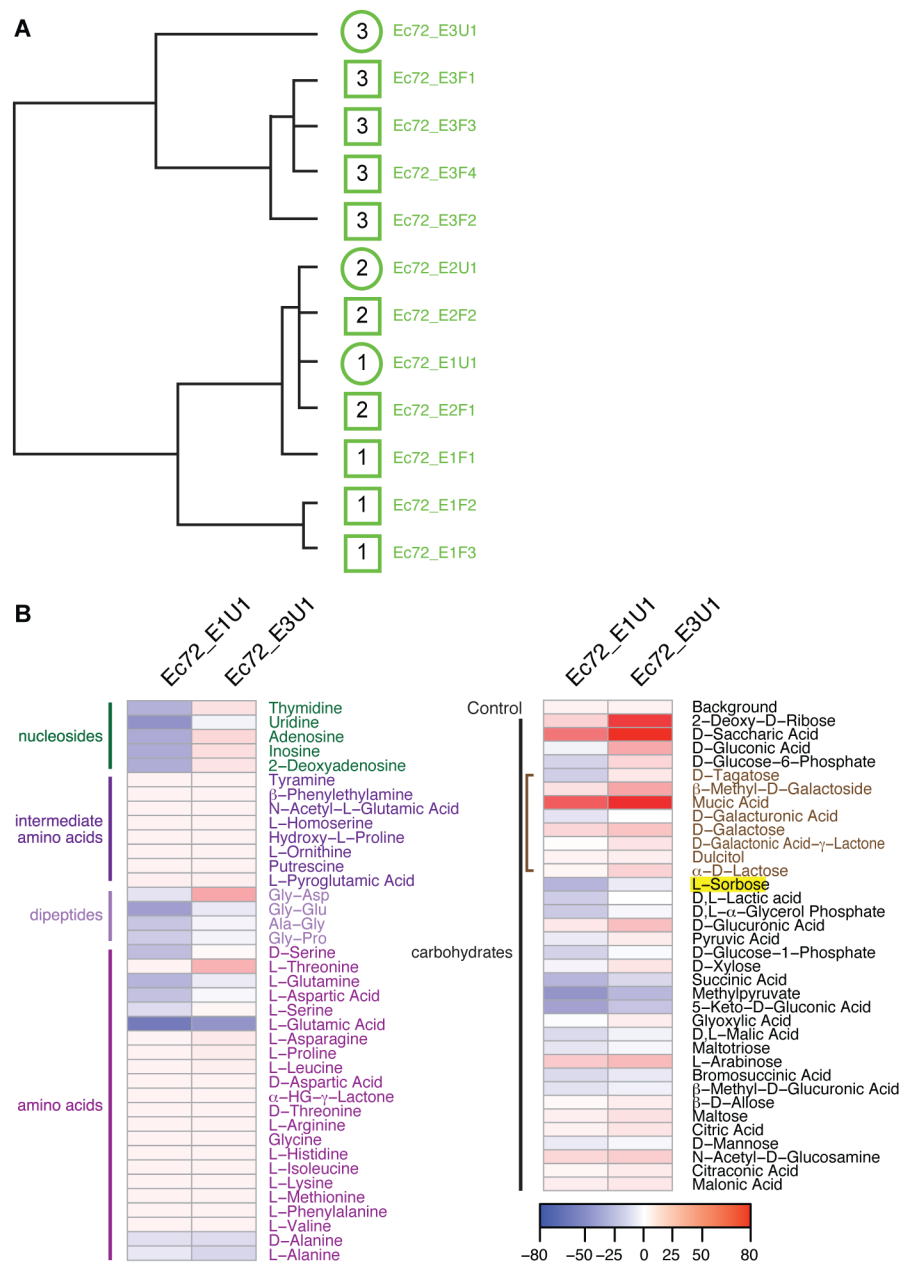
Fig. 4. Growth phenotypes of isolates from Patient 72

(A) Unsupervised hierarchical clustering of growth phenotypes as defined by Biolog phenotype microarrays. Squares represent fecal samples, circles denote urine samples, and the numbers inside squares and circles indicate the episode number from which the strain was derived. **(B)** Comparison of the growth phenotypes of strains Ec72_E1U1 and Ec72_E3U1 from UTI episodes 1 and 3, on carbohydrate, nucleoside and amino acid substrates. Carbohydrate substrates that result in greater than 10-fold growth of the episode 3 strain are shown. See Fig. S10 for other carbohydrates examined. Utilization of the bracketed carbohydrates requires genes involved in galactose metabolism. The color key denotes growth relative to the reference UPEC strain, UTI89.
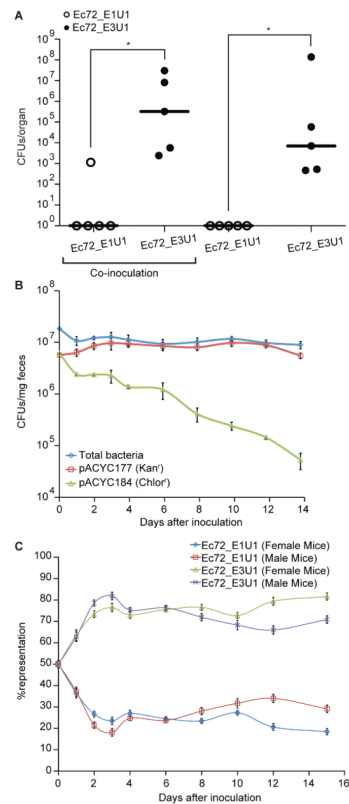
Fig. 5. *In vivo* fitness of urine isolates from Patient 72's UTI episodes 1 and 3
(A) Urine isolates obtained from UTI episodes 1 (isolate Ec72_E1U1) and 3 (Ec72_E3U1) were introduced separately or together into the bladders of female C3H/HeN mice. Data points represent CFUs/bladder in individual mice, and horizontal bars represent the median of CFUs/bladder of that strain in the experiments. Data on the y-axis are presented in log scale: therefore all 0s were plotted as 1 for visualization (*, $p<0.05$; two-tailed Mann-Whitney test). **(B)** The plasmid pACYC184 (Chlor[R]) was lost during a two-week colonization of the intestines of gnotobiotic mice or conferred a fitness disadvantage. CFU on LB agar plates without antibiotics represent all Ec72_E1U1 in the fecal samples. CFU on LB agar plates with kanamycin or chloramphenicol represent the Ec72_E1U1 carrying the corresponding antibiotic-resistance plasmid from the same fecal sample (data shown as mean +/− SEM). **(C)** FitSeq determination of the relative fitness of strains from Patient 72's UTI episodes 1 and 3 in the gut of gnotobiotic mice (data shown as mean ± SEM).