Published in final edited form as:

Nat Microbiol. 2022 May ; 7(5): 630–639. doi:10.1038/s41564-022-01107-x.

Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women

Colin J. Worby[1], **Henry L. Schreiber IV**[2,3,4], **Timothy J. Straub**[1], **Lucas R. van Dijk**[1,5], **Ryan A. Bronson**[1], **Benjamin S. Olson**[2,3], **Jerome S. Pinkner**[2], **Chloe L. P. Obernuefemann**[2], **Vanessa L. Muñoz**[2], **Alexandra E. Paharik**[2], **Philippe N. Azimzadeh**[2], **Bruce J. Walker**[6], **Christopher A. Desjardins**[1], **Wen-Chi Chou**[1], **Karla Bergeron**[7], **Sinéad B. Chapman**[1], **Aleksandra Klim**[7], **Abigail L. Manson**[1], **Thomas J. Hannan**[8], **Thomas M. Hooton**[9], **Andrew L. Kau**[3,10], **H. Henry Lai**[7,11], **Karen W. Dodson**[2,3], **Scott J. Hultgren**[*,2,3], **Ashlee M. Earl**[*,1]

[1]Infectious Disease and Microbiome Program, Broad Institute, Cambridge, MA 02142, USA

[2]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA

[3]Center for Women's Infectious Disease Research (CWIDR), Washington University School of Medicine, St. Louis, MO 63110, USA

[4]Division of Biology & Biological Engineering, California Institute of Technology, Pasadena, CA, USA

[5]Delft Bioinformatics Lab, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Netherlands

[6]Applied Invention, Cambridge MA, USA

[7]Department of Surgery, Division of Urologic Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA

[8]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

[9]Department of Medicine, University of Miami, Miami, Florida, USA

[10]Department of Medicine, Division of Allergy and Immunology. Washington University School of Medicine, St. Louis, MO 63110, USA

[*]To whom correspondence may be addressed (aearl@broadinstitute.org; hultgren@wustl.edu).

Author contributions

Study design HLS, KWD, SJH, AME

Study coordination HLS, KB, SBC, AK

Experiments performed HLS, JSP, CLPO, VLM, AEP

Data analysis CJW, HLS, TJS, LRvD, RAB, BSO, BJH, CAD, WCC

Consultation and supervision of analyses BJW, ALM, TJH, TMH, ALK, HHL, KWD, SJH, AME

Prepared the original draft CJW, ALM, KWD, SJH, AME

Review and approval of the final manuscript was provided by all authors.

Competing interests

The authors declare no competing interests.

[11.]Department of Anesthesiology, Washington University School of Medicine, St. Louis, MO 63110, USA

Abstract

Recurrent urinary tract infections (rUTIs) are a major health burden worldwide, with history of infection being a significant risk factor. While the gut is a known reservoir for uropathogenic bacteria, the role of the microbiota in rUTI remains unclear. We conducted a year-long study of women with (n=15) and without (n=16) history of rUTIs, from whom we collected urine, blood and monthly fecal samples for metagenomic and transcriptomic interrogation. During the study, 24 UTIs were reported, and additional samples collected during and after infection. The gut microbiome of individuals with a history of rUTI was significantly depleted in microbial richness and butyrate-producing bacteria compared to controls, reminiscent of other inflammatory conditions. However, *Escherichia coli* gut and bladder populations were comparable between cohorts in both relative abundance and phylogroup. Transcriptional analysis of peripheral blood mononuclear cells revealed expression profiles indicative of differential systemic immunity between cohorts. Altogether, these results suggest that rUTI susceptibility is in part mediated through the gut-bladder axis, comprising gut dysbiosis and differential immune response to bacterial bladder colonization, manifesting in symptoms.

Introduction

Urinary tract infections (UTIs) are among the most common bacterial infections worldwide and a significant cause of morbidity in females, with uropathogenic *Escherichia coli* (UPEC) being the primary causative agent [1]. One of the strongest risk factors for UTI is a history of prior UTIs [2], but the biological basis and risk factors for long-term recurrence remain unclear in otherwise healthy women. 20–30% of women diagnosed with a UTI will experience a recurrent UTI (rUTI), with some suffering six or more per year. Over one million women in the United States are referred to urologists each year because of rUTIs, and the rapid spread of antibiotic resistance in uropathogens is making treatment more challenging.

The gut is a reservoir for UPEC, and UTIs most commonly arise via the ascension of UPEC from the gut to the urinary tract [3–5]. Recent studies have explored the ‘gut microbiota-UTI axis’, showing that uropathogen abundance in the gut is a risk factor for UTI in kidney transplant patients [6], and that a ‘bloom’ in uropathogen gut abundance may precede infection [7]. Other studies have demonstrated differences in gut microbiome composition associated with children suffering UTIs [8], and with kidney transplant patients developing bacteriuria [9], compared to healthy controls. Furthermore, fecal microbiota transplants to treat *Clostridium difficile* infections may have the collateral effect of reducing the frequency of rUTI [10,11], suggesting that perturbation of the gut microbiota can modulate rUTI susceptibility.

It is increasingly accepted that the gut microbiota can play a role in conditions affecting distal organs – for instance, the gut-brain and gut-lung axes are the subject of ongoing research [12–15]. However, the gut-bladder axis – the spectrum of direct and indirect

interactions between gut flora and the bladder immune and/or infection status – remains uncharacterized, and the role of the gut microbiota in rUTI susceptibility is not well understood. No study has yet ascertained whether: i) gut dysbiosis is associated with rUTI susceptibility; ii) rUTI women have unique uropathogen dynamics within and between the gut and the bladder; or iii) microbiome-mediated immunological differences may be linked to rUTI susceptibility, as seen in other diseases [16].

Here, we present results from the UTI microbiome (UMB) project, a year-long clinical study of women with a history of rUTI and a matched cohort of healthy women. Our unique longitudinal study design allowed us to explore the importance and interdependence of the gut microbiota and *E. coli* strain dynamics in rUTI, susceptibility to infection, and host immune responses that may impact these dynamics. Using multi-omic techniques, we determined that: **i)** compared to healthy controls, women with a history of rUTI had a distinct, less diverse gut microbiota, depleted in butyrate producers and exhibiting characteristics of low-level inflammation; **ii**) differential immunological biomarkers suggest rUTI women may have a distinct immune state; **iii)** *E. coli* strains were transmitted from the gut to the bladder in both cohorts, though no UTI symptoms occurred in healthy controls; and **iv)** UTI-causing *E. coli* strains often persistently colonized the gut and were not permanently cleared by repeated antibiotic exposure. Thus, susceptibility to rUTI is in part mediated through a syndrome involving the gut-bladder axis, comprising a dysbiotic gut microbiome with reduced butyrate production and apparent alterations of systemic immunity. Our work shows that UPEC strains persist in the gut despite antibiotic treatment, which itself may exacerbate gut dysbiosis.

Results

Frequent antibiotic use and *E. coli* infections in rUTI cohort

Women with a history of rUTI were recruited to the UMB study, along with an age- and community-matched control cohort comprising healthy women (Methods). A total of 16 control and 15 rUTI women participated in the year-long study, providing monthly home-collected stool samples, as well as blood, urine and rectal swabs at enrollment and subsequent clinic visits for UTI treatment (Figure 1a). Participants completed monthly questionnaires on diet, symptoms, and behavior (Supp Data). There was a greater proportion of white women in the rUTI cohort, and self-reported antibiotic use was higher in this group in line with UTI treatment; otherwise, few dietary or behavioral differences were apparent (Extended Data Table 1).

A total of 24 UTIs occurred during the study, all in rUTI women, who each experienced 0–4 UTIs (Figure 1b). Nineteen were diagnosed by clinicians and five were inferred through self-reported symptoms and antibiotic use in the questionnaire during monthly sample collection. UTIs were typically treated with ciprofloxacin or nitrofurantoin. No significant temporal risk factors for UTI were identified amongst dietary or behavioral variables. Sexual intercourse is a well-known risk factor for UTI [2,17], and all 19 clinically diagnosed UTIs occurred following at least one reported sexual encounter in the previous two weeks (Extended Data Figure 1).

Urine samples collected at the time of clinical UTI diagnoses were plated on MacConkey agar; bacterial growth was detected (> 0 CFU/ml) from the majority (15/19; 79%, Supplementary Table 1). To determine the cause of infection, we sequenced 13 urine cultures, as well as uncultured urine, from all UTI diagnoses, defaulting to results from cultures when available. *E. coli* dominated 12/13 (92%) sequenced outgrowths; the remaining sample was dominated by *Klebsiella pneumoniae*. Sequencing uncultured urine from the remaining UTI samples identified uropathogens in a further four samples, including *E. coli* (2), *Enterococcus faecalis* and *Staphylococcus saprophyticus*, while two yielded no bacterial sequence (Supplementary Table 1). Based on sequencing, we defined 14 *E. coli* UTIs, comprising 82% of infections for which a bacterial cause could be inferred, broadly reflecting previous estimates of the proportion of all UTIs caused by *E. coli* [1].

rUTI gut depleted in microbial richness and butyrate-producers

It is increasingly recognized that the gut microbiota plays a role in a range of autoimmune and inflammatory diseases [18], as well as susceptibility to infection [16], and can alter inflammation in distal organs [19]. While previous studies have highlighted differential abundances of non-uropathogenic gut taxa as risk factors for bacteriuria in kidney transplant patients (reduced *Faecalibacterium* and *Romboutsia* [9]) and UTIs in children (reduced *Peptostreptococcaceae* [8]), it is unclear if these are risk factors for recurrence in otherwise healthy adult women. To explore this, we sequenced and analyzed the metagenomes of 367 longitudinal stool samples from both rUTI (n=197) and control (n=170) women (Figure 1b; Methods). Rectal swabs, collected during clinic visits, were not used to determine microbiome profiles.

There were broad differences in the gut microbiota composition between cohorts (Figure 2a–c). We fit linear mixed models with individual-level random effects to determine differences in diversity and composition between cohorts, adjusting for recent antibiotic use (Methods). Gut microbial richness was significantly lower, on average, in rUTI women (p=0.05, Figure 2c). At the phylum level, we saw elevated levels of *Bacteroidetes* (false discovery rate [FDR]=0.003) and a lower relative abundance of *Firmicutes* (FDR=0.02) in rUTI women. We identified 22 differentially abundant taxa (FDR<0.25) at lower taxonomic levels, 16 of which were depleted in rUTI women (Supplementary Table 2; Figure 2b), including *Faecalibacterium* as previously reported [9].

Several of the taxa reduced in the rUTI gut, including *Faecalibacterium*, *Akkermansia*, *Blautia* and *Eubacterium hallii*, are associated with short chain fatty acid (SCFA) production, including propionate and butyrate, which exert an anti-inflammatory effect in the gut through promotion of the intestinal barrier function and immunomodulation [20,21]. *Blautia* was additionally identified as the only taxon significantly depleted at the time of UTI relative to non-UTI samples (FDR=0.01). Cumulatively, SCFA producers, particularly butyrate producers, were significantly less abundant in rUTI women (p=0.001) (Figure 2d; Extended Data Figure 2). Four KEGG Orthogroups [22] representing components of butyrate production pathways were significantly reduced across the rUTI cohort (Supplementary Table 3). Functional analysis with HUMAnN2 [23] additionally revealed pathways depleted in the rUTI cohort, including those associated with sugar degradation and biosynthesis of

metabolite intermediates and amino acids (Supplementary Table 4), many of which were also found to be differentially abundant in a study of irritable bowel syndrome (IBS) patients with sugar malabsorption [24].

This loss of gut microbial richness, diversity, and butyrate-producing bacteria is also a hallmark of exposure to broad spectrum antibiotics, including ciprofloxacin [25–27], which was used to treat more than a third of UTIs in our study. Thus, we sought to determine whether antibiotic effects may contribute to the observed shifts in microbiome composition in rUTI women ('rUTI dysbiosis'). Though antibiotic exposure in the previous two weeks was associated with a significant reduction in microbial richness ($p=0.05$), this loss of richness was not sustained. Samples taken 2–6 weeks after antibiotic exposure were not significantly different from baseline levels ($p=0.2$). Furthermore, we saw no association between the reported number of antibiotic courses and average richness (Figure 2c), and no differences in the overall gut microbiome stability between cohorts, despite more frequent antibiotic treatment among UTI women (Extended Data Figure 3). We observed no differences in richness or in the abundance of butyrate producers between rUTI women with different antibiotic exposures (Extended Data Figure 4a–b). Within the rUTI group, the frequency of infections was not associated with microbial richness or the relative abundance of butyrate producers. The microbial richness of women suffering UTIs during the study did not differ significantly from that of rUTI women not reporting infections ($p=0.4$; Figure 2b–d). While we did not detect a lasting impact from individual antibiotic courses – there were few long-term trends among rUTI women over the study (Extended Data Figure 4c) – it is still possible that repeated antibiotic use over years may have contributed to the observed rUTI dysbiosis.

rUTI gut dysbiosis shares broad similarities with IBD

The depletion of butyrate-producing taxa and microbial richness, key characteristics of rUTI dysbiosis, are also observed in other gut inflammatory conditions, including nosocomial diarrhea [28], IBS [29], and inflammatory bowel disease (IBD) [20], particularly Crohn's disease [30], and thus may be indicative of gut inflammation in rUTI women. While IBD is a multifactorial disorder for which the causative role of gut microbes is incompletely understood [31], mouse models have helped demonstrate a causal relationship between gut dysbiosis and inflammation [32]. We compared our data to longitudinal gut microbiome data from adults with and without IBD in the Human Microbiome Project 2 (HMP2) study [33], which shared the same extraction and sequencing protocols (Methods). Relative to each study's control group, we found that the ten most significantly depleted species in the rUTI gut, including butyrate producers *F. prausnitzii* and *E. hallii*, were also depleted in the IBD gut. We further observed a significant overall correlation in the estimated change of species-level abundances associated with rUTI and IBD (Extended Data Figure 5), suggesting more general similarities.

There were also some notable differences. *Bacteroides*, significantly elevated in the rUTI group, did not differ between cohorts in the HMP2 study (Extended Data Figure 5), and were also decreased among IBD patients in other studies [34]. *E. coli* was significantly elevated in IBD patients in the HMP2 study, but showed no difference in average relative

abundance between our cohorts (Figure 2e). Diminished *Bacteroides* alongside elevated *Enterobacteriaceae* was also observed in patients with nosocomial diarrhea [28]. Diarrhea, also a symptom of IBD, is associated with reduced gut transit time and is known to enrich for organisms common in the upper gastrointestinal tract, including *Enterobacteriaceae* [35], at the expense of anaerobic organisms such as *Bacteroides* [36]. As such, rUTI women with low-level inflammation and no diarrhea may lack the depletion of *Bacteroides* and elevation of *Enterobacteriaceae* observed in diarrhea-associated conditions. It is also possible that the considerable differences in treatment regimens; i.e. antibiotics vs. anti-inflammatories, contribute to divergences of a common underlying inflammatory signal.

Differential host immune response potentially linked to rUTI

rUTI dysbiosis also shares similarities with immunological syndromes affecting distal sites. For example, depletion of butyrate producers has been associated with rheumatoid arthritis, a systemic autoimmune disease which can be partially ameliorated in animal models with oral butyrate supplementation [37,38]. Patients with chronic kidney disease also exhibit similar dysbiosis, including reduced *Parasutterella* and *Akkermansia*, the latter of which is inversely correlated with interleukin-10 levels, an anti-inflammatory cytokine [39]. We hypothesized that rUTI dysbiosis may also have an immunomodulatory role, potentially eliciting a differential immune response to bacterial invasion of the bladder. Thus, we explored immunological biomarkers from blood samples collected at enrollment and UTI, quantifying (i) a Luminex panel of human cytokines, chemokines, and growth factors involved in inflammation and T cell activation, and (ii) cell types and the transcriptional activity of peripheral blood mononuclear cells (PBMCs) (Methods).

Of the 39 Luminex analytes, one chemokine, plasma eotaxin-1, was higher in rUTI women vs. control women at enrollment, and is associated with intestinal inflammation [40]. Levels of eotaxin-1 are increased in colonic tissue of patients with active IBD [41]. Subsequent human eotaxin-1 ELISAs validated these results, highlighting an additional link to dysbiosis-driven perturbation of the immune state; though, since this result did not hold after adjusting for race, we could not rule out potential demographic confounders. Eotaxin-1 was also higher in blood plasma of rUTI women at the time of UTI vs. enrollment ($p=0.04$; Extended Data Figure 6b).

Our small cohort size provided limited statistical power to identify differential expression between cohorts based on PBMC RNA Seq data, and no large-scale differences were observed (Extended Data Figure 6a). However, we found two genes that were upregulated in the PBMCs of the rUTI cohort (FDR < 0.1), *ZNF266* and the long non-coding RNA *LINC00944* (Supplementary Table 5). *ZNF266* has been previously linked to urological health, as a potential PBMC biomarker for overactive bladder and incontinence in women [42]. *LINC00944* has been associated with inflammatory and immune-related signaling pathways, as well as tumor invading T lymphocytes in breast cancer, and markers for programmed cell-death [43]. Resting NK cells were significantly reduced at the time of UTI relative to baseline levels ($p=0.02$; Extended Data Figure 6c). NK cells help suppress bladder infection by UPEC in mice [44], so the loss of NK cells in the periphery may suggest a migration to the bladder at time of rUTI.

Gut & bladder *E. coli* dynamics similar between cohorts

Previous work has implicated gut dysbiosis and a depletion of butyrate-producing bacteria in enhanced susceptibility to gut colonization by pathogens, including *Salmonella* [45] and *C. difficile* [46]. While we could not quantify *absolute* species abundances, we observed no significant difference in the average relative abundance of *E. coli* between cohorts (Figure 2e), suggesting the rUTI dysbiotic gut is no more hospitable to *E. coli* colonization than controls. Further, we found no relationship between the relative abundances of *Escherichia* and butyrate producers in either cohort, suggesting that depletion of butyrate-producing bacteria does not enhance gut colonization by *Escherichia* (Extended Data Figure 7). We considered the possibility that a temporal increase, or bloom, in *E. coli* relative abundance is a rUTI risk factor. Of the samples collected in the 14 days preceding an *E. coli* UTI, 75% exhibited *E. coli* relative abundance at or above average levels in the gut (Extended Data Figure 8a–b). However, elevated *E. coli* levels were not predictive of UTIs; none of the 22 *E. coli* blooms (defined as *E. coli* relative abundance >10-fold higher than the intra-host mean) occurred in the two weeks prior to UTIs. Thänert et al. identified intestinal blooms of uropathogens preceding some UTIs, but similarly noted that blooms often occurred in the absence of infection [7], leading us to conclude that elevated levels of *E. coli* may facilitate transfer to the bladder but rarely manifest in infection. However, without frequent urine collection, we cannot rule out asymptomatic bladder colonization.

Though we did not detect differences in *E. coli* species dynamics, we hypothesized that rUTI dysbiosis may manifest in a qualitatively different *E. coli* population in the gut, contributing to increased rUTI susceptibility. We applied StrainGE [47] to explore *E. coli* strain-level diversity within stool metagenomes (Methods), and classified strains by phylogroup [48]. Patterns of strain carriage were similar in the rUTI (Figure 3) and control (Extended Data Figure 9) cohorts. Both the number of strains per sample and the phylogroup distribution were comparable between cohorts (Figure 4, Extended Data Figure 8c–d). While most *E. coli* strains (62%) were observed in one sample only, 22% were ‘persistent’, observed in at least one quarter of their carrier’s samples. Persistent strains were more likely to originate from phylogroups B2 and D (p=0.01), regardless of cohort, and were slightly more common in control women (OR=2.1 (0.9, 5.2), p=0.1), at odds with the hypothesis of differential colonization resistance to phylogroups associated with UPEC between cohorts.

We then applied StrainGE to all urine samples, seeking to elucidate differences in strain dynamics in the bladder. We found that 79% (11/14) of *E. coli* UTIs were caused by phylogroup B2 (n=7) or D (n=4) strains (Supplementary Table 1), approximately in line with previous studies [4,49]. Of the 24 healthy enrollment urine samples yielding sufficient bacterial DNA to be sequenced and profiled (Supplementary Table 6), we detected *E. coli* strains in 54% (13/24), including over half of samples (7/13) from control participants, despite the absence of symptoms. All but one of these strains also belonged to phylogroups B2 and D. Control urines carried *E. coli* strains that were phylogenetically similar to UTI-causing strains based on StrainGE predictions (Figure 4; Methods), despite divergent clinical outcomes.

Mapping urine metagenome assemblies to a curated virulence factor database showed that UTI-causing strains were enriched in virulence factors (including iron uptake systems (*sit*,

chu, *iro*, *ybt* operons), colibactin (*clb*), and type 6 secretion systems) relative to an *E. coli* species-wide database, though many of these were also present in the one urine sample from a control participant for which we had sufficient coverage to assess gene content (Methods, Supplementary Table 7). This transition of a likely urovirulent strain to the bladder of healthy women without eliciting UTI symptoms is consistent with previous studies which have been unable to identify genetic markers of urovirulence in mice [49], or consistently discriminate between UTI and asymptomatic bacteriuria strains in women [50]. Nevertheless, the divergence in clinical outcomes after bacterial bladder invasion may still arise due to phenotypic differences in *E. coli* strains reaching the bladder that are not readily apparent in genome comparisons. rUTI dysbiosis could have an impact on UPEC gene expression; it has been shown that higher SCFA levels are associated with down-regulation of *E. coli* virulence factors including fimbrial and flagellar genes [51]. However, such transcriptional analyses fall outside the scope of this study.

Antibiotic treatment fails to clear UTI-causing strains from gut

While it is well known that UTIs are most commonly caused by UPEC resident in the gut, their longitudinal dynamics of these strains within the gut are less well understood, despite the importance of such insights into developing rUTI prophylaxis. We applied StrainGE to all urine samples to identify UTI-causing strains and their gut dynamics, in particular at the time of UTI and after antibiotic exposure. Four rUTI women suffered multiple confirmed *E. coli* UTIs, though only one was a same strain recurrence (individual 8; Figure 3b). Comparisons of sequence data from urine samples and cultured rectal swabs from UTI clinic visits revealed that nearly all (11/12) *E. coli* UTIs, for which we had same-day rectal swabs, contained the same UTI strain, underscoring frequent gut to bladder transmission. The dominant *E. coli* strain in four of the rectal swab outgrowths was not the UTI-causing strain, suggesting some UTIs may be caused by minority strains. Only one UTI (individual 5, Figure 3) was caused by a strain never observed in another sample from that individual. This phylogroup B1 strain likely arose from a source other than the gut, such as the urinary tract or the vagina, also implicated as UPEC reservoirs [7,52].

We anticipated that antibiotic exposure - particularly ciprofloxacin - would impact gut carriage of *E. coli* strains, and may explain the lower frequency of persistent colonizers in the rUTI group. Indeed, *E. coli* strains were detected by StrainGE significantly less frequently in stool samples from the two weeks following antibiotic use (OR=0.3 (0.13, 0.68); p=0.004). However, many strains apparently cleared by antibiotics were observed again at later time points; in fact, none of the UTI-causing strains observed in the gut was permanently cleared following antibiotic exposure. It has previously been shown that coexistence of susceptible and resistant strains of the same lineage through acquisition/loss of mobile resistance elements can allow UPEC populations to rapidly adapt to repeated antibiotic exposure and persist in the gut [53]. While low-level persistence that is undetectable from sequencing data is a possibility, we plated a subset of post-treatment stool samples onto MacConkey agar to culture *E. coli*. In many cases, we observed no growth, suggesting absence rather than low-level persistence (Supplementary Table 8). Furthermore, profiling of 12 UTI-causing strains isolated from proximate stool samples demonstrated that the majority were susceptible to the antibiotics to which they were exposed (Supplementary Table 9).

While a single stool sample is not completely representative of the gut microbiota, this suggests that UTI-causing strains may be frequently reintroduced to the gut from alternative sources following antibiotic clearance of the bladder and gut.

Discussion

Our study design, data collection and culture-independent metagenomic sequencing approach allowed us to characterize dynamics of the gut-bladder axis in healthy and rUTI women. We propose that rUTI susceptibility is dependent, in part, on perturbation of the gut-bladder axis, which represents a previously undescribed syndrome, comprising gut dysbiosis and differential host immunology. While this study was not designed to identify causal links between gut dysbiosis, immune response and rUTI susceptibility, the proposed model is consistent with our findings and provides a benchmark to be tested in future studies. Compared to healthy controls, women suffering rUTI exhibited gut dysbiosis characterized by depleted levels of butyrate-producing bacteria and diminished microbial richness. This dysbiosis did not appear to impact *E. coli* dynamics within the gut; relative abundances and strain types were similar between cohorts, suggesting that gut carriage of urovirulent bacteria in itself is not a risk factor for rUTIs. Notably, *E. coli* was commonly identified in the urine of healthy women, including strains arising from UPEC-associated clades and harboring similar virulence factors. Based on our observations, rUTI gut dysbiosis is consistent with low-level gut inflammation, and is reminiscent of other disorders in which microbiome-mediated immunomodulation plays a role in disease severity.

Our study had a number of limitations. Firstly, due to the limited collection of urine samples in control women, it was not possible to robustly compare (i) the composition of the urine microbiome, and (ii) the frequency of (asymptomatic) strain transfer from gut to bladder between cohorts. Secondly, we did not assess the role of other potential reservoirs, such as the vagina, which could explain UTIs caused by strains never observed in the gut. Thirdly, while StrainGE offers a high-resolution view of *E. coli* strain dynamics in the gut and bladder, we cannot rule out the presence of additional, low abundance strains which could not be detected from the depth of metagenomic data generated. Finally, the small cohort size and infrequent blood sample collection provided limited power to assess differential expression in PBMCs. While we identified some indications of immunological differences between cohorts, our findings warrant further investigations to explore microbiome-host mucosal immune interactions in the context of rUTI susceptibility.

While identifying the origins of rUTI dysbiosis is outside the scope of this study, repeated antibiotic exposure is a plausible mechanism through which dysbiosis is maintained. The relatively short study period precluded us from establishing whether dysbiosis is the direct result of long-term antibiotic perturbation. In addition to the potentially detrimental impact of antibiotic use on the gut microbiota, we found that treatment also failed to clear UTI-causing strains from the gut in the long term. rUTI treatment protocols targeting UPEC strains in the gut with minimal disruption to other gut microbiota, such as small molecule therapeutics [54], may offer improved prospects. While more evidence is required to fully characterize the causal mechanisms between dysbiosis and infection, our work highlights the ineffectiveness and potential detrimental impact of current antibiotic therapies, as well

as the potential for microbiome therapeutics (e.g. fecal microbiota transplants [10]) to limit infections via restoration of a healthy bacterial community in the gut.

Methods

Study design & sample collection

Enrollment—This study was conducted with the approval and under the supervision of the Institutional Review Board of Washington University School of Medicine in St. Louis, MO. Women from the St. Louis, MO area reporting three or more UTIs in the past 12 months were recruited into the rUTI study arm, while women with no history of UTI (at most one UTI ever) were recruited into the control arm via the Department of Urological Surgery at Barnes-Jewish Hospital in St. Louis, MO. We excluded women who: i) had inflammatory bowel disease (IBD) or urological developmental defects (e.g., ureteral reflux, kidney agenesis, etc.), ii) were pregnant, iii) take antibiotics as prophylaxis for rUTI, and iv) were younger than 18 years or older than 45 at the time of enrollment. All participants provided informed consent. Microbiological information for previous UTIs was not available. A total of 16 control and 15 rUTI women aged between 18 and 45 were recruited to the study; participants were remunerated with gift cards for participation. 14 women in each cohort completed the entire study collection protocol; no participants who completed the study were excluded from downstream analyses. Participants who did not complete the study were included in cohort-level comparisons, but excluded from longitudinal analyses. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications, e.g. [55,56]. As an observational study with no intervention, with cohort membership based on predetermined criteria defined above, no subject randomization was required. Data collection and analysis were not performed blind to the conditions of the experiments.

Sample collection & storage—Participants provided blood and urine samples, as well as rectal swabs, at the initial clinic visit. UTIs were diagnosed during clinic visits; additional UTIs (not presenting at the study clinic) were inferred based on symptoms (painful urination, increased urgency/frequency of urination, cloudy urine) and antibiotic consumption reported in the monthly questionnaire. Women visiting the clinic during the study with UTI symptoms provided rectal swabs, blood and urine samples, and were requested to submit stool samples as soon as possible (within 24 hours) after the clinic visit, as well as at a two week follow-up time point.

All participants provided monthly stool samples for 12 months. Samples were collected at home, and submitted via mail following procedures developed in the Human Microbiome Project 2 [33]. Briefly, participants collected a fresh fecal sample in a disposable toilet hat and then aliquoted two teaspoon-sized scoops of stool each into a tube containing phosphate buffered saline and a tube containing 100% ethanol. Samples were overnight to the Broad Institute where they were stored at −80C until sample processing. All stool samples were shipped Monday to Thursday within each week to limit samples long term exposure to ambient temperature; samples were stored in patients' home freezers until shipment, if necessary. Questionnaires were completed with all monthly and clinical sample collections;

these captured self-reported antibiotic and drug use, dietary intake, sexual intercourse and UTI symptoms. Participants who did not provide stool samples and questionnaires at the beginning of each month were given phone call or email reminders to provide samples.

Sample processing

Blood sample preparation—A total of 15 mL of blood was collected from each patient during initial enrollment and UTI visits. The blood was stored on ice for less than 30 minutes and then mixed with an equal amount PBS with 2% fetal bovine serum (FBS). Peripheral blood mononuclear cell (PBMCs) were then isolated using SepMate PBMC isolation tubes (Stemcell Technologies) with Ficoll-Paque PLUS density gradient medium (Cytiva). Serum was collected during the PBMC isolation process and stored at −80C until use. PBMCs were washed with PBS plus 2% FBS and pelleted via centrifugation at 10,000 × g at room temperature for 5 minutes. PBMC cell pellets were then flash frozen and stored at −80C until RNA extraction.

Rectal swab and urine preparation—Rectal swabs were collected in the clinic and stored on ice for less than 30 minutes. Rectal swabs were washed in 2 mL of PBS. 1 mL of PBS was centrifuged at 10,000 × g at room temperature for 2 minutes and the PBS supernatant was removed. The bacterial/fecal pellet was then flash frozen and stored at −80C until DNA extraction. The remaining 1 mL was then used to make serial dilutions and then plated on both Luria Broth (LB) and MacConkey agar and incubated overnight at 37°C to quantify colony forming units (CFUs). After bacterial enumeration, bacteria from MacConkey and LB plates were scraped to collect bacterial outgrowths. Bacterial cells were washed with PBS, pelleted at 10,000 × g at room temperature for 2 minutes, flash frozen and then stored at −80°C until DNA extraction.

Mid-stream urine samples were collected in sterile containers and stored on ice for less than 30 minutes. 10 mL of urine was centrifuged at 10,000 × g at room temperature for 5 minutes. The resulting pellet was washed in PBS, pelleted again, and then flash frozen and stored at −80°C until DNA extraction. 1 mL of urine was used to make serial dilutions and then plated onto both LB and MacConkey and incubated overnight at 37°C to enumerate CFUs. After outgrowth, the plates were scraped to collect bacterial colonies, which were then washed with PBS, pelleted at 10,000 × g at room temperature for 2 minutes, flash frozen and then stored at −80°C until DNA extraction.

RNA Extraction - PBMCs—RNA was extracted from stored PBMCs using TRIzol Reagent (cat. no. 15596–026 and 15596–018; Life Technologies), according to the manufacturer's protocol. Briefly, 0.75 mL of TRIzol was added per 0.25 mL of sample and cells were lysed by several rounds of pipetting. Samples were incubated for five minutes at room temperature. Chloroform was added to the samples at the recommended concentration and samples were incubated shaking for 15 seconds and set to rest for 2–3 minutes at room temperature. After incubation, samples were centrifuged at 12,000 × *g* for 15 minutes at 4°C. The aqueous phase was collected for RNA isolation. RNA was precipitated using 100% isopropanol and incubated at room temperature for 10 minutes, followed by centrifugation at 12,000 × *g* for 10 minutes at 4°C. The precipitated RNA was washed according to

the protocol using 75% ethanol and resuspended in RNase-free water. Extracted RNA was stored at −80°C until further use.

DNA Extraction – Rectal Swabs & Urine—DNA was extracted from rectal swabs and urine samples plated on MacConkey agar using the Wizard Genomic DNA Purification Kit (Promega), according to the manufacturer's protocol. Briefly, samples were resuspended in 600 uL of Nuclei Lysis solution and incubated at 80°C for five minutes, then cooled to room temperature. RNase solution was added to samples and incubated for 15 minutes at 37°C, then cooled to room temperature. 200 uL of Protein Precipitation solution was added to the RNase-treated sample, vortexed for 20 seconds, and incubated on ice for 5 minutes. After incubation, samples were centrifuged for 3 minutes at 16,000 × *g* and the supernatant was transferred to a 1.5 mL microcentrifuge tube containing 600 uL of isopropanol. Samples were gently mixed and centrifuged for 2 minutes at 16,000 × *g*. The supernatant was removed and the DNA pellet was washed with 70% ethanol. Samples were centrifuged for 2 minutes at 16,000 × *g,* ethanol was aspirated and DNA pellets were air-dried for 15 minutes. The DNA pellet was rehydrated with DNA Rehydration solution and incubated at 65°C for 1 hour. Extracted DNA was stored at 4C for short-term storage and at 80°C for long-term storage until further use.

DNA Extraction – Stool—Total nucleic acid from stool was extracted following the HMP2 protocol [33], the basis of which is the Chemagic MSM I with the Chemagic DNA Blood Kit-96 from Perkin Elmer. DNA samples were quantified using a fluorescence-based PicoGreen assay.

WMS sequencing & sequence data processing—Libraries were constructed from extracted DNA from stool, urine, rectal swabs, and plate scrapes using the NexteraXT kit (Illumina). Then, libraries were sequenced on a HiSeq 2500 (Illumina) in 101 bp paired-end read mode and/or a HiSeq X10 (Illumina) in 151 bp paired-end read mode. Sequence data was then demultiplexed. Samples that were sequenced multiple times on different runs were pooled together. Reads were processed with KneadData (v0.7.2, https://huttenhower.sph.harvard.edu/kneaddata/) to remove adapter sequence and trim low base qualities (with Trimmomatic), as well as to remove human-derived sequences (by aligning to human genome with bowtie2).

Luminex assays—Custom Luminex magnetic bead assay kit was obtained from R&D systems (product LXSAHM). Analytes from Human Inflammation and Human T Cell Response panels were chosen for the custom kit of 39 analytes: CXCL1/GROalpha, IL-1alpha, M-CSF/CSF1, LIF, Ltalpha/TNF-b, MIF, APRIL, CCL11/Eotaxin, CCL4/MIP-1b, CXCL8/IL-8, IFN-g, IL-1b, IL-11, IL-13, IL-17A, IL-18, IL-21, IL-27, IL-31, IL-4, IL-6, MMP-1, TNF-a, BAFF/BlyS, CCL2/MCP-1, CX3CL1/Fractalkine, CXCL5/ENA-78, GM-CSF, IL-10, IL-12p70, IL-15, IL-17E/IL-25, IL-2, IL-22, IL-28A/INF-12, IL-33, IL-5, IL-7, MMP-3. Detection of the analytes in human plasma samples was performed using the Curiox DropArray system for miniaturization of magnetic bead multiplex kits. Plasma samples were diluted 2-fold for the assay. Results were read and quantified using a BioPlex multiplex plate reader and Microplate Manager software (v5).

Eotaxin ELISA—Plasma eotaxin (CCL11) levels from rUTI and control patients were measured using the Eotaxin (CCL11) Human Simple Step ELISA kit (cat. No. Ab185985; Abcam), according to the manufacturer's protocol. Briefly, plasma samples were diluted into sample diluent and 50 uL of sample and 50 uL of antibody cocktail were added to 96 well plate strips. Plates were sealed and incubated shaking for one hour at room temperature. Wells were washed three times with 1x wash buffer and inverted to remove excess liquid. 100 uL of TMB substrate was added to each well; plates were covered to protect from light and incubated shaking for 10 minutes. Stop solution (100 uL) was added to each well and plates were incubated shaking for one minute. The OD_{450} was measured and recorded to determine the concentration of Eotaxin in pg/mL.

Sequence data analysis

Community profiling and metrics—Bacterial community composition was determined using MetaPhlAn2 (v2.7.0 with db v20) [57] on KneadData-processed sequences. Functional profiling was performed by HUMAnN2 (v2.8.1, database downloaded in October 2016) [23] on KneadData-processed sequences. Diversity metrics and Bray-Curtis distances were derived from the MetaPhlAn2 relative abundance output using the *vegan* package in R [https://cran.r-project.org/web/packages/vegan/].

PBMC RNASeq analysis—Sequences from PBMC extracted mRNA were aligned to the human reference genome (hg19, Bioproject PRJNA31257) using the STAR aligner [58]. Picard-Tools (https://broadinstitute.github.io/picard/) was used to mark duplicate reads. Read counts per gene were generated with subread featureCounts [59]. Read counts were normalized into Counts Per Million (CPM) using edgeR [60]. This normalized read count matrix was then used as input for CIBERSORT using the LM22 signature gene set [61]. Results from CIBERSORT reported the relative abundance of 22 different immune cell types, including both PBMC and non-PBMC cell types, and it was used to remove three samples that were contaminated with 5% or greater of non-PBMC cell types. The CIBERSORT filtered set of samples was used to perform differential gene expression analysis using DESeq2 [62]. Baseline healthy control samples were compared to baseline rUTI samples. Due to limited sample numbers and potential confounding, we included only samples collected from caucasian women in this analysis. Results driven by single outlying data points were not considered.

E. coli strain profiling—In order to track *E. coli* strain dynamics we used Strain Genome Explorer (StrainGE), which we extensively benchmarked for use on low abundance species in the context of typical Illumina sequencer error [47]. We applied the StrainGST module of StrainGE to identify representative *E. coli* strains in all stool, urine and rectal swabs, using an *E. coli* reference database generated from RefSeq complete genomes, as detailed in that paper. Strains mapping to the same representative reference genome in this database typically have an ANI of at least 99.9%. To provide further evidence that same-strain calls from sample pairs from the same host were indeed matches, we ran the StrainGR module of StrainGE, which calculates alignment-based similarity metrics. We used benchmarked thresholds to determine strain matches; strain pairs with a common callable genome >0.5%,

Jaccard gap similarity >0.95 and average callable nucleotide identity >99.95% were deemed matches.

Determination of UTI-causing strains—Urine samples provided at the time of UTI diagnosis were plated on MacConkey agar. Sequence data was generated from DNA extracted from uncultured urine, and/or outgrowth on selective media. The cause of UTI was deemed to be the most abundant uropathogen, using outgrowth data where available, uncultured urine otherwise. Species were determined to be uropathogens based on UTI prevalence studies, e.g. [1].

Determination of virulence factors—Urine metagenomes for which *E. coli* represented the dominant species were assembled using SPAdes [63]. To detect virulence factors in *E. coli* references (see StrainGST section above) and assembled genomes from study samples, we used the Virulence Factor Database (VFDB) for *E. coli* and the type 6 secretion system (T6SS) database (SecReT6) in genome-wide BLAST+ searches. Though VFDB contains T6SS genes, we removed them in favor of the T6SS-specific database for a T6SS-specific analytical pipeline. Other VFDB hits from blastn were filtered for ≥90% identity and ≥90% coverage. All *E. coli* genomes were separated by phylogroup for enrichment analysis, where Fisher's Exact test was used to determine the significance of virulence factor enrichment in a certain phylogroup. T6SS hits were filtered for ≥90% identity and ≥90% coverage and the system was considered present where at least 12 different adjacent T6SS genes were present. Again, an enrichment analysis was performed using Fisher's Exact test to determine the significance of T6SS presence in certain phylogroups.

Statistical testing & models

rUTI risk factors—We used questionnaire responses to determine if any dietary or behavioral factors were associated with rUTI. We first compared the proportion of participants in each cohort who responded positively to binary variables (e.g. dairy consumption, alcohol etc. in the previous two weeks) in more than 50% or responses, and used a Fisher's Exact test to determine significance. We next fit mixed effects logistic regression models to determine temporal risk factors for UTIs. Samples collected within 3 days of UTI diagnosis were classified as 'time of UTI'; this binary variable was fit as a function of host (random effects term) and each dietary or behavioral response variable collected in the questionnaire. Variables with limited or no variance were excluded.

Identifying differences at the cohort level and time of UTI—We fit mixed effects linear regression models to compare the structure, diversity and function of the gut microbiome between cohorts, following similar approaches employed by previous studies, e.g. [64]. For this purpose, we used sequence data from all collected stool samples, but did not include rectal samples. An arcsine square root transformation was applied to relative abundance values. Features (transformed relative abundances, diversity, microbial richness) were fit as a function of host (random effects term), cohort (categorical variable), and terms for antibiotic use and race (categorical variable) to adjust for potential confounding effects. To assess change in relative abundances at relevant timepoints, we also fit models including covariates for 'pre-UTI' (14 days preceding UTI diagnosis), 'time of UTI' (three

days either side of UTI diagnosis), or post antibiotics (<14 days post antibiotic exposure) as binary variables. All taxa with more than 10% non-zero values were fitted using the *lme4* function in R. Significance of associations was determined using Wald's test, and p-values were adjusted for multiple hypothesis testing using Benjamini-Hochberg correction at each taxonomic level.

The relative abundance of SCFA producers was additionally compared between cohorts; butyrate- and propionate-producing species were determined based on functional capacity to produce butyrate and propionate [65]. These species' relative abundances were then aggregated and compared as above.

We compared the stability of the microbiome between cohorts by assessing the distributions of within-host pairwise Bray-Curtis (BC) dissimilarities between individuals. Since rUTI women had, on average, slightly more frequent sampling than control women, due to the additional follow-up samples after UTI diagnoses, this metric may be biased towards smaller values in this cohort. However, we observed no significant trend between BC dissimilarity and time between samples, suggesting no detectable long-term trends. Furthermore, we detected no difference in the distribution of time-adjusted BC distances (BC divided by number of days between samples) between cohorts.

IBD comparisons—To compare rUTI dysbiosis to an IBD gut state, we downloaded MetaPhlAn2 output from the HMP2 study [33], (ibdmdb.org). We extracted longitudinal samples from adults with IBD (diagnosis='UC' or 'CD') and non-IBD controls (diagnosis='nonIBD'). We fit linear mixed effects models with standardized relative abundances as a function of host (random effects term), race (race='white'; binary term) and recent antibiotic use. Fitted coefficients for the IBD and rUTI cohorts are then plotted in Extended Data Figure 5.

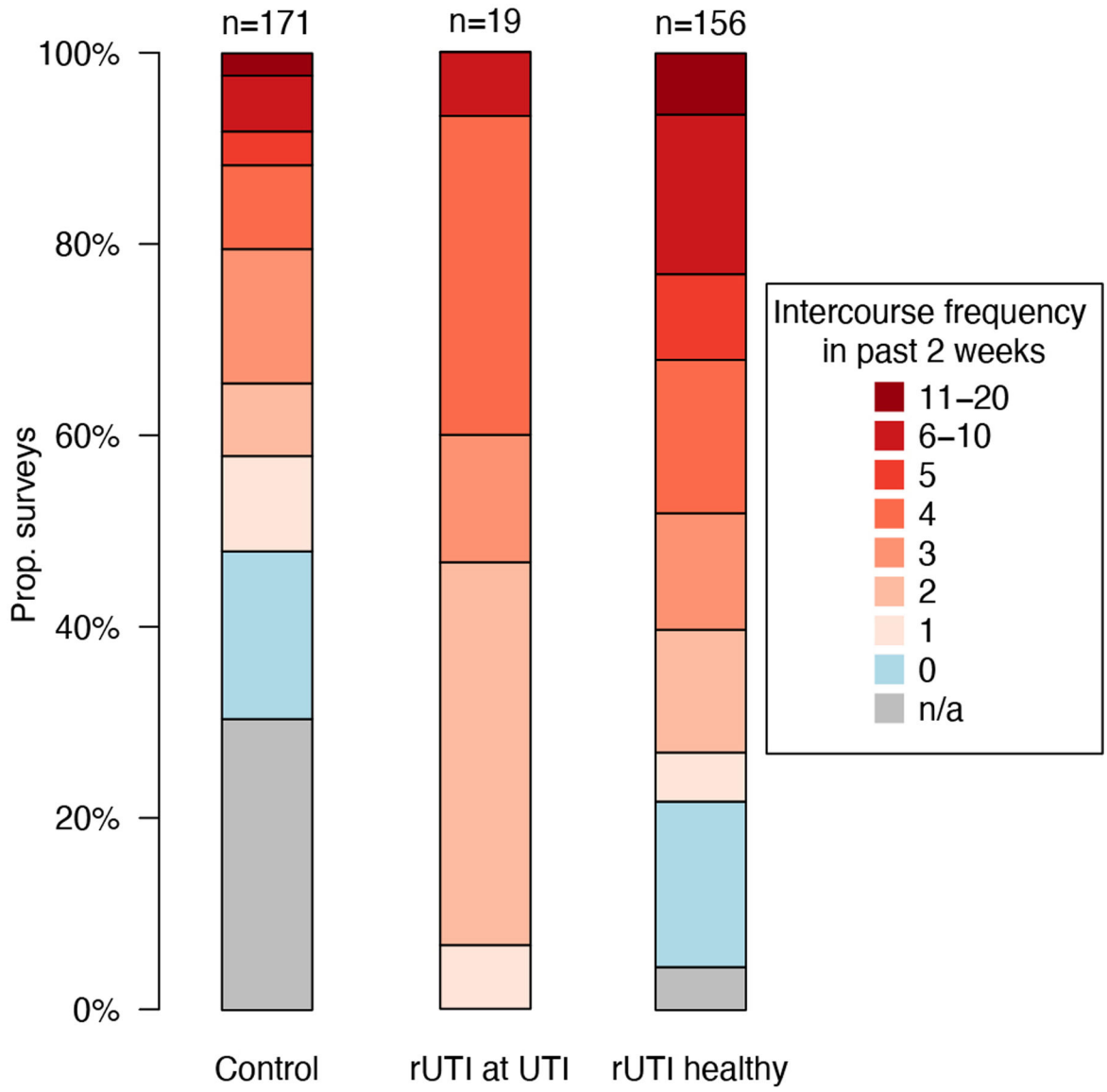
Data availability

Metagenomic sequence data are available from the Sequence Read Archive under Bioproject PRJNA400628. PBMC RNASeq data are available from the database of Genotypes and Phenotypes (dbGaP) under project number phs002728. Questionnaire data, output files from MetaPhlan2, Humann2, StrainGE are available from github.com/cworby/UMB-study.

Code availability

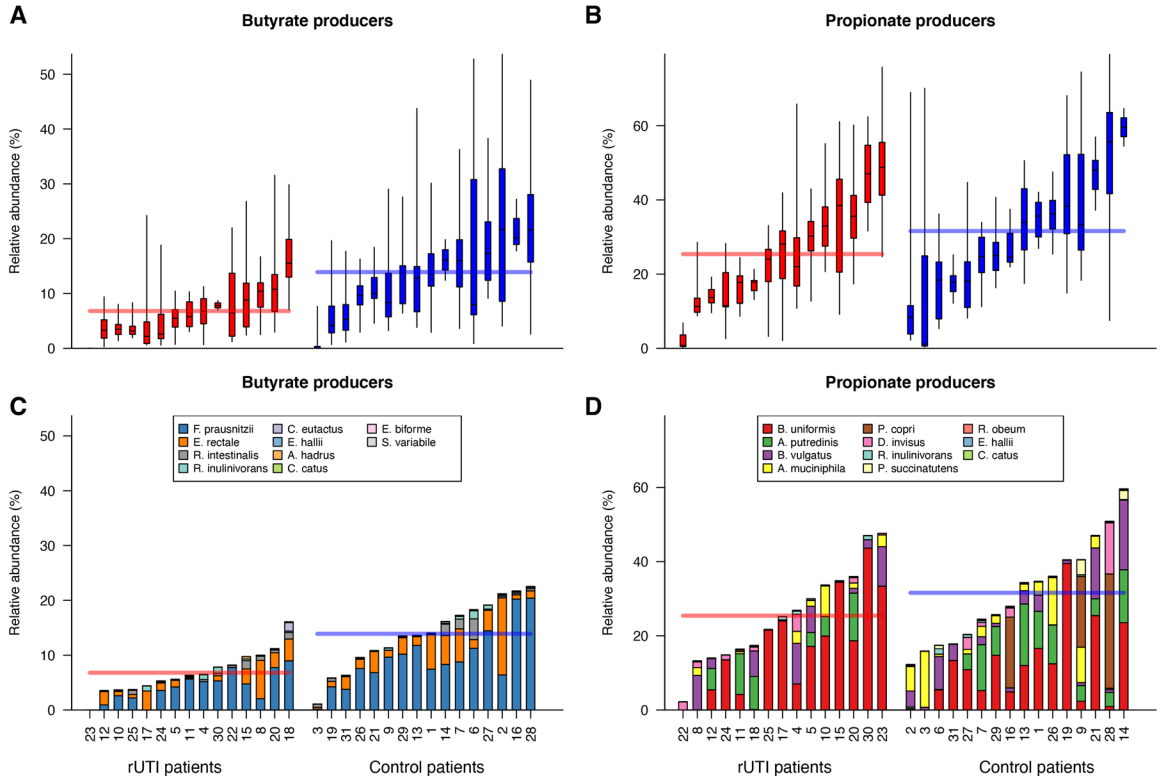
Custom R scripts to analyze outputs are available from github.com/cworby/UMB-study.

Extended Data



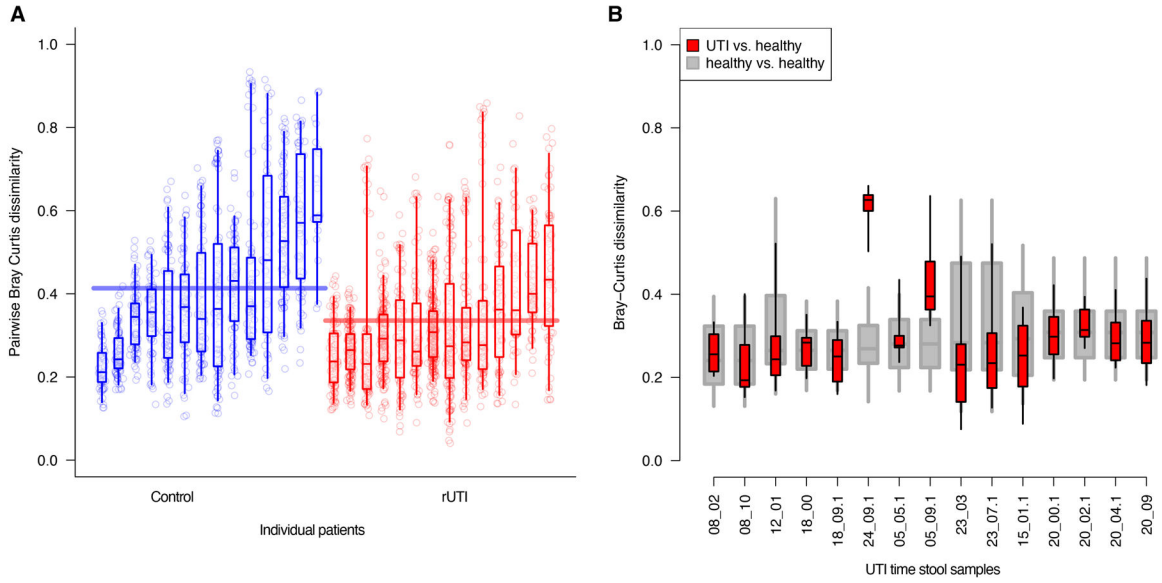
Extended Data Fig. 1. Sex precedes all clinical UTI events

Survey reports of intercourse frequency in the previous two weeks. Responses are partitioned by (i) control women, (ii) rUTI women at time of UTI, and (iii) rUTI women at non-UTI time points.



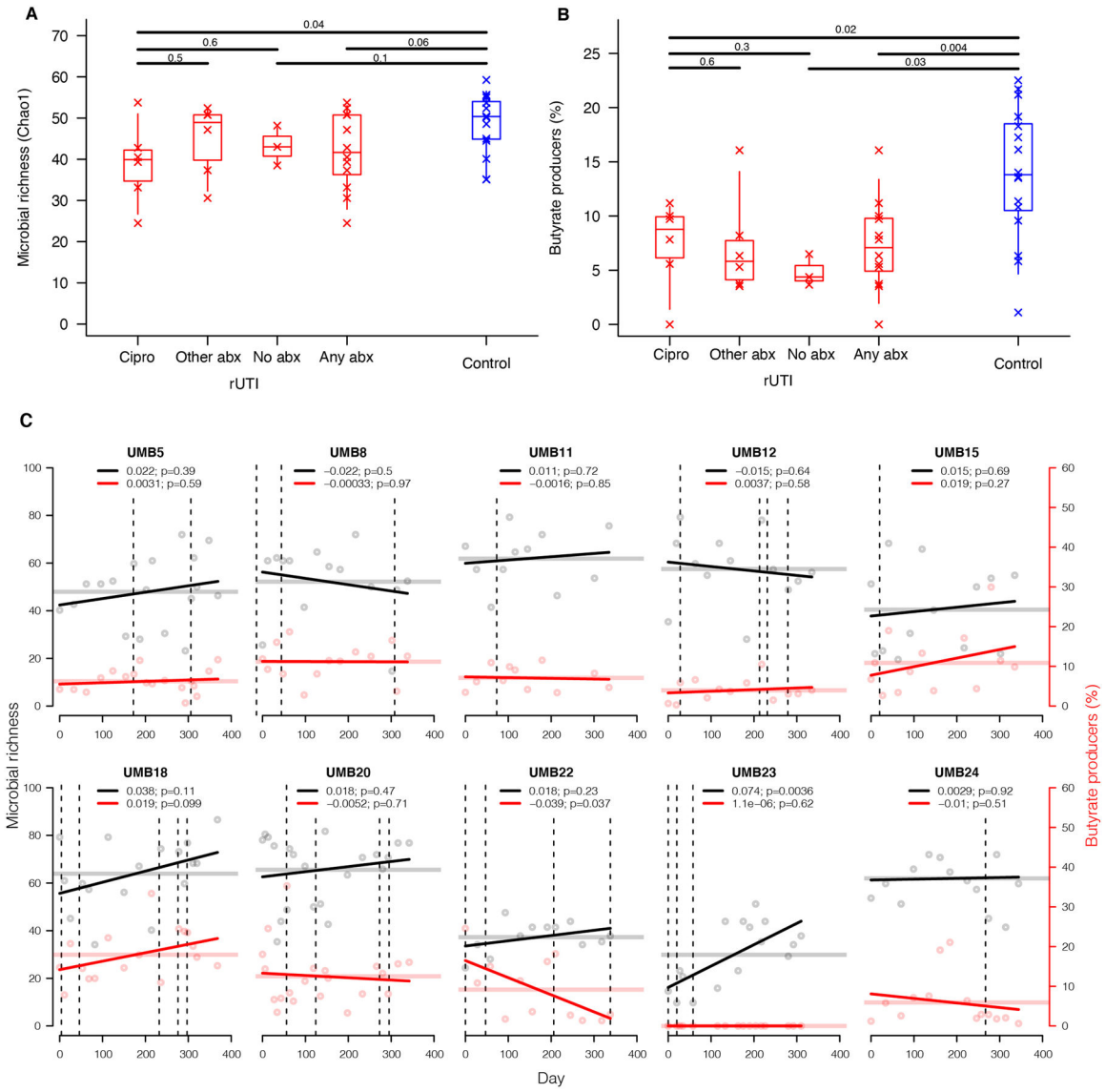
Extended Data Fig. 2. SCFA producing bacteria are depleted in the rUTI gut

Cumulative relative abundances of (a) butyrate and (b) propionate producing bacterial species in rUTI and control samples. Box plots display the median (center line), 25th and 75th percentiles (box), as well as the 5th and 95th percentiles (whiskers). Within-host average relative abundances of individual species for (c) butyrate and (d) propionate producers are also shown. Horizontal lines denote the mean relative abundance in rUTI (red) and control (blue) women.



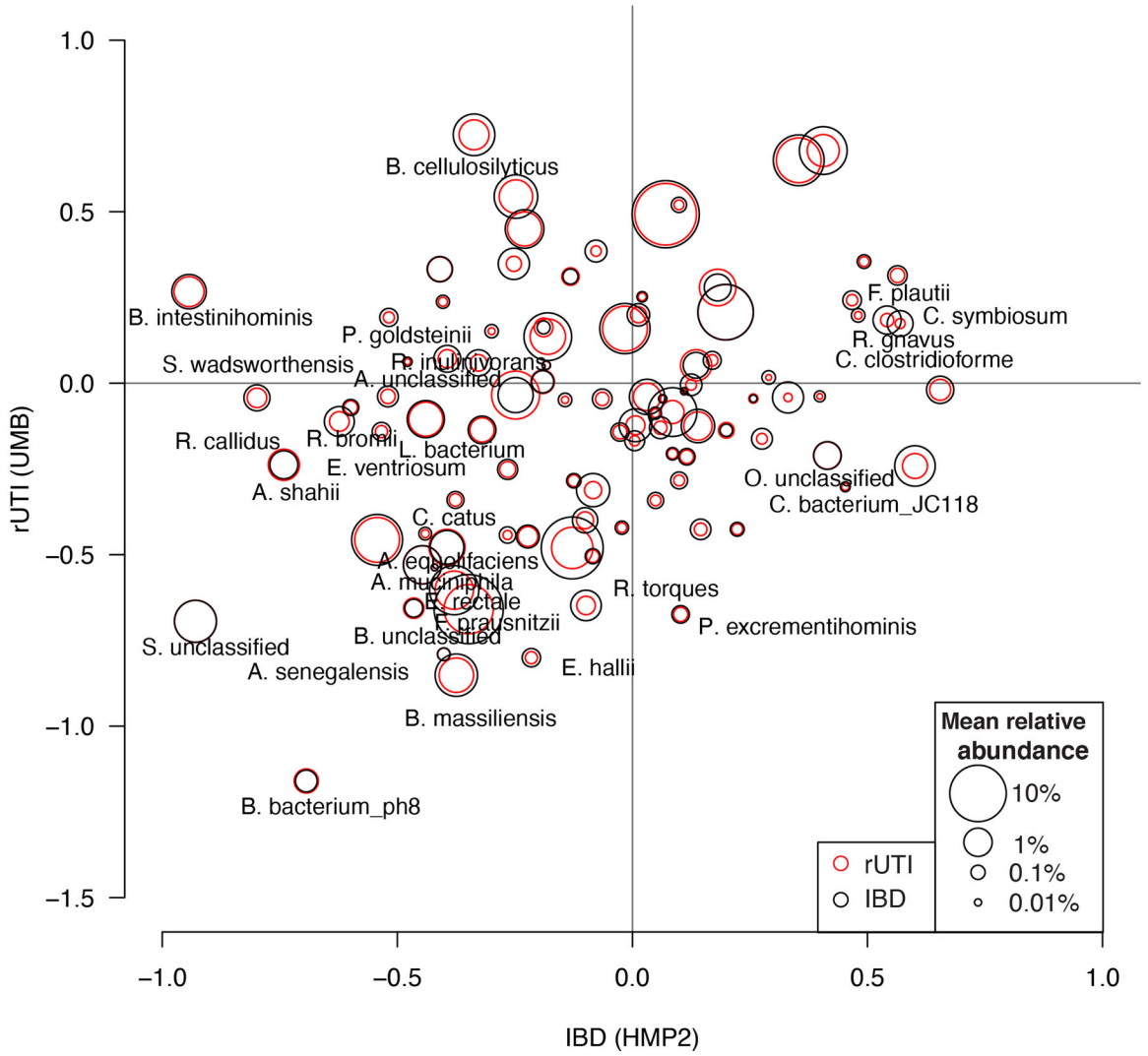
Extended Data Fig. 3. Bray Curtis dissimilarity across stool samples

(a) For each patient, the distribution of Bray-Curtis dissimilarities between all stool samples, ordered by increasing mean patient values within each cohort. (b) Bray-Curtis distributions between samples taken at the time of UTI vs. healthy time points (red), compared to all pairwise healthy sample comparisons. Box plots show the median (center line), 25th and 75th percentiles (box), as well as the 5th and 95th percentiles (whiskers).



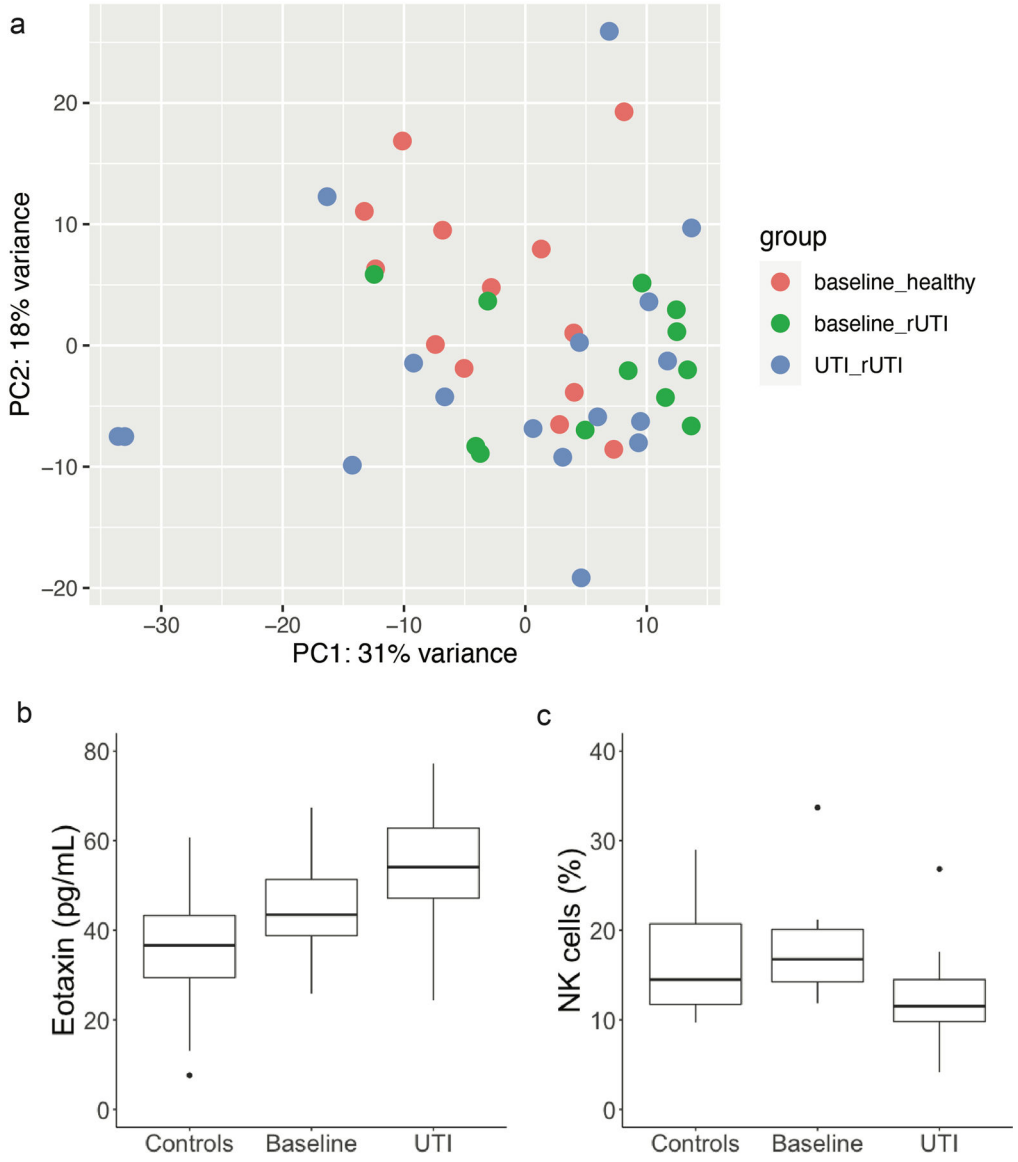
Extended Data Fig. 4. rUTI dysbiosis is not driven by antibiotic use during the study

We grouped rUTI women according to their antibiotic exposures at any point during the UMB study; (i) ciprofloxacin (n=6) (ii) non-ciprofloxacin antibiotics (n=6); (iii) no antibiotics (n=3); (iv) any antibiotics (n=12). Groups were compared against each other and against the control cohort (n=16) for (a) overall microbial richness and (b) relative abundance of butyrate producers. Crosses represent mean values for individuals, boxplots denote the IQR and 95% central quantiles for each group. Wilcoxon rank sum tests (two-sided) were applied to group pairs to derive p-values. (c) Temporal trends of microbial richness (black) and relative abundance of butyrate producers (red) in all rUTI participants using antibiotics during the study. For each individual, linear models were fit to observations (points) over time; fitted trends are shown, with coefficients & p values reported at the top of each panel. Dashed vertical lines denote antibiotic usage. Participant mean values are represented by horizontal lines.



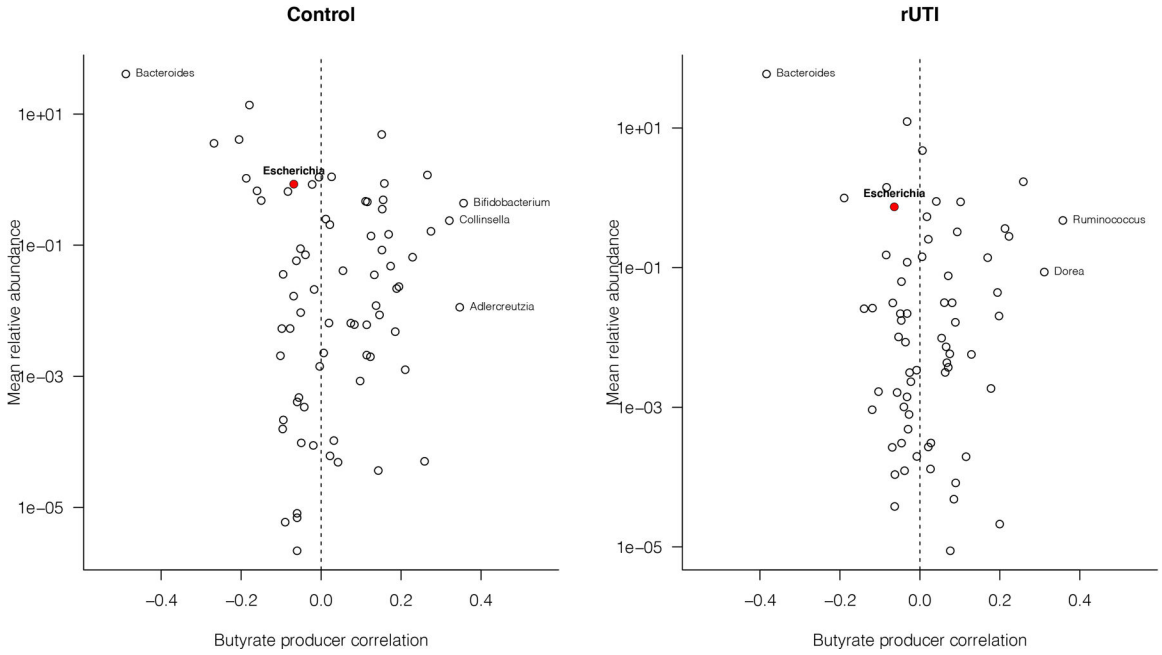
Extended Data Fig. 5. Most species depleted in the rUTI gut are also depleted in the IBD gut

We compared discriminatory taxa in rUTI women to those in IBD patients using data from adult participants in the HMP2 study [33]. For each study, we fitted mixed effects models to standardized Metaphlan2 relative abundances as a function of categorical disease group (rUTI or IBD respectively, vs. each study's control cohort), including covariates for race and antibiotic use. The disease group coefficients are plotted against each other for each species, with circle pairs representing the average relative abundance in each study. Species with uncorrected p values <0.05 in either study are labelled. Species not present in at least 10% of samples in either study are excluded. IBD comprises patients with either CD or UC.



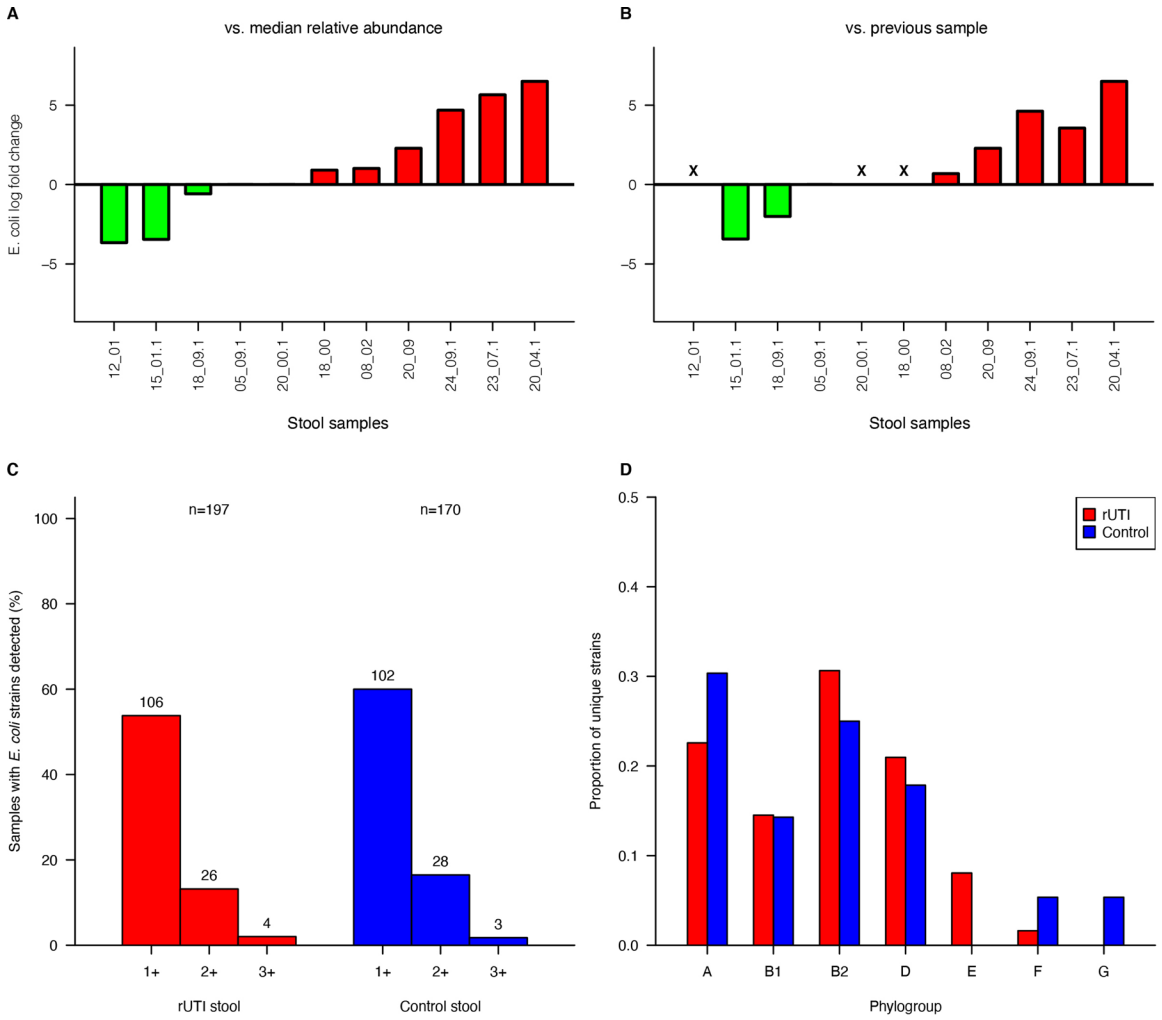
Extended Data Fig. 6. Immunological differences between cohorts

(a) PCA plot of gene expression across cohorts, based on PBMC RNA Seq data. Samples are partitioned into healthy controls (n=13), rUTI patient baseline (enrollment; n=12) and rUTI patient at time of UTI (n=17). (b) Plasma eotaxin-1 levels in control women, and rUTI women at healthy enrollment and time of UTI. (c) Relative abundance of NK cells in control and rUTI women based on CIBERSORT output. Box plots display the median (center line), 25th and 75th percentiles (box), as well as data points within 1.5 IQR of the upper & lower quartiles (whiskers), and outliers beyond this range (dots).



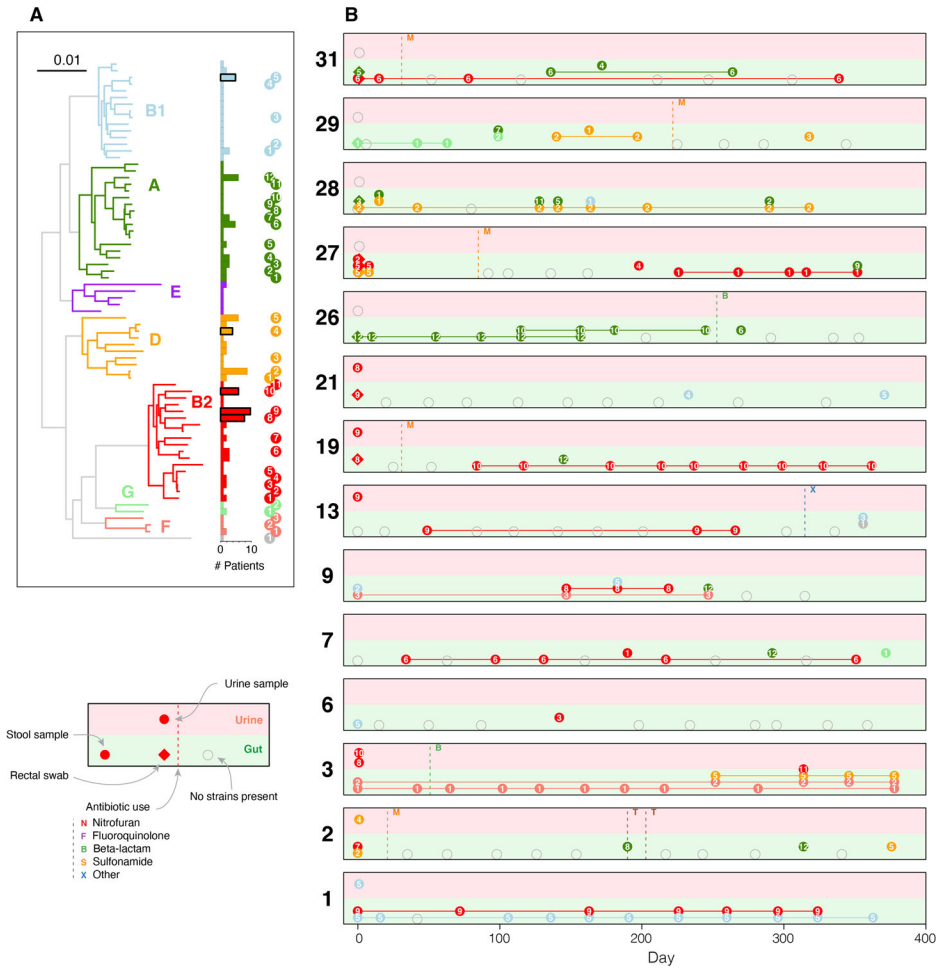
Extended Data Fig. 7. Limited relationship between non SCFA-producing taxa with butyrate producers

For all non SCFA-producing genera detected across all samples, the correlation coefficient between its relative abundance and the relative abundance of butyrate producers was calculated and plotted against its mean relative abundance across (a) control (n=170) and (b) rUTI (n=197) samples. Genera with an absolute correlation coefficient greater than 0.25 are labeled, along with Escherichia, represented by the red point.



Extended Data Fig. 8. *coli* relative abundance around the time of UTI and phylogroup distributions

E. For all stool samples taken within 3 days of a UTI event, the log fold change is given relative to (a) the median *E. coli* relative abundance in the corresponding patient, excluding samples taken at the time of UTI, and (b) the relative abundance of *E. coli* in the preceding stool sample. 'X' denotes samples for which there was no prior sample available. (c) Number of detected *E. coli* strains by sample type. (d) Number of detected StrainGST reference strains vs. relative abundance of *E. coli.*



Extended Data Fig. 9. Strain dynamics in control women

Strain dynamics within all control participants; analogous to Figure 3. (a) Phylogenetic tree comprising strains called by StrainGE across all stool and urine samples, colored by phylogroup. Bars show number of unique participants with at least one strain observation; bars are bolded if the strain was identified in at least one urine sample. Each strain identified in control women is uniquely identifiable by the phylogroup (color) and ID (numeral) indicated right. (b) Each panel represents longitudinal strain dynamics within one patient. Numerals refer to strain identifiers in (a). All fecal strains are connected to their most recent previous observation in fecal samples. Diamonds denote clinical rectal swabs. Strains identified in urine outgrowth depicted if available; otherwise raw urine strains are shown. Fecal or urine samples with no detected *E. coli* strains represented by open grey symbols. Vertical dashed lines represent self-reported antibiotic use.

Extended Data Table 1
Cohort Characteristics

Demographic, behavioral and dietary characteristics of the rUTI and control women who completed the year-long study. Fisher's exact tests (two-sided) were used to compare frequencies between cohorts.

	rUTI (n=14)	Control (n=14)	p
Age (mean, years)	28.6	29.3	0.77
Race=white	12 (86%)	6 (46%)	0.046
No. UTIs during study (mean)	1.6	0	
Intercourse frequency (per week; mean)	2	1.6	0.22
Antibiotic use during study			
Cumulative antibiotic use (doses per patient)	2.6	0.9	0.04
Total doses:			
Nitrofuran (macrobid)	11	0	
Fluoroquinolone (ciprofloxacin)	8	0	
Beta-lactam (incl. amoxicillin, cephalexin)	6	3	
Macrolide (azithromycin)	1	5	
Sulfonamide (bactrim, sulfamethoxazole)	3	0	
Tetracycline (doxycycline)	0	2	
Unspecified	7	2	
Usually consumes*:			
Tea or coffee (no sugar)	11 (79%)	3 (21%)	0.01
Soft drinks, tea/coffee with sugar	10 (71%)	12 (86%)	0.65
Diet soft drinks, tea/coffee with sugar substitute	6 (43%)	2 (14%)	0.21
Fruit juice	9 (64%)	6 (43%)	0.45
Alcohol	11 (79%)	10 (71%)	1
Yoghurt/active bacterial culture	9 (64%)	7 (50%)	0.7
Dairy	14 (100%)	14 (100%)	1
Probiotic (not yoghurt)	3 (21%)	0 (0%)	0.22
Fruit	14 (100%)	14 (100%)	1
Vegetables	14 (100%)	13 (93%)	1
Beans (inc. tofu, soy)	10 (71%)	8 (57%)	0.69
Processed meats	9 (64%)	12 (86%)	0.38
Red meat	9 (64%)	13 (92%)	0.16
White meat	11 (79%)	14 (100%)	0.22
Shellfish	6 (43%)	4 (29%)	0.69
Fish	7 (50%)	8 (57%)	1
NSAIDs	11 (79%)	7 (50%)	0.24

* recorded recent consumption of item in at least 50% of questionnaire responses during study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under grant no. U19AI110818 to the Broad Institute, from the National Institutes of Health Mucosal Immunology Studies Team consortium under grant no. U01AI095542 to Washington University and the National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Department of Health and Human Services, under Grant Number R01DK121822 to the Broad Institute and Washington University. BSO was supported by grants from the National Institutes of Health, USA (T32GM007067 and T32GM139774). This work was also supported by funds from the Center for Women's Infectious Disease Research (cWIDR) at Washington University School of Medicine.

We would like to acknowledge members of the Broad's Bacterial Genomics group and Hera Vlamakis for helpful conversations. We thank Brian Haas for assistance with PBMC RNA-Seq analysis as well as the Multi-Omics Core and Genomics Platform at the Broad Institute for sample processing and data generation.

References

1. Flores-Mireles AL, Walker JN, Caparon M & Hultgren SJ Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13, 269–284 (2015). [PubMed: 25853778]
2. Hooton TM, et al. A prospective study of risk factors for symptomatic urinary tract infection in young women. N Engl J Med 335, 468–474 (1996). [PubMed: 8672152]
3. Yamamoto S, et al. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. J Urol 157, 1127–1129 (1997). [PubMed: 9072556]
4. Nielsen KL, Dynesen P, Larsen P & Frimodt-Moller N Faecal Escherichia coli from patients with E. coli urinary tract infection and healthy controls who have never had a urinary tract infection. J Med Microbiol 63, 582–589 (2014). [PubMed: 24464694]
5. Jantunen ME, Saxen H, Lukinmaa S, Ala-Houhala M & Siitonen A Genomic identity of pyelonephritogenic Escherichia coli isolated from blood, urine and faeces of children with urosepsis. J Med Microbiol 50, 650–652 (2001). [PubMed: 11444777]
6. Magruder M, et al. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. Nat Commun 10, 5521 (2019). [PubMed: 31797927]
7. Thanert R, et al. Comparative Genomics of Antibiotic-Resistant Uropathogens Implicates Three Routes for Recurrence of Urinary Tract Infections. mBio 10(2019).
8. Paalanne N, et al. Intestinal microbiome as a risk factor for urinary tract infections in children. Eur J Clin Microbiol Infect Dis 37, 1881–1891 (2018). [PubMed: 30006660]
9. Magruder M, et al. Gut commensal microbiota and decreased risk for Enterobacteriaceae bacteriuria and urinary tract infection. Gut Microbes 12, 1805281 (2020). [PubMed: 32865119]
10. Tariq R, et al. Fecal Microbiota Transplantation for Recurrent Clostridium difficile Infection Reduces Recurrent Urinary Tract Infection Frequency. Clin Infect Dis 65, 1745–1747 (2017). [PubMed: 29020210]
11. Wang T, Kraft CS, Woodworth MH, Dhere T & Eaton ME Fecal Microbiota Transplant for Refractory Clostridium difficile Infection Interrupts 25-Year History of Recurrent Urinary Tract Infections. Open Forum Infect Dis 5, ofy016 (2018). [PubMed: 29450212]
12. Mayer EA, Tillisch K & Gupta A Gut/brain axis and the microbiota. J Clin Invest 125, 926–938 (2015). [PubMed: 25689247]
13. Cryan JF, et al. The Microbiota-Gut-Brain Axis. Physiol Rev 99, 1877–2013 (2019). [PubMed: 31460832]
14. Budden KF, et al. Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol 15, 55–63 (2017). [PubMed: 27694885]

15. Dang AT & Marsland BJ Microbes, metabolites, and the gut-lung axis. Mucosal Immunol 12, 843–850 (2019). [PubMed: 30976087]
16. Lazar V, et al. Aspects of Gut Microbiota and Immune System Interactions in Infectious Diseases, Immunopathology, and Cancer. Front Immunol 9, 1830 (2018). [PubMed: 30158926]
17. Scholes D, et al. Risk factors for recurrent urinary tract infection in young women. J Infect Dis 182, 1177–1182 (2000). [PubMed: 10979915]
18. Clemente JC, Manasson J & Scher JU The role of the gut microbiome in systemic inflammatory disease. BMJ 360, j5145 (2018). [PubMed: 29311119]
19. Belkaid Y & Hand TW Role of the microbiota in immunity and inflammation. Cell 157, 121–141 (2014). [PubMed: 24679531]
20. Parada Venegas D, et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Front Immunol 10, 277 (2019). [PubMed: 30915065]
21. Liu H, et al. Butyrate: A Double-Edged Sword for Health? Adv Nutr 9, 21–29 (2018). [PubMed: 29438462]
22. Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457–462 (2016). [PubMed: 26476454]
23. Franzosa EA, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods 15, 962–968 (2018). [PubMed: 30377376]
24. Mack A, et al. Changes in gut microbial metagenomic pathways associated with clinical outcomes after the elimination of malabsorbed sugars in an IBS cohort. Gut Microbes 11, 620–631 (2020). [PubMed: 31809634]
25. Palleja A, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol 3, 1255–1265 (2018). [PubMed: 30349083]
26. Zaura E, et al. Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. mBio 6, e01693–01615 (2015). [PubMed: 26556275]
27. Rooney AM, et al. Each Additional Day of Antibiotics Is Associated With Lower Gut Anaerobes in Neonatal Intensive Care Unit Patients. Clin Infect Dis 70, 2553–2560 (2020). [PubMed: 31367771]
28. Schubert AM, et al. Microbiome data distinguish patients with Clostridium difficile infection and non-C. difficile-associated diarrhea from healthy controls. mBio 5, e01021–01014 (2014). [PubMed: 24803517]
29. Pozuelo M, et al. Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome. Sci Rep 5, 12693 (2015). [PubMed: 26239401]
30. Geirnaert A, et al. Butyrate-producing bacteria supplemented in vitro to Crohn's disease patient microbiota increased butyrate production and enhanced intestinal epithelial barrier integrity. Sci Rep 7, 11450 (2017). [PubMed: 28904372]
31. Ni J, Wu GD, Albenberg L & Tomov VT Gut microbiota and IBD: causation or correlation? Nat Rev Gastroenterol Hepatol 14, 573–584 (2017). [PubMed: 28743984]
32. Schaubeck M, et al. Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. Gut 65, 225–237 (2016). [PubMed: 25887379]
33. Integrative Human Microbiome Project Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe 16, 276–289 (2014). [PubMed: 25211071]
34. Zhou Y & Zhi F Lower Level of Bacteroides in the Gut Microbiota Is Associated with Inflammatory Bowel Disease: A Meta-Analysis. Biomed Res Int 2016, 5828959 (2016). [PubMed: 27999802]
35. Duvallet C, Gibbons SM, Gurry T, Irizarry RA & Alm EJ Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 8, 1784 (2017). [PubMed: 29209090]
36. Asnicar F, et al. Blue poo: impact of gut transit time on the gut microbiome using a novel marker. Gut (2021).

37. Takahashi D, et al. Microbiota-derived butyrate limits the autoimmune response by promoting the differentiation of follicular regulatory T cells. EBioMedicine 58, 102913 (2020). [PubMed: 32711255]
38. Rosser EC, et al. Microbiota-Derived Metabolites Suppress Arthritis by Amplifying Aryl-Hydrocarbon Receptor Activation in Regulatory B Cells. Cell Metab 31, 837–851 e810 (2020). [PubMed: 32213346]
39. Li F, Wang M, Wang J, Li R & Zhang Y Alterations to the Gut Microbiota and Their Correlation With Inflammatory Factors in Chronic Kidney Disease. Front Cell Infect Microbiol 9, 206 (2019). [PubMed: 31245306]
40. Adar T, Shteingart S, Ben Ya'acov A, Bar-Gil Shitrit A & Goldin E From airway inflammation to inflammatory bowel disease: eotaxin-1, a key regulator of intestinal inflammation. Clin Immunol 153, 199–208 (2014). [PubMed: 24786916]
41. Adar T, et al. The Importance of Intestinal Eotaxin-1 in Inflammatory Bowel Disease: New Insights and Possible Therapeutic Implications. Dig Dis Sci 61, 1915–1924 (2016). [PubMed: 26874691]
42. Cheung W, Bluth M, Khan S, Johns C & Bluth M Peripheral blood mononuclear cell gene array profiles in female patients with involuntary bladder contractions. Advances in Genomics and Genetics 1, 3–7 (2011).
43. de Santiago PR, et al. Immune-related IncRNA LINC00944 responds to variations in ADAR1 levels and it is associated with breast cancer prognosis. Life Sci 268, 118956 (2021). [PubMed: 33383047]
44. Gur C, et al. Natural killer cell-mediated host defense against uropathogenic E. coli is counteracted by bacterial hemolysinA-dependent killing of NK cells. Cell Host Microbe 14, 664–674 (2013). [PubMed: 24331464]
45. Rivera-Chavez F, et al. Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. Cell Host Microbe 19, 443–454 (2016). [PubMed: 27078066]
46. Antharam VC, et al. Intestinal dysbiosis and depletion of butyrogenic bacteria in Clostridium difficile infection and nosocomial diarrhea. J Clin Microbiol 51, 2884–2892 (2013). [PubMed: 23804381]
47. van Dijk L, et al. StrainGE: A toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol 23, 74 (2022). [PubMed: 35255937]
48. Clermont O, Bonacorsi S & Bingen E Rapid and simple determination of the Escherichia coli phylogenetic group. Appl Environ Microbiol 66, 4555–4558 (2000). [PubMed: 11010916]
49. Schreiber H.L.t., et al. Bacterial virulence phenotypes of Escherichia coli and host susceptibility determine risk for urinary tract infections. Sci Transl Med 9(2017).
50. Garretto A, et al. Genomic Survey of E. coli From the Bladders of Women With and Without Lower Urinary Tract Symptoms. Front Microbiol 11, 2094 (2020). [PubMed: 33013764]
51. Zhang S, et al. Short Chain Fatty Acids Modulate the Growth and Virulence of Pathosymbiont Escherichia coli and Host Response. Antibiotics (Basel) 9(2020).
52. Stapleton AE The Vaginal Microbiota and Urinary Tract Infection. Microbiol Spectr 4(2016).
53. Forde BM, et al. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary tract infection. Nat Commun 10, 3643 (2019). [PubMed: 31409795]
54. Cusumano CK, et al. Treatment and prevention of urinary tract infection with orally active FimH inhibitors. Sci Transl Med 3, 109ra115 (2011).
55. Dethlefsen L & Relman DA, Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A. 108 Suppl 1, 4554–61 (2011). [PubMed: 20847294]
56. Turnbaugh PJ, et al. , An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–31 (2006). [PubMed: 17183312]
57. Truong DT, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12, 902–903 (2015). [PubMed: 26418763]
58. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

59. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014). [PubMed: 24227677]

60. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010). [PubMed: 19910308]

61. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12, 453–457 (2015). [PubMed: 25822800]

62. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]

63. Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477 (2012). [PubMed: 22506599]

64. Lloyd-Price J, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019). [PubMed: 31142855]

65. Louis P & Flint HJ Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol 19, 29–41 (2017). [PubMed: 27928878]

66. Levi I, et al. Potential role of indolelactate and butyrate in multiple sclerosis revealed by integrated microbiome-metabolome analysis. Cell Rep Med 2, 100246 (2021). [PubMed: 33948576]

67. Alcock BP, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48, D517–D525 (2020). [PubMed: 31665441]

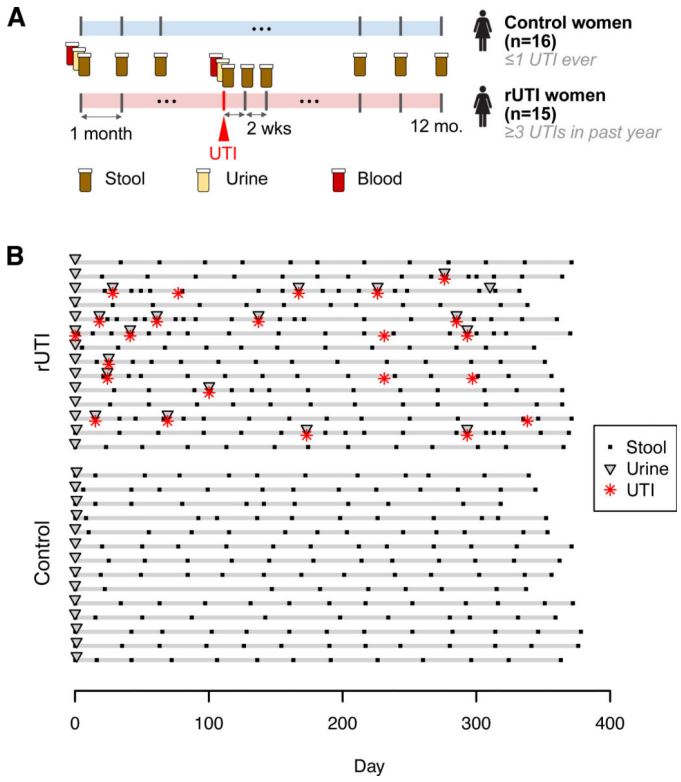


Figure 1. Study design and sample collection for the UMB study.

(a) Stool samples were collected monthly from rUTI and control patients. Stool, urine and blood plasma samples were collected upon enrollment and subsequent UTI clinic visits. Biweekly stool samples were requested following UTI diagnoses. (b) Stool and urine samples collected from all rUTI and control participants (excluding one rUTI and two control participants who dropped out of the study prior to completion). Each participant's enrollment timeline is represented by horizontal gray lines, with stool (black dots) and urine (triangles) sample collection times denoted. Red symbols denote diagnosed and inferred UTI events.

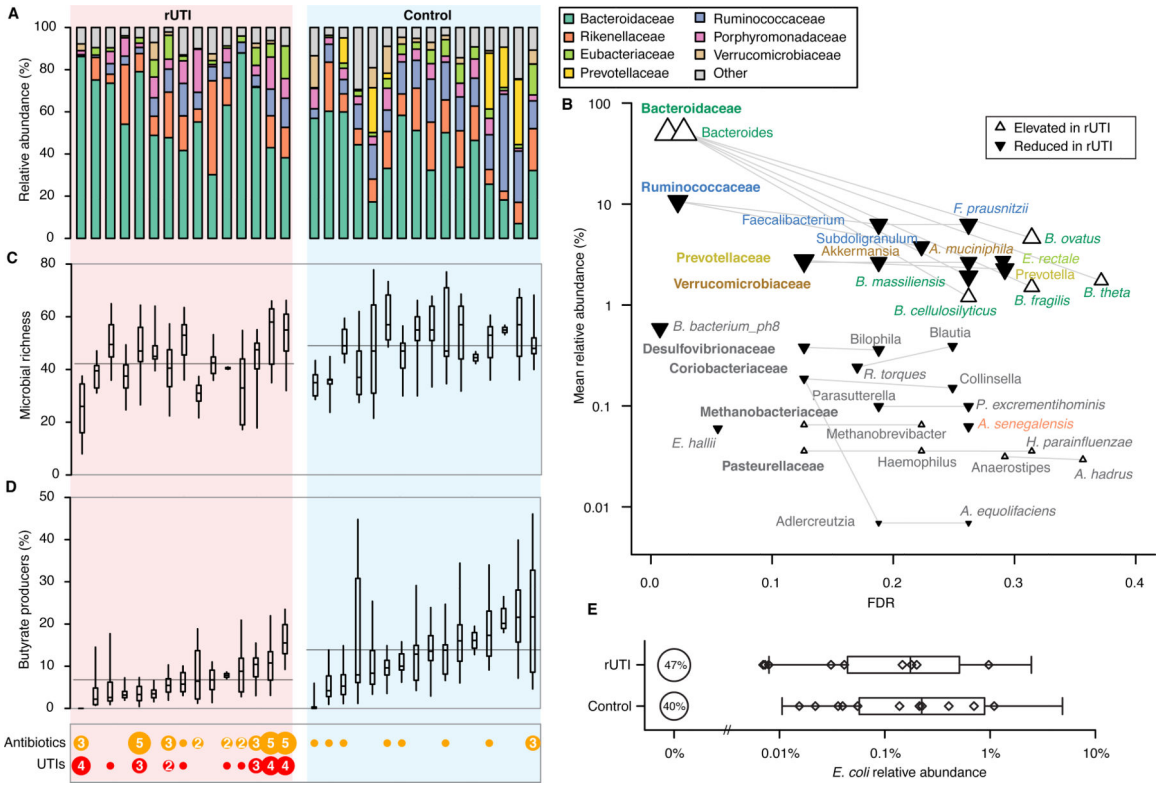


Figure 2. rUTI women have a distinct gut microbiome.

(a) Average relative abundances of bacterial families for each patient in the rUTI (left) and the control (right) cohorts. (b) Significance and rUTI group effect size for selected taxa. Each point represents one taxon; its effect size and direction (symbol) for rUTI vs. control, false discovery rate (FDR) and mean relative abundance across all samples. Taxonomic relationships are represented by lines. FDR values calculated independently at each taxonomic level. Bold, regular and italic text denote family, genus and species levels (c) Microbial richness distributions and (d) cumulative relative abundance of butyrate-producing species for each study participant. Plot displays the median (center line), 25th and 75th percentiles (box), as well as the 5th and 95th percentiles (whiskers) for each individual. Horizontal lines represent group-level mean of individual means. Antibiotic use and UTI occurrence for each study participant is shown at the bottom left; symbol size and numerals denote the number of UTIs/reported antibiotic courses. (e) Relative abundance of *E. coli* in each cohort. Symbols denote median relative abundances of individual patients; box plots display the median (center line), 25th and 75th percentiles (box), as well as the 5th and 95th percentiles (whiskers) of non-zero values.

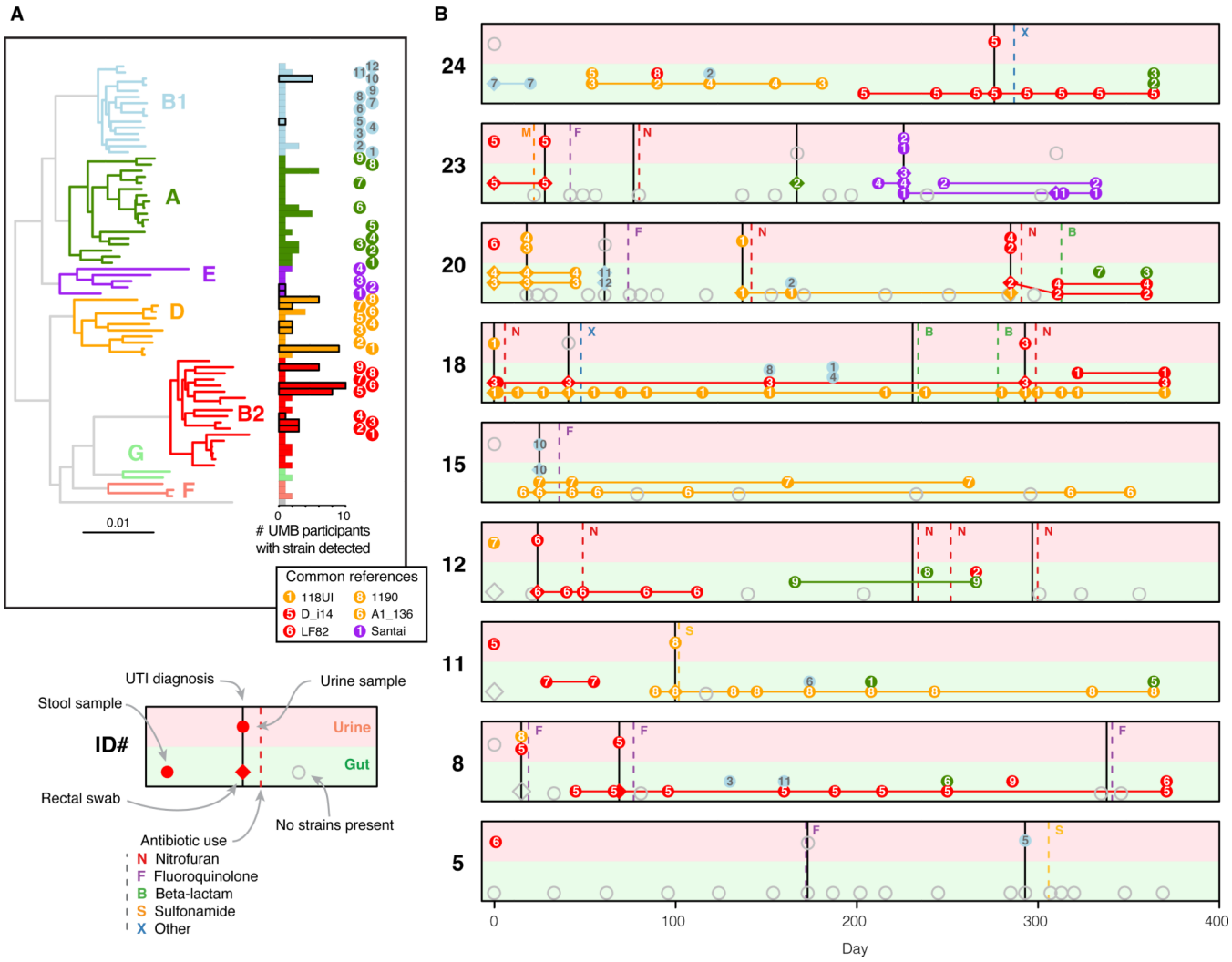


Figure 3. Frequent gut-bladder transmission and strain persistence in rUTI patients.

Strain dynamics within all participants with *E. coli* UTIs. (a) Phylogenetic tree comprising strains called by StrainGE across all stool and urine samples, colored by phylogroup. Bars show number of unique participants with at least one strain observation; bars are bolded if the strain was identified in at least one urine sample. Each strain identified in rUTI patients is uniquely identifiable by the phylogroup (color) and ID (numeral) indicated right. (b) Each panel represents longitudinal strain dynamics within one patient. Numerals refer to strain identifiers in (a). All fecal strains are connected to their most recent previous observation in fecal samples. Diamonds denote clinical rectal swabs. Strains identified in urine outgrowth depicted if available; otherwise raw urine strains are shown. Fecal or urine samples with no detected *E. coli* strains represented by open grey symbols. Vertical dashed lines represent self-reported antibiotic use, solid black lines denote UTI events.

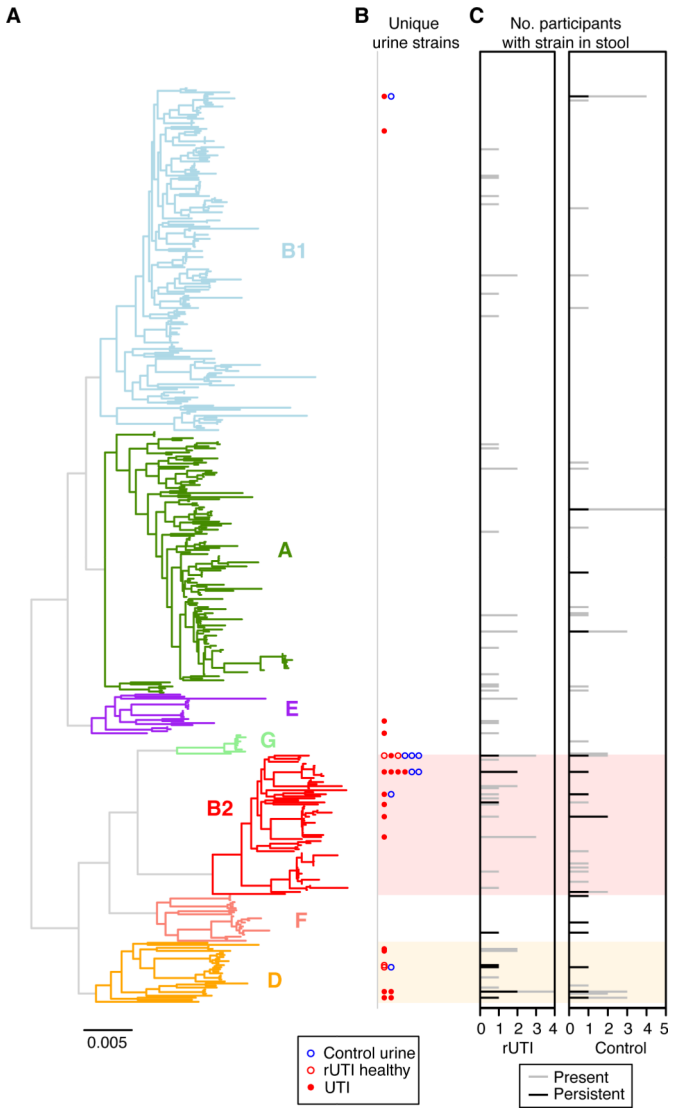


Figure 4. Phylogenetic distribution of *E. coli* strains identified in all stool and urine samples.

(a) The phylogenetic tree of StrainGE reference strains colored and annotated by phylogroup. (b) Unique *E. coli* strains identified in urine samples are marked alongside the corresponding reference strain. Filled circles represent UTI-causing strains, blue circles denote strains identified in control hosts. (c) The total number of rUTI (left) and control (right) women with the corresponding strain present in stool samples. Black bars denote the number of women for whom the strain was persistent in the gut.