Check for updates

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 ✉] and Ashlee M. Earl [1 ✉]

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 rUTI, from whom we collected urine, blood and monthly faecal samples for metagenomic and transcriptomic interrogation. During the study 24 UTIs were reported, with 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 with 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.

U rinary tract infections (UTI) are among the most common bacterial infections worldwide and a major cause of morbidity in females, with uropathogenic *E. coli* (UPEC) being the primary causative agent[1]. One of the strongest risk factors for UTI is a history of previous UTIs[2], but the biological basis and risk factors for long-term recurrence remain unclear in otherwise healthy women. Between 20 and 30% of women diagnosed with a UTI will experience recurrent UTIs (rUTIs), 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 with healthy controls. Furthermore, faecal microbiota transplants used 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 (1) gut dysbiosis is associated with rUTI susceptibility, (2) women with rUTI have unique uropathogen dynamics within and between the gut and the bladder or (3) 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 (1) compared with 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; (2) differential immunological biomarkers suggest that rUTI women may have a distinct immune state; (3) *E. coli* strains were transmitted from the gut to the bladder in both cohorts, although no UTI symptoms occurred in healthy controls; and (4) 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 controls and 15 women with rUTI participated in the year-long study, providing both monthly home-collected stool samples and blood, urine and rectal swabs at enrolment and subsequent clinic visits for UTI treatment (Fig. 1a). Participants completed monthly questionnaires on diet, symptoms and behaviour (Supplementary 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 behavioural differences were apparent (Extended Data Table 1).

A total of 24 UTIs occurred during the study, all in women with rUTI who each had experienced between no and four UTIs (Fig. 1b). Nineteen were diagnosed by clinicians while 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 behavioural 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 2 weeks (Extended Data Fig. 1).

Urine samples collected at the time of clinical UTI diagnoses were plated on MacConkey agar, with bacterial growth detected (>0 colony-forming units (CFU) ml⁻¹) in 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 being dominated by *Klebsiella pneumoniae*. Sequencing of uncultured urine from the remaining UTI samples identified uropathogens in a further four samples, including *E. coli* (two), *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 is 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 patients having undergone kidney transplant (reduced *Faecalibacterium* and *Romboutsia*[9]) and UTIs in children (reduced
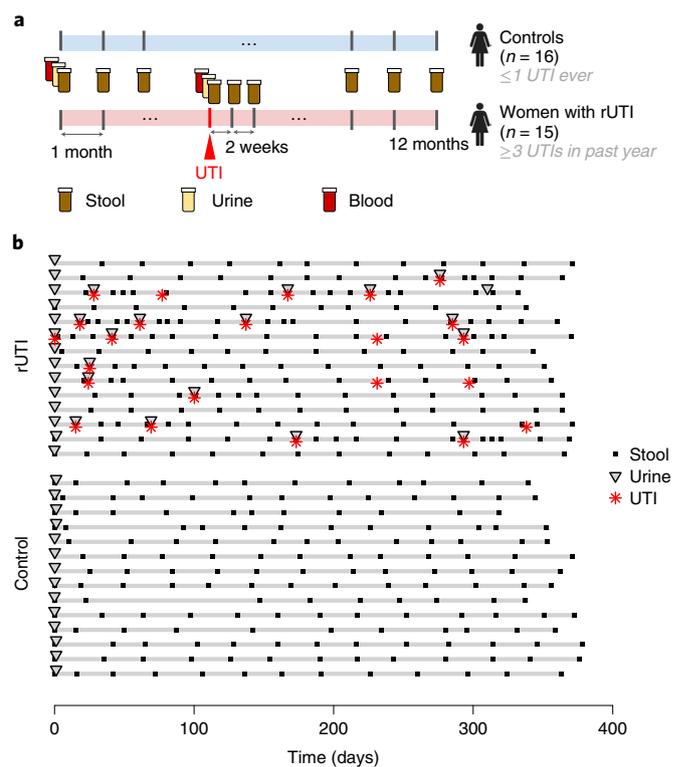

Fig. 1 | Study design and sample collection for the UMB study. a, Stool samples were collected monthly from both patients with rUTI and controls. Stool, urine and blood plasma samples were collected upon enrolment and subsequent UTI clinic visits. Biweekly stool samples were requested following UTI diagnoses. **b**, Stool and urine samples were collected from all participants (excluding one with rUTI and two control participants who dropped out of the study before completion). Each participant's enrolment timeline is represented by horizontal grey lines, with stool (black dots) and urine (triangles) sample collection times denoted. Red symbols denote diagnosed and inferred UTI events.

Peptostreptococcaceae[8]), it is unclear whether these are risk factors for recurrence in otherwise healthy adult women. To explore this, we sequenced and analysed the metagenomes of 367 longitudinal stool samples from both women with rUTI (*n* = 197) and controls (*n* = 170) (Fig. 1b and Methods). Rectal swabs, collected during clinic visits, were not used to determine microbiome profiles.

There were broad differences in gut microbiota composition between cohorts (Fig. 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 women with rUTI (*P* = 0.05; Fig. 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 women with rUTI. We identified 22 differentially abundant taxa (FDR < 0.25) at lower taxonomic levels, 16 of which were depleted in women with rUTI (Supplementary Table 2 and Fig. 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 the promotion of 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,
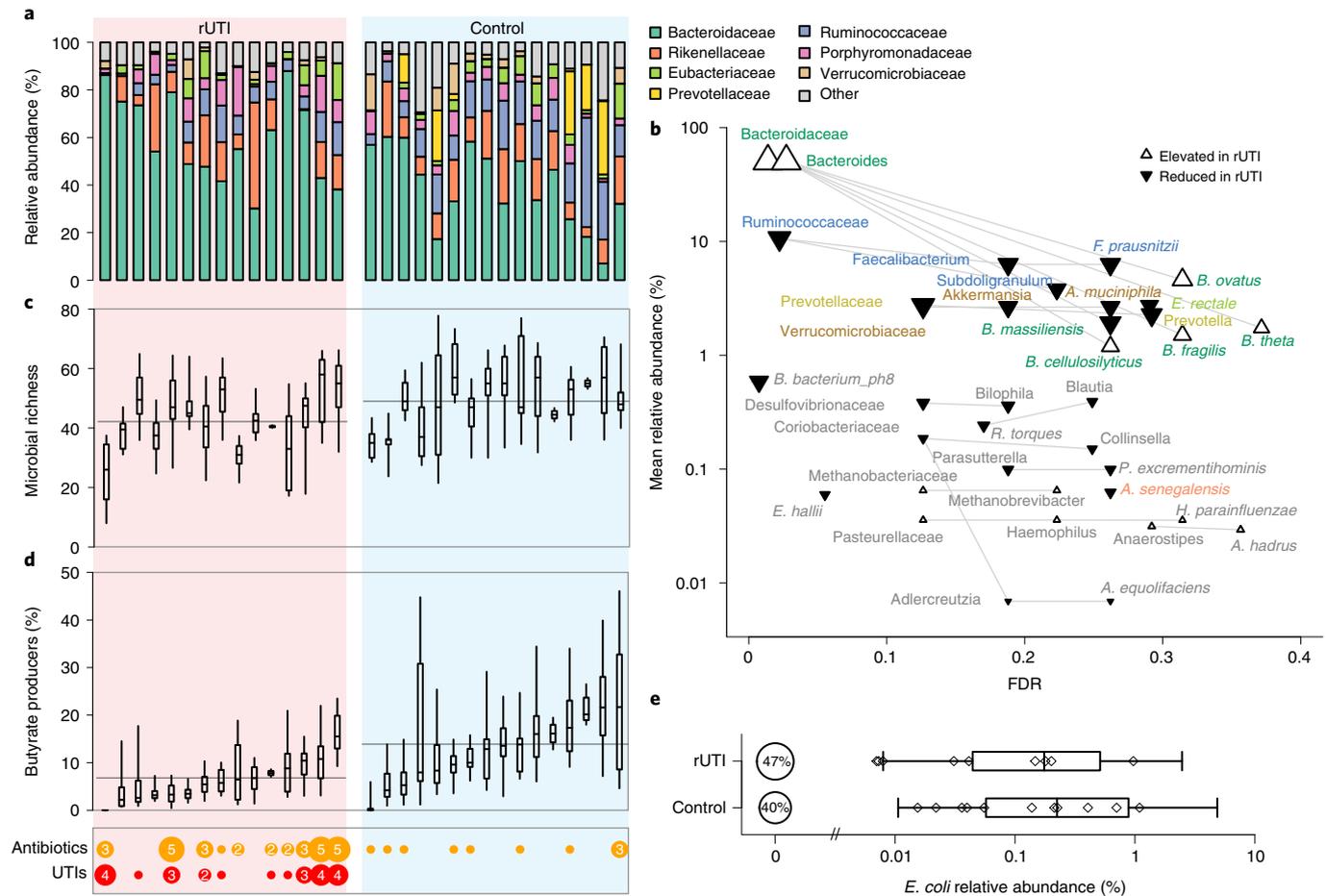
Fig. 2 | Women with rUTI have a distinct gut microbiome. a, Average relative abundances of bacterial families for each patient in the rUTI (left) and control (right) cohorts. **b**, Significance and rUTI group effect size for selected taxa. Each point represents one taxon: its effect size and direction (symbol) for rUTI versus control, FDR and mean relative abundance across all samples. Taxonomic relationships are represented by lines. FDR values were calculated independently at each taxonomic level. Bold, regular and italic text denote family, genus and species levels, respectively. **c,d**, Microbial richness distributions (**c**) and cumulative relative abundance (**d**) of butyrate-producing species for each study participant. Plots display the median (centre line), 25th and 75th percentiles (box) and 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 are shown at 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 (centre line), 25th and 75th percentiles (box) and 5th and 95th percentiles (whiskers) of non-zero values.

particularly butyrate producers, were significantly less abundant in women with rUTI ($P = 0.001$; Fig. 2d and Extended Data Fig. 2). Four Kyoto Encyclopedia of Genes and Genomes orthogroups[22] representing components of butyrate production pathways were significantly reduced across the rUTI cohort (Supplementary Table 3). Functional analysis with HUMAnN2 (ref. [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 patients with irritable bowel syndrome (IBS) involving 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 one-third of UTIs in our study. Thus, we sought to determine whether antibiotic effects might contribute to the observed shifts in microbiome composition in women with rUTI ('rUTI dysbiosis'). Although antibiotic exposure in the previous 2 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 (Fig. 2c), and no differences in overall gut microbiome stability between cohorts, despite more frequent antibiotic treatment among women with UTI (Extended Data Fig. 3). We observed no differences in either richness or the abundance of butyrate producers between women with rUTI under different antibiotic exposures (Extended Data Fig. 4a,b). Within the rUTI group, the frequency of infections was associated with neither microbial richness nor the relative abundance of butyrate producers. The microbial richness of women suffering UTIs during the study did not differ significantly from that of women with rUTI not reporting infections ($P = 0.4$; Fig. 2b–d). While we did not detect a lasting impact from individual antibiotic courses—there were few long-term trends among women with rUTI over the study (Extended Data Fig. 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 inflammatory bowel disease. The depletion of butyrate-producing taxa and

microbial richness, key characteristics of rUTI dysbiosis, is also observed in other gut inflammatory conditions, including nosocomial diarrhoea[28], IBS[29] and inflammatory bowel disease (IBD)[20], particularly Crohn's disease[30], and thus may be indicative of gut inflammation in women with rUTI. 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 *Faecalibacterium 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 Fig. 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 Fig. 5), and was also decreased among patients with IBD in other studies[34]. *E. coli* was significantly elevated in patients with IBD in the HMP2 study, but showed no difference in average relative abundance between our cohorts (Fig. 2e). Diminished *Bacteroides* alongside elevated Enterobacteriaceae was also observed in patients with nosocomial diarrhoea[28]. Diarrhoea, 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, women with rUTI and low-level inflammation and no diarrhoea may lack the depletion of *Bacteroides* and elevation of Enterobacteriaceae observed in diarrhoea-associated conditions. It is also possible that the considerable differences in treatment regimens—that is, antibiotics versus 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 that 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 levels of interleukin-10, 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 enrolment and UTI, quantifying (1) a Luminex panel of human cytokines, chemokines and growth factors involved in inflammation and T cell activation, and (2) 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 women with rUTI versus controls at enrolment, 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; however, since this result did not hold after adjusting for race, we could not rule out potential demographic confounders. Eotaxin-1 was also higher in the blood plasma of women with rUTI at the time of UTI versus enrolment ($P=0.04$; Extended Data Fig. 6b).

Our small cohort size provided limited statistical power to identify differential expression between cohorts based on PBMC RNA sequencing (RNA-seq) data, and no large-scale differences were observed (Extended Data Fig. 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 previously been 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 tumour-invading T lymphocytes in breast cancer and markers for programmed cell death[43]. Resting natural killer (NK) cells were significantly reduced at the time of UTI relative to baseline levels ($P=0.02$; Extended Data Fig. 6c). NK cells help suppress bladder infection by UPEC in mice[44], so the loss of these cells in the periphery may suggest migration to the bladder concurrent with rUTI.

Gut and bladder *E. coli* dynamics are similar between cohorts. Previous work has implicated gut dysbiosis and 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 (Fig. 2e), suggesting that the rUTI dysbiotic gut is no more hospitable to *E. coli* colonization than that of 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 Fig. 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 Fig. 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 >tenfold higher than intrahost mean) occurred in the 2 weeks before 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.

Although 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 (Fig. 3) and control (Extended Data Fig. 9) cohorts. Both the number of strains per sample and phylogroup distribution were comparable between cohorts (Fig. 4 and Extended Data Fig. 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 controls (odds ratio (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 enrolment urine samples yielding sufficient bacterial DNA for sequencing and profiling (Supplementary Table 6), we detected *E. coli* strains in 54% (13/24),
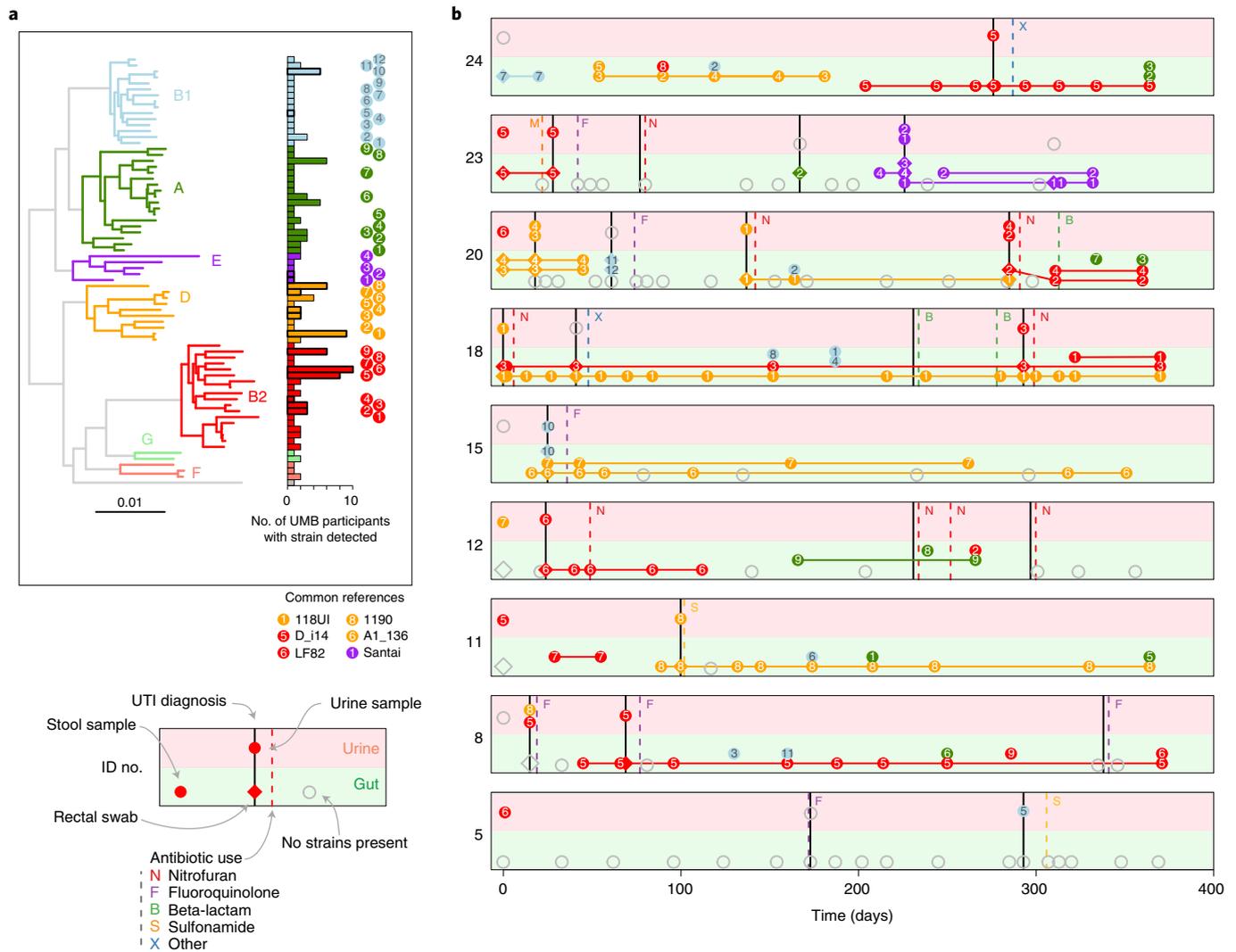
Fig. 3 | Frequent gut–bladder transmission and strain persistence in patients with rUTI. a,b, Strain dynamics within all participants with *E. coli* UTIs. **a**, Phylogenetic tree comprising strains called by StrainGE across all stool and urine samples, coloured by phylogroup. Bars show number of unique 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 patients with rUTI is uniquely identifiable by phylogroup (colour) and ID (numeral), indicated to the right. **b**, Each panel represents longitudinal strain dynamics within one patient. Numerals refer to strain identifiers in **a**. All faecal strains are connected to their most recent previous observation in faecal samples. Diamonds denote clinical rectal swabs. Strains identified in urine outgrowth depicted if available; otherwise raw urine strains are shown. Faecal or urine samples with no detected *E. coli* strains are represented by open grey symbols. Vertical dashed lines represent self-reported antibiotic use, solid black lines denote UTI events.

including over half of samples (7/13) from control participants, despite the absence of symptoms. All but one of these strains also belonged to phylogroup B2 or D. Control urine samples carried *E. coli* strains that were phylogenetically similar to UTI-causing strains based on StrainGE predictions (Fig. 4 and Methods), despite divergent clinical outcomes.

Mapping of 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 and Supplementary Table 7). This transition of a probable urovirulent strain to the bladder of healthy women without eliciting UTI symptoms is consistent with previous studies that were unable to identify genetic markers of urovirulence in mice[49], or to 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 downregulation 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, the longitudinal dynamics of these strains within the gut are less well understood despite the importance of such insights into development of 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
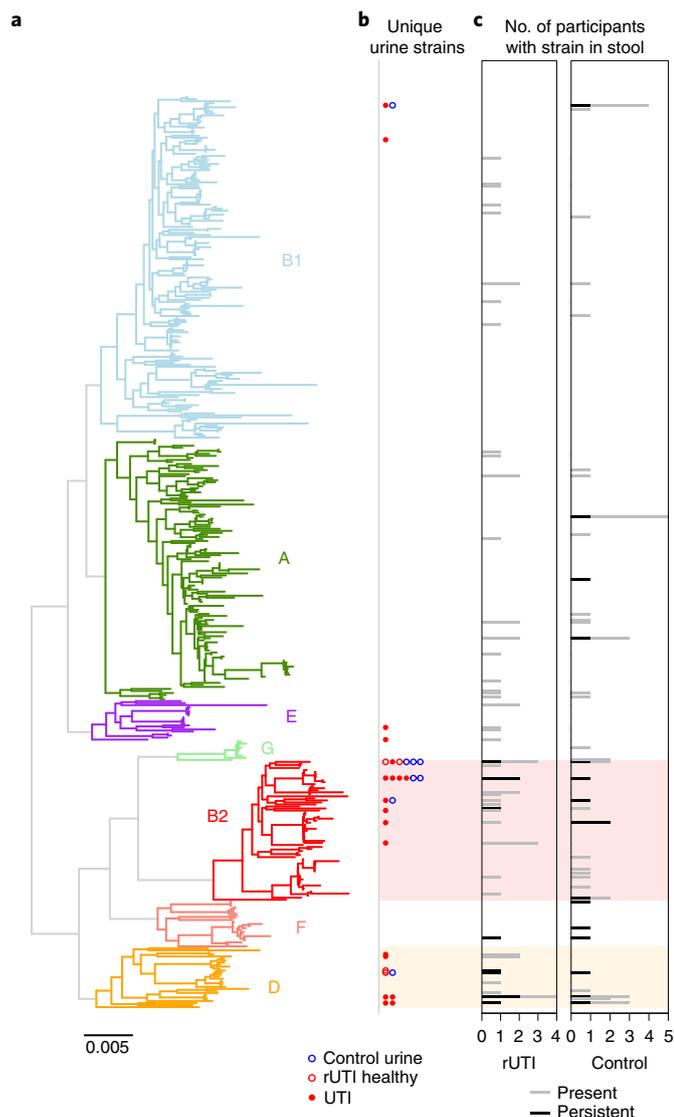
Fig. 4 | Phylogenetic distribution of *E. coli* strains identified in all stool and urine samples. a, Phylogenetic tree of StrainGE reference strains, coloured and annotated by phylogroup. **b**, Unique *E. coli* strains identified in urine samples are marked alongside the corresponding reference strain. Filled circles represent UTI-causing strains, blue circles denote strains identified in control hosts. **c**, Total number of women with rUTI (left) and controls (right) with the corresponding strain present in stool samples. Black bars denote the number of women for whom the strain was persistent in the gut.

exposure. Four women with rUTI suffered multiple confirmed *E. coli* UTIs, though only one was a same-strain recurrence (individual no. 8; Fig. 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 that some UTIs may be caused by minority strains. Only one UTI (individual no. 5; Fig. 3) was caused by a strain never observed in another sample from that individual. This phylogroup B1 strain probably 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 2 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 rapid adaptation by UPEC populations to repeated antibiotic exposure and persistence 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 both healthy women and those with rUTI. 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 from 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 rUTI. Notably, *E. coli* was commonly identified in the urine of healthy women, including strains arising from UPEC-associated clades and harbouring 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 has a number of limitations. First, due to the limited collection of urine samples in the control group, it was not possible to robustly compare (1) the composition of the urine microbiome and (2) the frequency of (asymptomatic) strain transfer from gut to bladder between cohorts. Second, 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. Third, 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 that 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 identification of 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 (for example, faecal microbiota transplants[10]) to limit infections via restoration of a healthy bacterial community in the gut.

Methods

Study design and sample collection. *Enrolment.* 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 (1) had IBD or urological developmental defects (for example, ureteral reflux, kidney agenesis and so on), (2) were pregnant, (3) were taking antibiotics as prophylaxis for rUTI or (4) were younger than 18 years or older than 45 years at the time of enrolment. All participants provided informed consent. Microbiological information for previous UTIs was not available. A total of 16 control and 15 women with rUTI and aged between 18 and 45 years were recruited to the study; participants were remunerated with gift cards for participation. Fourteen 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 predetermine sample sizes, but our sample sizes are similar to those reported in previous publications (for example, refs. [55,56]). As an observational study with no intervention, and with cohort membership based on the 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 and 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 h) after the clinic visit, as well as at a 2-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 HMP2 (ref. [33]). Briefly, participants collected a fresh faecal sample in a disposable toilet hat and then aliquoted two teaspoon-sized scoops of stool into one tube containing PBS and another containing 100% ethanol. Samples were kept overnight at the Broad Institute, where they were stored at −80 °C until sample processing. All stool samples were shipped Monday to Thursday within each week to limit their 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, frequency of sexual intercourse and UTI symptoms. Participants who did not provide stool samples and questionnaires at the beginning of each month were given a 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 enrolment and UTI visits. Blood was stored on ice for <30 min and then mixed with an equal amount PBS and 2% foetal bovine serum (FBS). 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 −80 °C until use. PBMCs were washed with PBS plus 2% FBS and pelleted via centrifugation at 10,000*g* at room temperature for 5 min. PBMC cell pellets were then flash-frozen and stored at −80 °C until RNA extraction.

Rectal swab and urine preparation. Rectal swabs were collected in the clinic and stored on ice for <30 min followed by washing in 2 ml of PBS, then 1 ml of PBS was centrifuged at 10,000*g* at room temperature for 2 min and PBS supernatant removed. The bacterial/faecal pellet was then flash-frozen and stored at −80 °C

until DNA extraction. The remaining 1 ml was then used to make serial dilutions and plated on both Luria broth (LB) and MacConkey agar and incubated overnight at 37 °C to quantify CFU. 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 min, 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 <30 min; 10 ml of urine was centrifuged at 10,000*g* at room temperature for 5 min. The resulting pellet was washed in PBS, pelleted again and then flash-frozen and stored at −80 °C until DNA extraction. One millilitre of urine was used to make serial dilutions, plated onto both LB and MacConkey and incubated overnight at 37 °C to enumerate CFU. After outgrowth, plates were scraped to collect bacterial colonies, which were then washed with PBS, pelleted at 10,000*g* at room temperature for 2 min, flash-frozen and stored at −80 °C until DNA extraction.

RNA extraction: PBMCs. RNA was extracted from stored PBMCs using TRIzol Reagent (nos. 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 5 min at room temperature. Chloroform was added to samples at the recommended concentration followed by incubation with shaking for 15 s, and samples were rested for 2–3 min at room temperature. After incubation, samples were centrifuged at 12,000*g* for 15 min at 4 °C. The aqueous phase was collected for RNA isolation. RNA was precipitated using 100% isopropanol and incubated at room temperature for 10 min, followed by centrifugation at 12,000*g* for 10 min at 4 °C. Precipitated RNA was washed according to a protocol using 75% ethanol and resuspended in RNase-free water. Extracted RNA was stored at −80 °C until further use.

DNA extraction: rectal swabs and 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 μl of Nuclei Lysis solution, incubated at 80 °C for 5 min then cooled to room temperature. RNase solution was added to samples and incubated for 15 min at 37 °C, then cooled to room temperature; 200 μl of Protein Precipitation solution was added to the RNase-treated sample, vortexed for 20 s and incubated on ice for 5 min. After incubation, samples were centrifuged for 3 min at 16,000*g* and the supernatant transferred to a 1.5-ml microcentrifuge tube containing 600 μl of isopropanol. Samples were gently mixed and centrifuged for 2 min at 16,000*g*. Supernatant was removed and the DNA pellet washed with 70% ethanol. Samples were centrifuged for 2 min at 16,000*g*, ethanol was aspirated and DNA pellets were air-dried for 15 min. DNA pellets were rehydrated with DNA Rehydration solution and incubated at 65 °C for 1 h. Extracted DNA was stored at 4 C for short-term storage or 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.

Whole-metagenome sequencing and sequence data processing. Libraries were constructed from extracted DNA from stool, urine, rectal swabs and plate scrapes using the NexteraXT kit (Illumina). Libraries were then sequenced on a HiSeq 2500 (Illumina) in 101-base pair (bp) paired-end read mode and/or a HiSeq X10 (Illumina) in 151-bp paired-end read mode. Sequence data were then demultiplexed. Samples sequenced multiple times on different runs were pooled. Reads were processed with KneadData (v.0.7.2, https://huttenhower.sph.harvard.edu/kneaddata/) to remove adaptor sequences and trim low base qualities (with Trimmomatic), as well as to remove human-derived sequences (by alignment to the human genome with bowtie2).

Luminex assays. The Custom Luminex magnetic bead assay kit was obtained from R&D systems (no. 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 and MMP-3. Detection of analytes in human plasma samples was performed using the Curiox DropArray system for miniaturization of magnetic bead multiplex kits. Plasma samples were diluted twofold for the assay. Results were read and quantified using a BioPlex multiplex plate reader and Microplate Manager software (v.5).

Eotaxin ELISA. Plasma eotaxin (CCL11) levels from patients with rUTI and controls were measured using the Eotaxin (CCL11) Human Simple Step ELISA kit (Abcam, no. ab185985), according to the manufacturer's protocol. Briefly, plasma samples were diluted in sample diluent and 50 μl each of sample and antibody

cocktail were added to 96-well plate strips. Plates were sealed and incubated with shaking for 1 h at room temperature. Wells were washed three times with 1× wash buffer and inverted to remove excess liquid; 100 µl of TMB substrate was added to each well, then plates were covered to protect from light and incubated with shaking for 10 min. Stop solution (100 µl) was added to each well and plates were incubated with shaking for 1 min. Optical density (OD$_{450}$) was measured and recorded to determine the concentration of Eotaxin (in pg ml^{-1}).

Sequence data analysis. *Community profiling and metrics.* Bacterial community composition was determined using MetaPhlAn2 (v.2.7.0 with db v20)[57] on KneadData-processed sequences. Functional profiling was performed using HUMAnN2 (v.2.8.1, database downloaded October 2016)[23] on KneadData-processed sequences. Diversity metrics and Bray–Curtis (BC) distances were derived from MetaPhlAn2 relative abundance output using the *vegan* package in R (https://cran.r-project.org/web/packages/vegan/).

PBMC RNA-seq analysis. Sequences from PBMC-extracted messenger RNA 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 to counts per million 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 was used to remove three samples contaminated with ≥5% non-PBMC cell types. The CIBERSORT-filtered set of samples was used to perform differential gene expression analysis using DESeq2 (ref. [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. 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 van Dijk et al.[47]. Strains mapping to the same representative reference genome in this database typically have an average nucleotide identity (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 were 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 or uncultured urine otherwise. Species were determined to be uropathogens based on UTI prevalence studies (for example, ref. [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 (*E. coli* strain profiling) 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. Although VFDB contains T6SS genes, we removed these in favour 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 and models. *rUTI risk factors.* We used questionnaire responses to determine whether any dietary or behavioural factors were associated with rUTI. We first compared the proportion of participants in each cohort who responded positively to binary variables (for example, dairy or alcohol consumption and so on in the previous 2 weeks) in >50% or responses, and used 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 behavioural response variable collected in the questionnaire. Variables with limited or no variance were excluded.

Identification of 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 (for example, ref. [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 time points, we also fit models including covariates for 'pre-UTI' (14 days preceding UTI diagnosis), 'time of UTI' (3 days either side of UTI diagnosis) or 'post antibiotics' (<14 days post antibiotic exposure) as binary variables. All taxa with >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 BC dissimilarities between individuals. Since women with rUTI had undergone, on average, slightly more frequent sampling than controls, due to 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 with 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 plotted in Extended Data Fig. 5.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Metagenomic sequence data are available from the Sequence Read Archive under Bioproject PRJNA400628. PBMC RNA-seq data are available from the database of Genotypes and Phenotypes (dbGaP) under project no. phs002728. Questionnaire data and output files from MetaPhlan2, Humann2 and StrainGE are available from github.com/cworby/UMB-study. Source data are provided with this paper.

Code availability

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

References

1. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
2. Hooton, T. M. et al. A prospective study of risk factors for symptomatic urinary tract infection in young women. *N. Engl. J. Med.* **335**, 468–474 (1996).
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).
4. Nielsen, K. L., 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).
5. Jantunen, M. E., 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).
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).
7. Thänert, R. et al. Comparative genomics of antibiotic-resistant uropathogens implicates three routes for recurrence of urinary tract infections. *mBio* **10**, e01977-19 (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).

9. Magruder, M. et al. Gut commensal microbiota and decreased risk for Enterobacteriaceae bacteriuria and urinary tract infection. *Gut Microbes* **12**, 1805281 (2020).

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).

11. Wang, M., Kraft, C. S., Woodworth, M. H., Dhere, T. & Eaton, M. E. Fecal microbiota transplant for refractory *Clostridium difficile* infection interrupts 25-year history of recurrent urinary tract infections. *Open Forum Infect. Dis.* **5**, ofy016 (2018).

12. Mayer, E. A., Tillisch, K. & Gupta, A. Gut/brain axis and the microbiota. *J. Clin. Invest.* **125**, 926–938 (2015).

13. Cryan, J. F. et al. The microbiota-gut-brain axis. *Physiol. Rev.* **99**, 1877–2013 (2019).

14. Budden, K. F. et al. Emerging pathogenic links between microbiota and the gut-lung axis. *Nat. Rev. Microbiol.* **15**, 55–63 (2017).

15. Dang, A. T. & Marsland, B. J. Microbes, metabolites, and the gut-lung axis. *Mucosal Immunol.* **12**, 843–850 (2019).

16. Lazar, V. et al. Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. *Front. Immunol.* **9**, 1830 (2018).

17. Scholes, D. et al. Risk factors for recurrent urinary tract infection in young women. *J. Infect. Dis.* **182**, 1177–1182 (2000).

18. Clemente, J. C., Manasson, J. & Scher, J. U. The role of the gut microbiome in systemic inflammatory disease. *Brit. Med. J.* **360**, j5145 (2018).

19. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).

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).

21. Liu, H. et al. Butyrate: a double-edged sword for health? *Adv. Nutr.* **9**, 21–29 (2018).

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–D462 (2016).

23. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).

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).

25. Palleja, A. et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **3**, 1255–1265 (2018).

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**, e01693-15 (2015).

27. Rooney, A. M. 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).

28. Schubert, A. M. et al. Microbiome data distinguish patients with *Clostridium difficile* infection and non-*C. difficile*-associated diarrhea from healthy controls. *mBio* **5**, e01021-14 (2014).

29. Pozuelo, M. et al. Reduction of butyrate- and methane-producing microorganisms in patients with irritable bowel syndrome. *Sci. Rep.* **5**, 12693 (2015).

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).

31. Ni, J., Wu, G. D., Albenberg, L. & Tomov, V. T. Gut microbiota and IBD: causation or correlation? *Nat. Rev. Gastroenterol. Hepatol.* **14**, 573–584 (2017).

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).

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).

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).

35. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).

36. Asnicar, F. et al. Blue poo: impact of gut transit time on the gut microbiome using a novel marker. *Gut* **70**, 1665–1674 (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).

38. Rosser, E. C. et al. Microbiota-derived metabolites suppress arthritis by amplifying aryl-hydrocarbon receptor activation in regulatory B cells. *Cell Metab.* **31**, 837–851(2020).

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).

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).

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).

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. *Adv. Genomics Genet.* **1**, 3–7 (2011).

43. de Santiago, P. R. et al. Immune-related lncRNA LINC00944 responds to variations in ADAR1 levels and it is associated with breast cancer prognosis. *Life Sci.* **268**, 118956 (2021).

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).

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).

46. Antharam, V. C. et al. Intestinal dysbiosis and depletion of butyrogenic bacteria in *Clostridium difficile* infection and nosocomial diarrhea. *J. Clin. Microbiol.* **51**, 2884–2892 (2013).

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).

48. Clermont, O., Bonacorsi, S. & Bingen, E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**, 4555–4558 (2000).

49. Schreiber, H. L. et al. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. *Sci. Transl. Med.* **9**, eaaf1283 (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).

51. Zhang, S. et al. Short chain fatty acids modulate the growth and virulence of pathosymbiont *Escherichia coli* and host response. *Antibiotics (Basel)* **9**, 462 (2020).

52. Stapleton, A. E. The vaginal microbiota and urinary tract infection. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.uti-0025-2016 (2016).

53. Forde, B. M. et al. Population dynamics of an *Escherichia coli* ST131 lineage during recurrent urinary tract infection. *Nat. Commun.* **10**, 3643 (2019).

54. Cusumano, C. K. et al. Treatment and prevention of urinary tract infection with orally active FimH inhibitors. *Sci. Transl. Med.* **3**, 109ra115 (2011).

55. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108**, 4554–4561 (2011).

56. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

57. Truong, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).

58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

60. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

61. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

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).

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

65. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota. *Environ. Microbiol.* **19**, 29–41 (2017).

Acknowledgements

Author contributions

Study design was undertaken by H.L.S., K.W.D., S.J.H. and A.M.E. Study coordination was carried out by H.L.S., K.B., S.B.C. and A.K. Experiments were performed by H.L.S., J.S.P., C.L.P.O., V.L.M. and A.E.P. Data analysis was undertaken by C.J.W., H.L.S., T.J.S., L.R.v.D., R.A.B., B.S.O., B.J.H., C.A.D. and W.-C.C. Consultation and supervision of analyses were the responsibility of B.J.W., A.L.M., T.J.H., T.M.H., A.L.K., H.H.L., K.W.D., S.J.H. and A.M.E. C.J.W., A.L.M., K.W.D., S.J.H. and A.M.E. prepared the original draft. Review and approval of the final manuscript was provided by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01107-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01107-x.
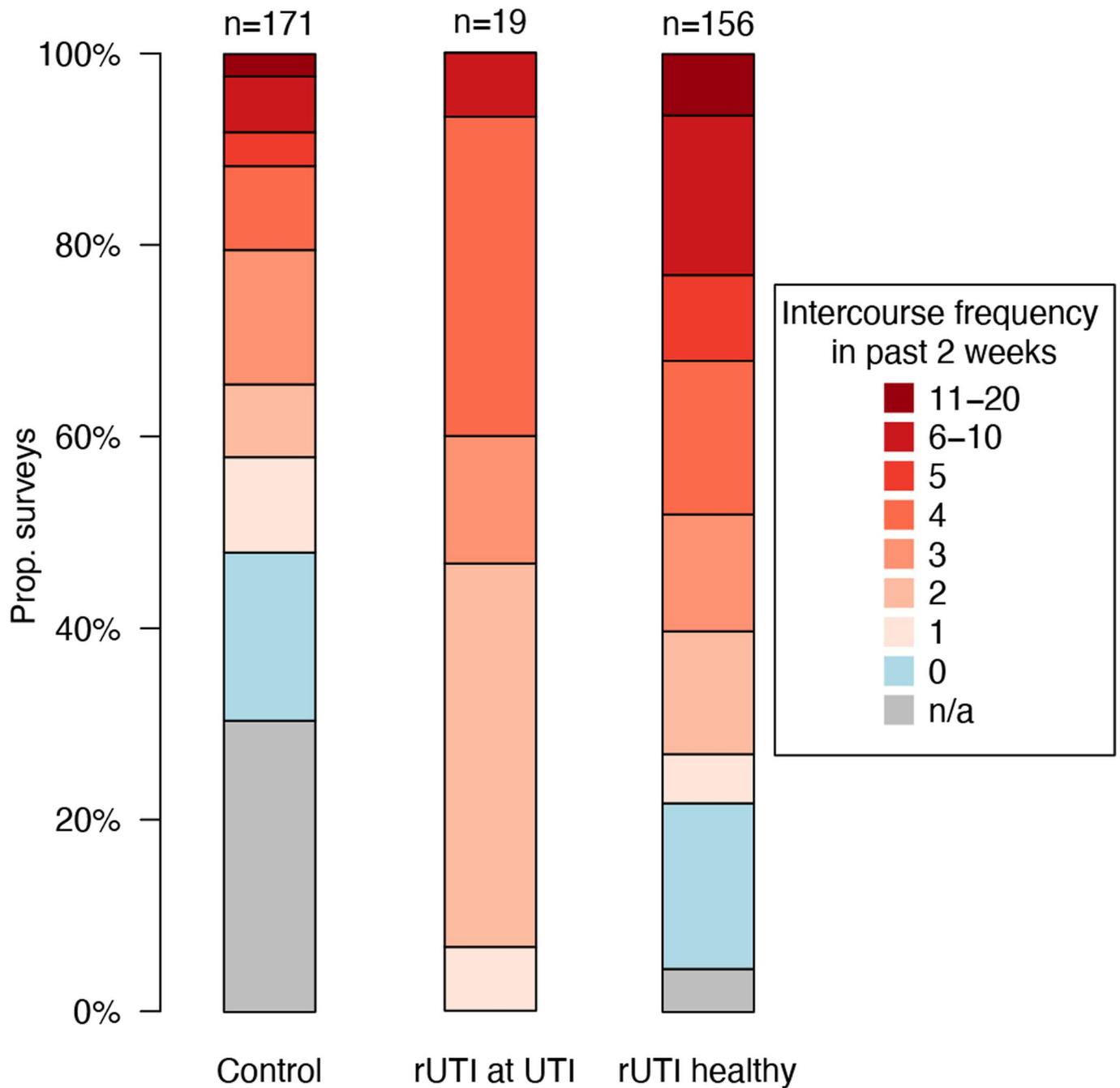
Correspondence and requests for materials should be addressed to Scott J. Hultgren or Ashlee M. Earl.

Peer review information *Nature Microbiology* thanks John Lee, Alice McHardy and Mark Schembri for their contribution to the peer review of this work.
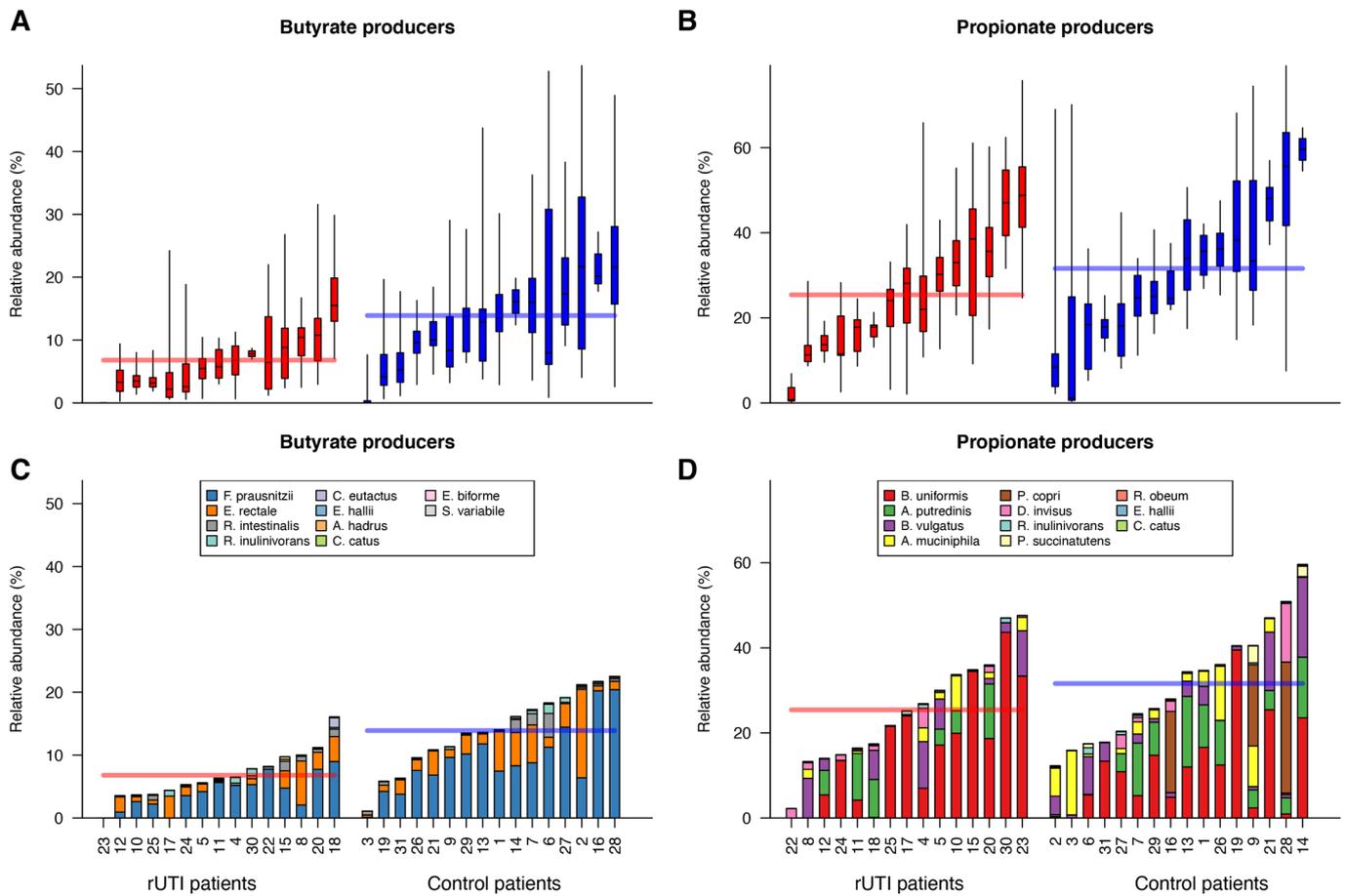
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
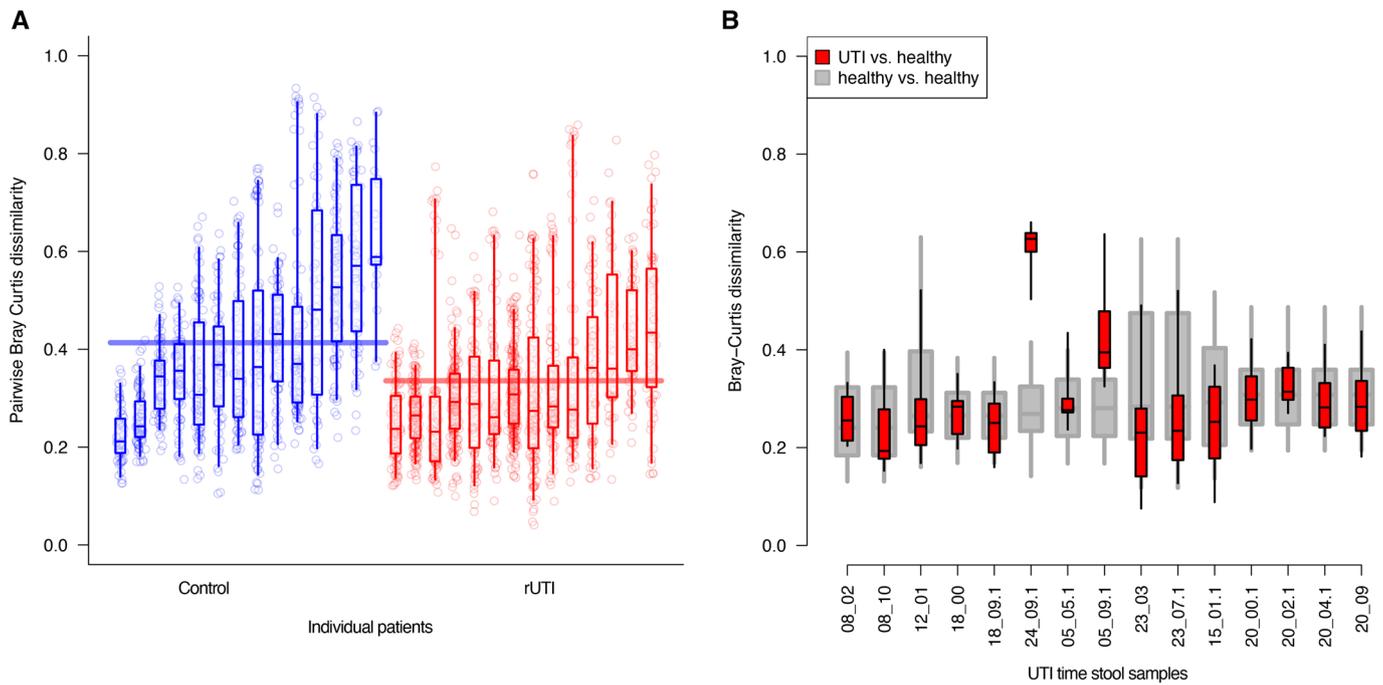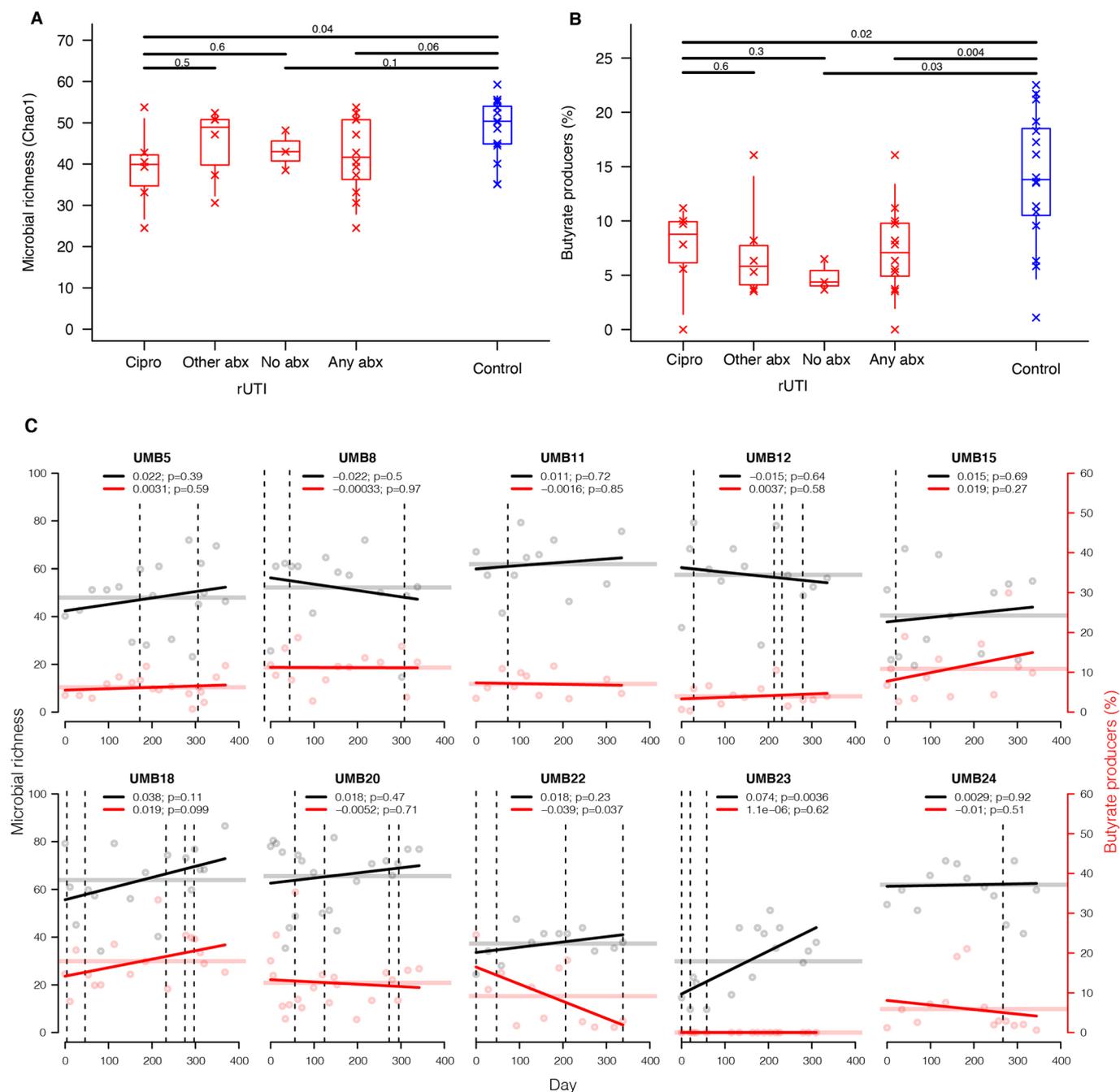
Extended Data Fig. 1 | Sex precedes all clinical UTI events. Survey reports of intercourse frequency in the previous two weeks. Responses are partitioned by (i) control women, (ii) rUTI women at time of UTI, and (iii) rUTI women at non-UTI time points.
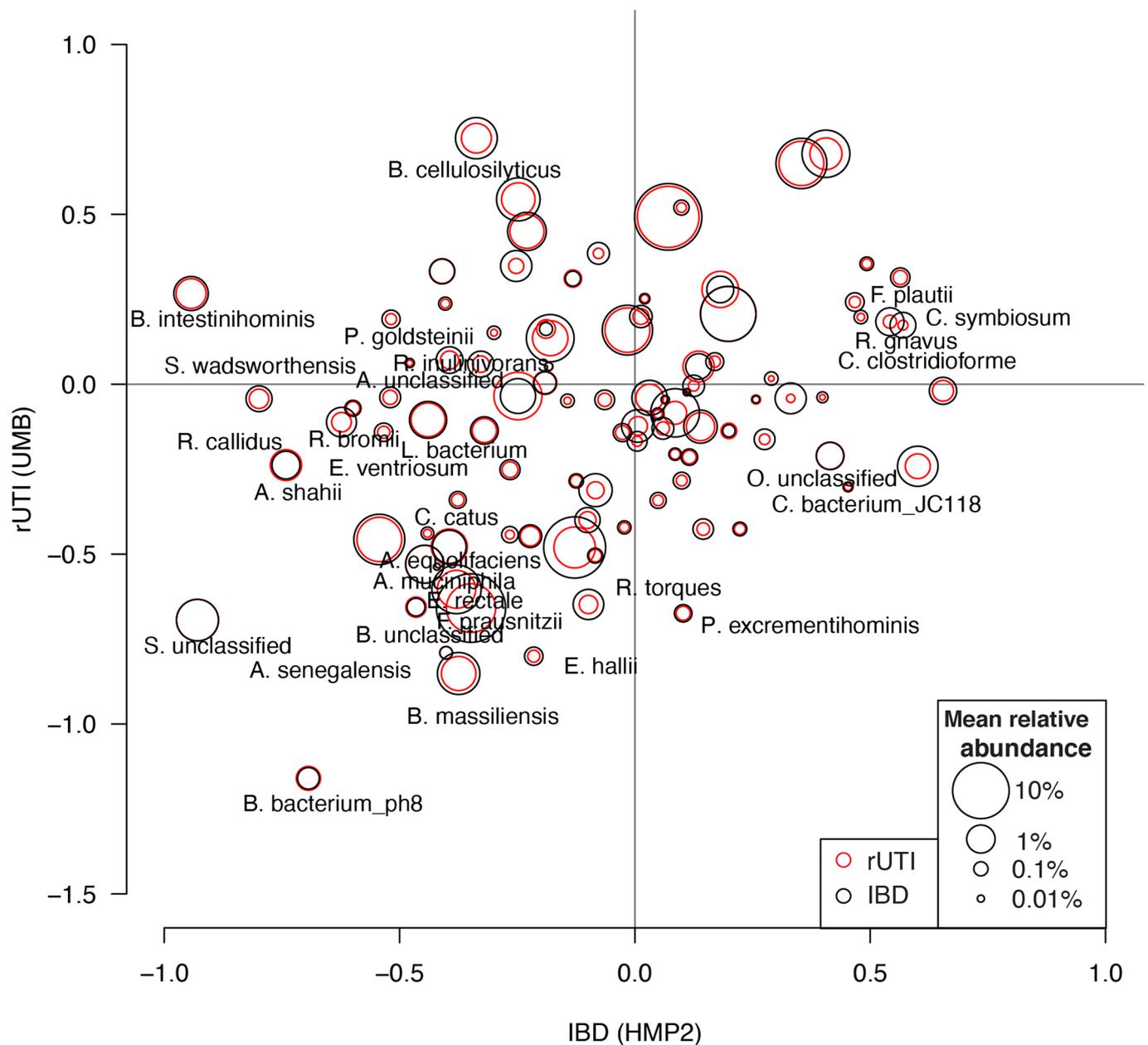
Extended Data Fig. 2 | SCFA producing bacteria are depleted in the rUTI gut. Cumulative relative abundances of (**a**) butyrate and (**b**) propionate producing bacterial species in rUTI and control samples. Box plots display the median (center line), 25th and 75th percentiles (box), as well as the 5th 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.
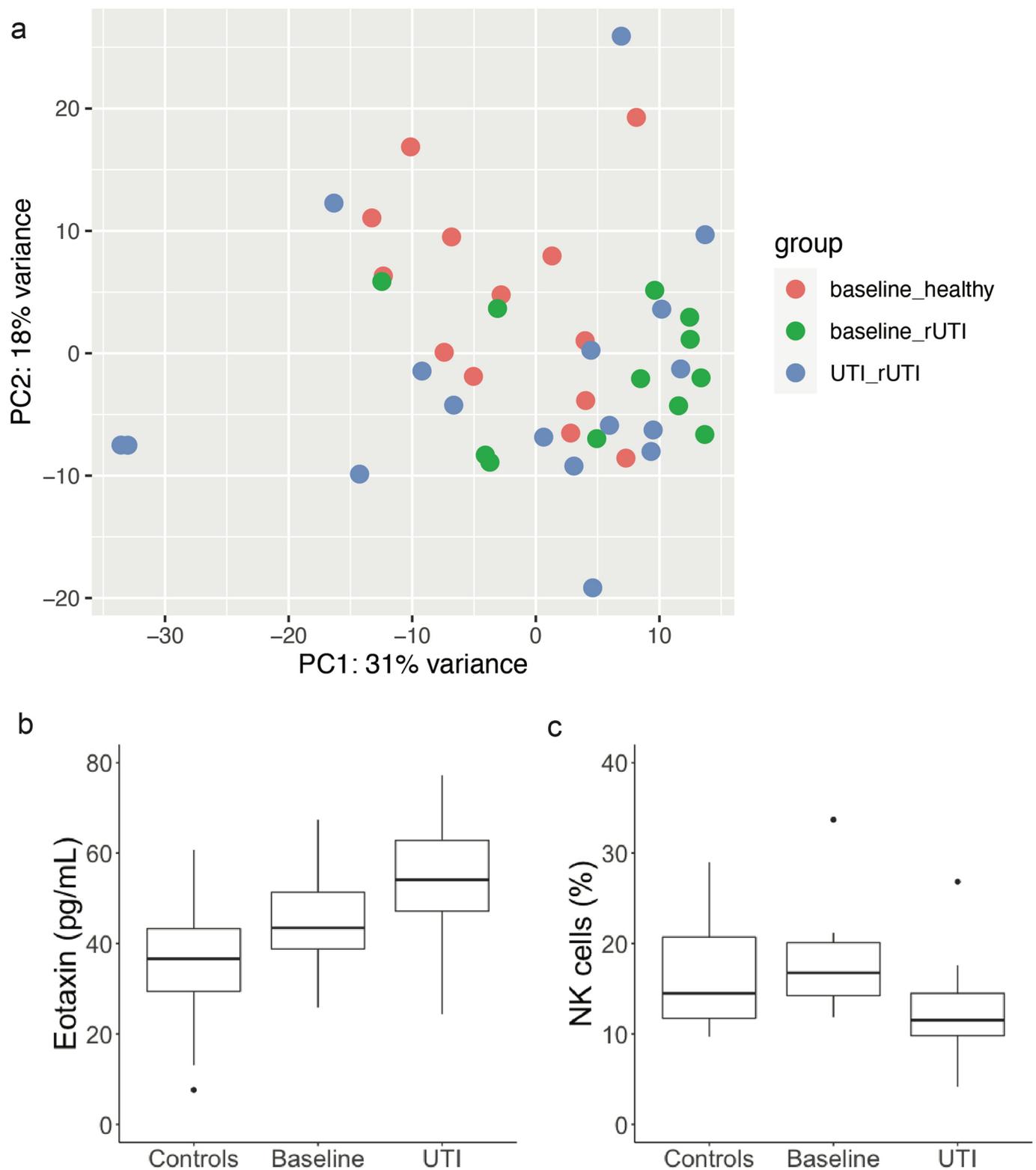
Extended Data Fig. 3 | Bray Curtis dissimilarity across stool samples. (a) For each patient, the distribution of Bray-Curtis dissimilarities between all stool samples, ordered by increasing mean patient values within each cohort. (**b**) Bray-Curtis distributions between samples taken at the time of 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).
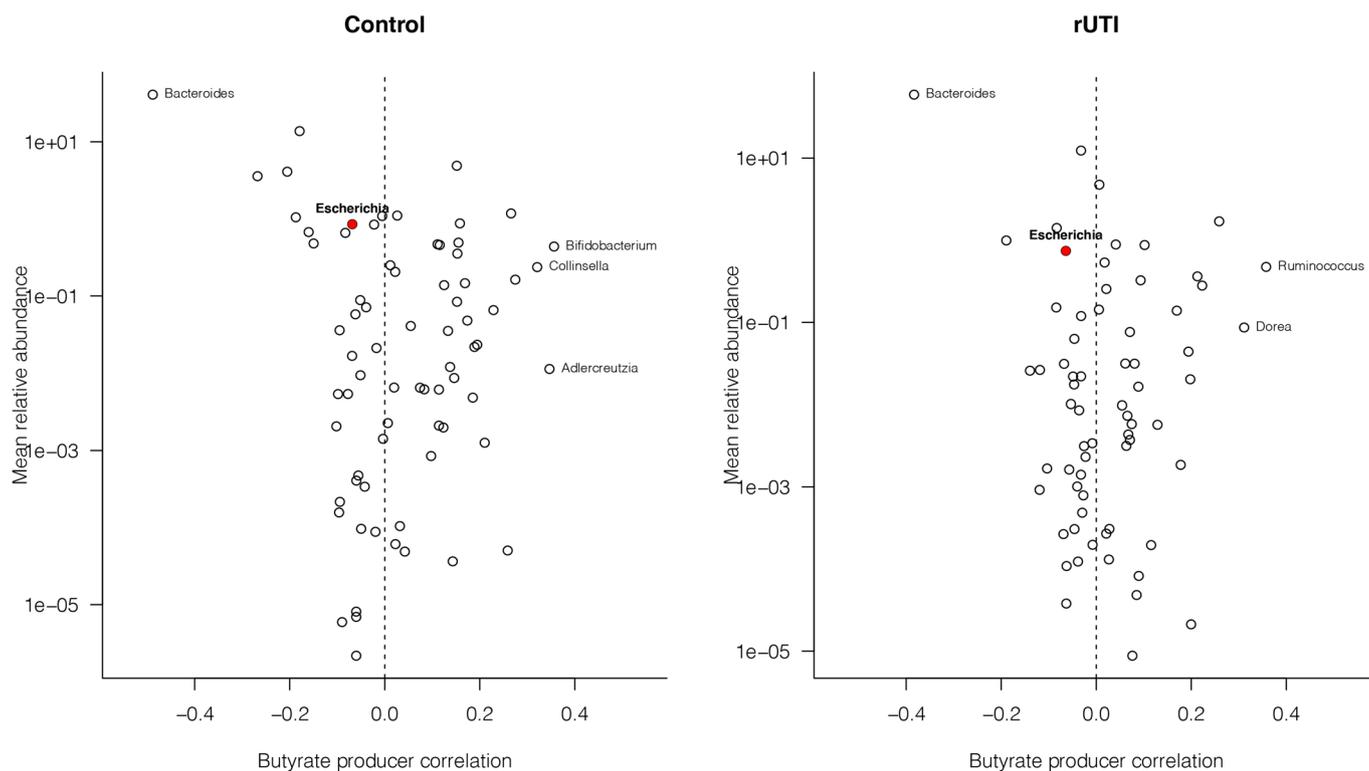
Extended Data Fig. 4 | rUTI dysbiosis is not driven by antibiotic use during the study. We grouped rUTI women according to their antibiotic exposures at any point during the UMB study; (i) ciprofloxacin (n=6) (ii) non-ciprofloxacin antibiotics (n=6); (iii) no antibiotics (n=3); (iv) any antibiotics (n=12). 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.
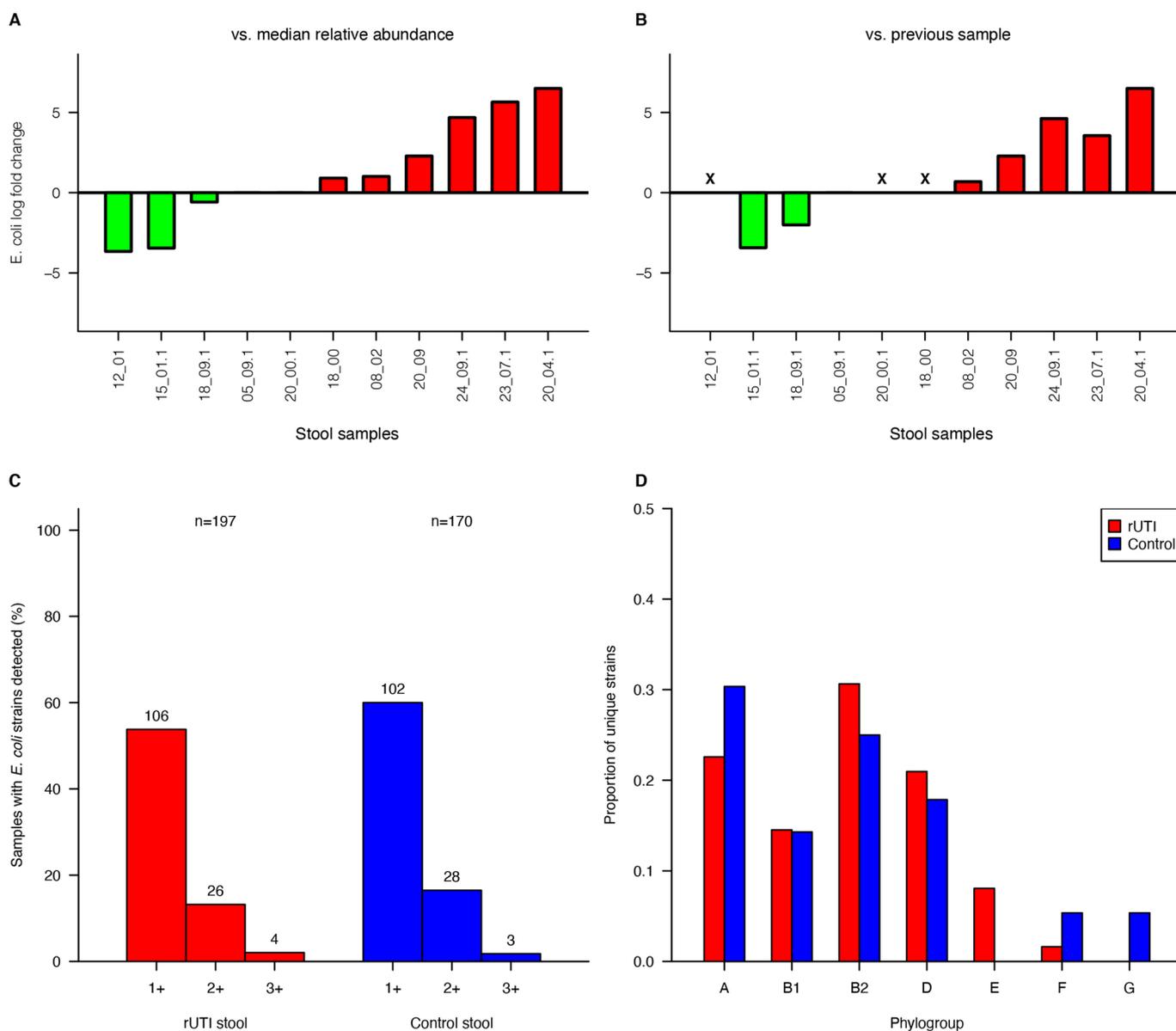
Extended Data Fig. 5 | Most species depleted in the rUTI gut are also depleted in the IBD gut. We compared discriminatory taxa in rUTI women to those in IBD patients using data from adult participants in the HMP2 study[33]. For each study, we fitted mixed effects models to standardized Metaphlan2 relative 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.
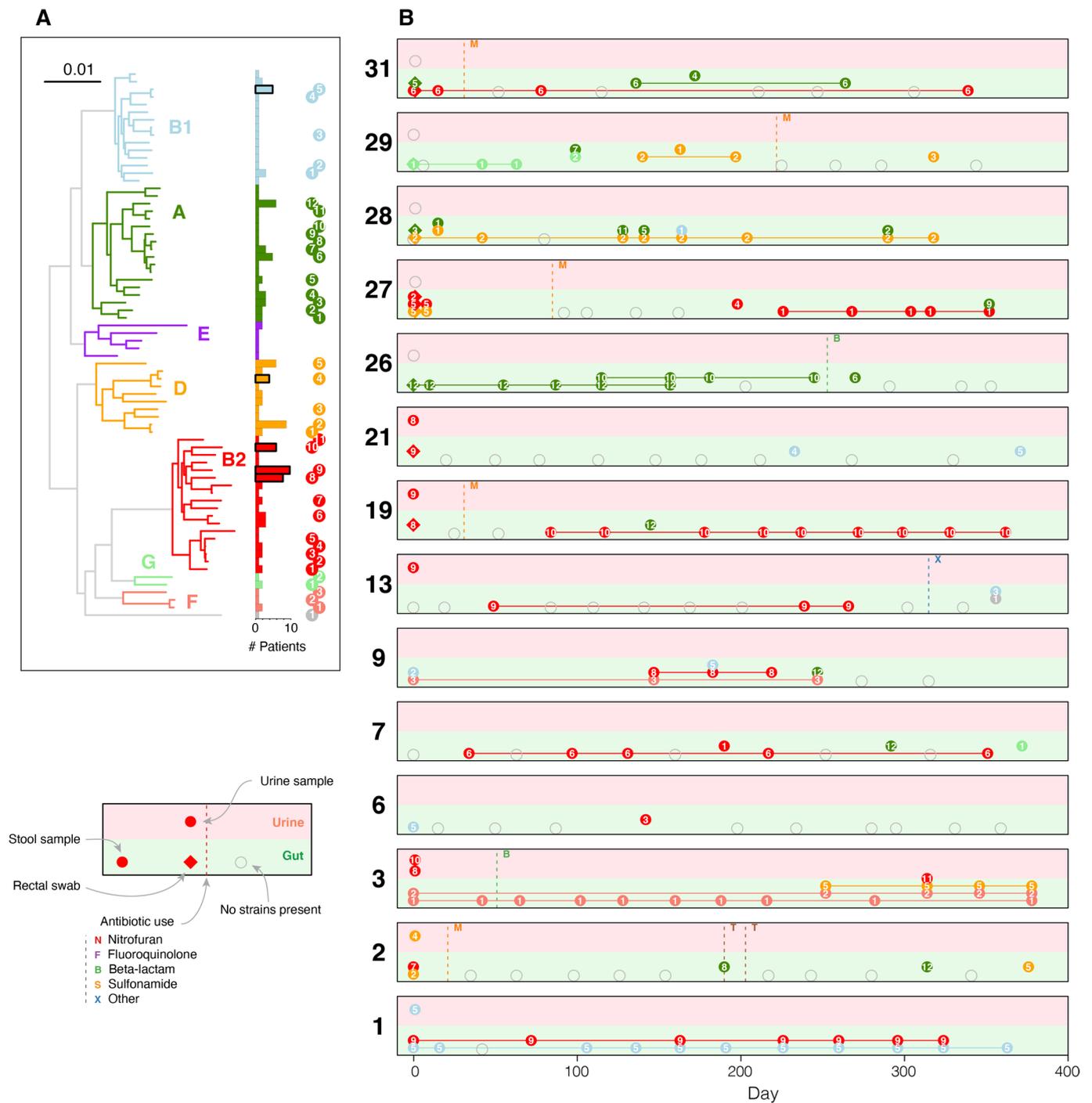
Extended Data Fig. 6 | Immunological differences between cohorts. (**a**) PCA plot of gene expression across cohorts, based on PBMC RNA Seq data. Samples are partitioned into healthy controls (n = 13), rUTI patient baseline (enrollment; n = 12) and rUTI patient at time of UTI (n = 17). (**b**) Plasma 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 | *E. coli* **relative abundance around the time of UTI and phylogroup distributions.** 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 Fig. 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 (colour) 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. * recorded recent consumption of item in at least 50% of questionnaire responses during study

	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

nature portfolio

Corresponding author(s): SJ Hultgren; AM Earl

Last updated by author(s): Mar 1, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Metagenomic sequence data were analyzed using Metaphlan2 (v2.7), Humann2 (v2.8.1). RNA-Seq data were analyzed using CIBERSORT. were Luminex assay results were read and quantified using a BioPlex multiplex plate reader and Microplate Manager software (v5). Statistical analysis was conducted using R v4.0.3 and R packages vegan (v2.5), DESeq (v1.30), lme4 (v1.1-27). Custom code to run analyses and generate figures is available at github.com/cworby/UMB-study

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Metagenomic sequence data are available from the Sequence Read Archive under Bioproject PRJNA400628. Questionnaire data, outout files from Metaphlan2, Humann2 and StrainGE are available from github.com/cworby/UMB-study. PBMC RNASeq data are PBMC RNASeq data are available from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002728.v1.p1. Human genome hg19 used in PBMC RNAseq analysis available under Bioproject PRJNA31257.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size calculation for longitudinal microbiome analyses is not straightforward, we note that each high-risk patient represents their own controlled experiment. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous longitudinal microbiome studies which were able to detect significant effects, e.g. Dethlefsen & Relman PNAS 2011. 108 Suppl 1: p.4554-61; Turnbaugh et al., Nature, 2006. 444(7122): p. 1027-31.
Data exclusions	No data were excluded from our study.
Replication	Our study was an observational cohort study and no replication was performed, although we have described the recruitment process and sampling strategy sufficiently such that the study may be replicated.
Randomization	Our study was an observational study with no intervention and cohorts based on pre-determined criteria; as such, no randomization was required. Control participants were age-matched to rUTI participants, and few dietary differences existed between the cohorts based on survey responses. We adjusted for race in cohort comparisons of microbiome structure.
Blinding	Our study was an observational study with no intervention and cohorts based on pre-determined criteria; as such, blinding was not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	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.
Recruitment	rUTI women were recruited based on clinical history via the Department of Urological Surgery, along with age-matched control women with no history of rUTI. Flyers were posted around Wash U Medical School, Wash U in St. Louis campus, and the Barnes-Jewish Hospital. Participants were remunerated midway through the study, and at the end of the study upon completion, via gift cards. Self selection biases may therefore exist; in particular we did not collect socio-economic data on participants. However, given age matching and the similarity in self-reported dietary habits between cohorts, we do not anticipate any such bias to have a significant impact on the composition of the gut microbiome.
Ethics oversight	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

Note that full information on the approval of the study protocol must also be provided in the manuscript.