


RESEARCH

Open Access



# Cataloging the phylogenetic diversity of human bladder bacterial isolates

Jingjie Du<sup>1,5</sup>, Mark Khemmani<sup>1</sup>, Thomas Halverson<sup>1</sup>, Adriana Ene<sup>2</sup>, Roberto Limeira<sup>3</sup>, Lana Tinawi<sup>3</sup>, Baylie R. Hochstedler-Kramer<sup>1</sup>, Melline Fontes Noronha<sup>1</sup>, Catherine Putonti<sup>2,4</sup> and Alan J. Wolfe<sup>1,3\*</sup> 

\*Correspondence:  
awolfe@luc.edu

<sup>1</sup> Department of Microbiology & Immunology, Stritch School of Medicine, Loyola University Chicago, Maywood, IL 60153, USA

<sup>2</sup> Bioinformatics Program, Loyola University Chicago, Chicago, IL 60660, USA

<sup>3</sup> Loyola Genomics Facility, Stritch School of Medicine, Loyola University Chicago, Maywood, IL 60153, USA

<sup>4</sup> Department of Biology, Loyola University Chicago, Chicago, IL 60660, USA

<sup>5</sup> Present address: Division of Nutritional Science, Cornell University, Ithaca, NY 14850, USA

## Abstract

**Background:** Although the human bladder is reported to harbor unique microbiota, our understanding of how these microbial communities interact with their human hosts is limited, mostly owing to the lack of isolates to test mechanistic hypotheses. Niche-specific bacterial collections and associated reference genome databases have been instrumental in expanding knowledge of the microbiota of other anatomical sites, such as the gut and oral cavity.

**Results:** To facilitate genomic, functional, and experimental analyses of the human bladder microbiota, we present a bladder-specific bacterial isolate reference collection comprising 1134 genomes, primarily from adult females. These genomes were culled from bacterial isolates obtained by a metaculturomic method from bladder urine collected by transurethral catheterization. This bladder-specific bacterial isolate reference collection includes 196 different species, including representatives of major aerobes and facultative anaerobes, as well as some anaerobes. It captures 72.2% of the genera found when re-examining previously published 16S rRNA gene sequencing of 392 adult female bladder urine samples. Comparative genomic analysis finds that the taxonomies and functions of the bladder microbiota share more similarities with the vaginal microbiota than the gut microbiota. Whole-genome phylogenetic and functional analyses of 186 bladder *Escherichia coli* isolates and 387 gut *Escherichia coli* isolates support the hypothesis that phylogroup distribution and functions of *Escherichia coli* strains differ dramatically between these two very different niches.

**Conclusions:** This bladder-specific bacterial isolate reference collection is a unique resource that will enable bladder microbiota research and comparison to isolates from other anatomical sites.

**Keywords:** *Escherichia coli*, Human bladder, Lower urinary tract symptoms, Metaculturomics, Transurethral catheterization, Urobiome, Urinary microbiome



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

High-throughput DNA sequencing and enhanced culture-based investigations have found bacterial DNA and live bacteria, respectively, in catheterized (bladder) urine deemed culture-negative by the standard urine culture method [1–13]. The expanded quantitative urine culture (EQUC) protocol and similar enhanced (metaculturomic) methods have enabled researchers to isolate species detected by 16S rRNA gene sequencing in both individuals with and without urinary tract symptoms [7, 8, 11, 13–15]. While the genetic diversity of some species has been extensively investigated, e.g., *Escherichia coli* [16, 17], many species have few or no genomic sequences. Furthermore, new urinary species are still being discovered, e.g., *Lactobacillus mulieris* [18], while others are being reclassified in light of new genomic data, e.g., *Gardnerella vaginalis* [19] and *Aerococcus urinae* [20].

Previously, we published an analysis of 149 genomes of 78 different species isolated from the female bladder [14]. Several of the taxa found within the bladder microbiota are also inhabitants of the vaginal microbiota [14, 21, 22]. This led us to posit that the bacterial communities of these two anatomical sites may be connected [14]. A recent 16S rRNA gene sequencing survey of paired bladder and vaginal samples identified more bacterial genera within the bladder than within the vagina, suggesting a greater bacterial diversity within the urinary tract [22]. In contrast, little overlap in species was reported between the microbiota of the urinary and gastrointestinal tracts [14], despite prior evidence that the gastrointestinal tract is the source of *E. coli* that cause urinary tract infections (UTIs) [23–26]. Furthermore, our prior work found that key metabolic pathways specific to the female urogenital environment are not found in bacterial genomes isolated from the gastrointestinal tract, suggesting that these two microbiotas are not tightly connected [14].

Capturing the species present, and the strain diversity of those species, is fundamental for the success of gene marker surveys, as well as shotgun metagenomic studies. In contrast to the gastrointestinal tract and oral cavity, very few shotgun metagenomic sequencing studies of the urinary microbiota have been conducted, and then only relatively recently [27–34]. Most of these studies analyzed voided urine, which often contains both urinary and genital microbes and thus cannot report on the bladder microbiota. Available genomes representative of the genetic and species diversity within the urinary bladder are essential for understanding the bladder microbiota and their association with symptoms and response to treatment.

In this study, we combined EQUC of urine obtained primarily from adult females by transurethral catheterization with large-scale whole-genome sequencing to generate a comprehensive bladder-specific bacterial isolate genome reference collection. To assess the completeness of this catalog, we compared the taxa found here to those identified in a reexamination of previously published shotgun metagenomic and high-throughput 16S rRNA gene sequencing surveys of bladder urine. We compared the genomes of this new culture collection to previously sequenced gut and vaginal isolates and identified taxonomies and functional similarities between the bladder and vagina but not the gut. Finally, we identified phylogenetic and functional differences between *E. coli* strains isolated from the bladder and gut.

## Results

### Demographics

In an effort to assemble a genomic reference collection representative of the phylogenetic diversity of the bladder, participants of this study were recruited as part of previous and ongoing IRB-approved studies. Out of a total of 5619 bladder isolates in our urinary bacteria collection, we selected 1050 strains for whole-genome sequencing. Some isolates came from the same participant; in such cases, the isolates were always from different species. Our objective was to sequence at least one genome for each species identified by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS). For species commonly found in either symptomatic or asymptomatic individuals, we sequenced more than 20 isolates. These species include *Escherichia coli*, *Klebsiella pneumoniae*, *Enterococcus faecalis*, *Streptococcus anginosus*, *Streptococcus agalactiae*, *Aerococcus urinae*, *Lactobacillus crispatus*, and *Lactobacillus jensenii*.

Using a metaculturomic method called expanded quantitative urine culture (EQUC) [7, 8], we isolated these bladder strains from the urine obtained by transurethral catheterization of 96 asymptomatic controls and 377 symptomatic individuals. The symptomatic individuals had been diagnosed with urinary tract infection (UTI,  $n=212$ ), recurrent urinary tract infection (RUTI,  $n=54$ ), overactive bladder (OAB,  $n=76$ ), stress urinary incontinence (SUI,  $n=10$ ), bladder cancer ( $n=12$ ), bladder and bowel dysfunction (BBD,  $n=3$ ), kidney stone ( $n=1$ ), interstitial cystitis/bladder pain syndrome (IC/BPS,  $n=1$ ), pelvic organ prolapse (POP,  $n=5$ ), or unknown urinary tract symptoms ( $n=3$ ). Participants were from 6 self-reported races: White/Caucasian ( $n=340$ ), Black ( $n=57$ ), Hispanic ( $n=41$ ), Asian ( $n=5$ ), Native American ( $n=3$ ), Middle Eastern ( $n=2$ ), and no reply ( $n=25$ ). Most participants were females ( $n=454$ ); only 19 were males. Participants ranged in age from 4 weeks to 99 years old (Table 1). Participant metadata for each genome are listed in Additional file 2: Table S1.

### The bladder-specific isolate genomes are diverse

Despite best efforts, many clinical isolates cannot be completely purified. Besides the primary, originally identified species, some isolates contained one or more secondary, minority species; in another system, these minority species have been called hitchhikers [35]. Of our 1050 isolates, 218 consisted of more than 1 species. In total, we obtained 1134 high-quality genomes (>90% completeness, <5% contamination), including 781 genomes produced as part of this current effort and 353 genomes from our collection that were previously deposited (Additional file 1: Fig. S1, Additional file 2: Table S1). The overall qualities of these genome assemblies were high: the median completeness level was 99.39%, and the median estimated contamination was 0.86%. The median positional coverage was  $160\times$ , and the median contigs L50 (the number of contigs needed to cover 50% of the genome assembly) was 14.24 (Fig. 1A–D).

Most genomes ( $n=1096$ ) were from females; only 38 were from males (Additional file 2: Table S1). This collection of genomes represented 8 phyla, 10 classes, 21 orders, 39 families, 89 genera, and 196 species, of which 12 species are unique to males (Fig. 1E). This includes 7 genomes for which neither the NCBI taxonomy database nor the Genome

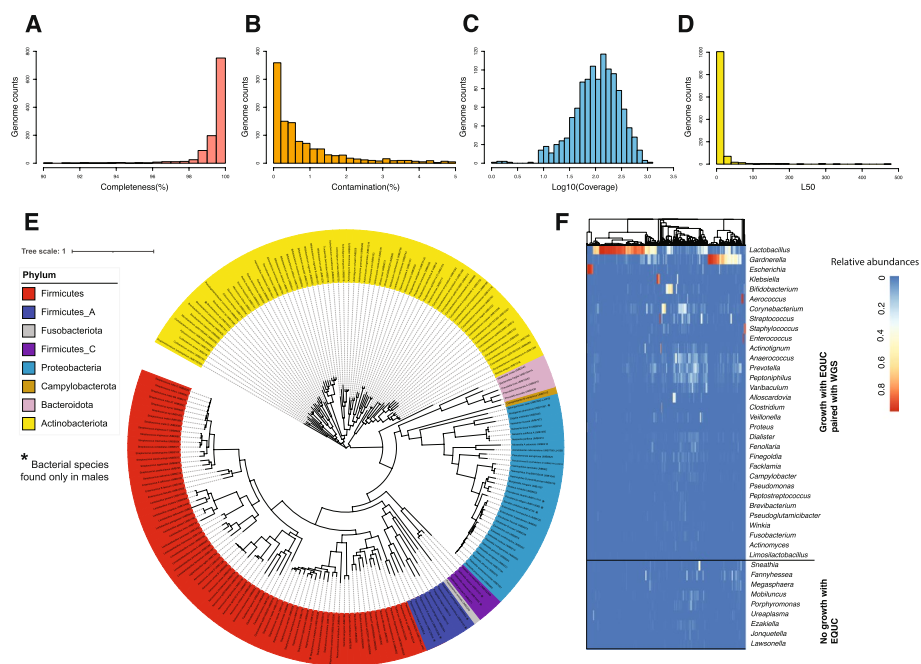
**Table 1** Demographics of participants in this study

		Symptomatic individuals (n = 377)	Asymptomatic controls (n = 96)
<b>Sex</b>	Female	366	88
	Male	11	8
<b>Age</b>	< 18	2	5
	18–40	28	23
	40–65	109	45
	> 65	227	15
<b>Disease states</b>	UTI	212	N/A
	RUTI	54	N/A
	OAB	76	N/A
	SUI	10	N/A
	Bladder cancer	12	N/A
	BBD	3	N/A
	Kidney stone	1	N/A
	IC/BPS	1	N/A
	POP	5	N/A
	With symptoms	3	N/A
	<b>Self-reported race</b>	White	282
Black		41	16
Hispanic		29	12
Asian		5	0
Native American		2	1
Middle Eastern		2	0

Taxonomy Database (GTDB) could assign a species (Additional file 2: Table S1). These 7 genomes (from UMB1203, UMB1308A, UMB1298, UMB7805-LC452B, UMB4589-SE434, UMB1308B, and UMB10442) represent isolates of the genera *Brevibacterium*, *Corynebacterium*, *Paenibacillus\_B*, *Sphingomonas*, and *Streptococcus*.

Pairwise whole-genome average nucleotide identity (ANI) comparison of the 196 species met a 95–96% threshold (Additional file 1: Fig. S2), the generally accepted ANI threshold (94–96%) [36, 37] for distinguishing between two bacterial species. More species were isolated from symptomatic participants ( $n = 169$ , Additional file 1: Fig. S3) than from asymptomatic participants ( $n = 76$ , Additional file 1: Fig. S4). This skew likely resulted from the simple fact that 79% of the genomes were from symptomatic participants.

To determine how well our genome collection represents the bacterial diversity captured by culture-independent methods, we re-analyzed 16S rRNA gene sequencing data previously obtained from 392 female bladder urine samples [3, 5, 38–40] (Additional file 2: Table S2). We found that our sequenced genomes represented 72.2% (26/36) of the most abundant genera (>0.1% relative abundance) captured by 16S sequence analysis (Fig. 1F). We also analyzed publicly available metagenomes from 42 urine samples collected using catheters across two studies [28, 31]. Our sequenced genomes represented 77.8% (21/27) of the bacterial genera and 52.8% (28/53) of the bacterial species (Additional file 2: Table S3).



**Fig. 1** Overall information of the bladder microbiota culture collection. **A–D** Distribution of genome assembly, completeness, contamination, coverage (on a log10 scale), and L50 values, respectively. **E** Phylogenomic tree of the 196 bacterial species (assigned by GTDB) represented in the bladder genome collection. A single genome per species was selected out of the 1134 high-quality isolated genomes (>90% completeness, <5% contamination) to reconstruct the multiple sequence alignment of 71 core bacterial genes. Bacterial species are colored by phylum. \*The bacterial species only found in males. **F** The diversity of isolated genomes was compared with the total diversity observed in the 16S sequence data (as shown in genus level) obtained from 392 bladder urine samples

Among the 50 genera detected by either 16S rRNA gene sequencing or metagenomics sequencing, 12 genera lack strains within our genome collection: *Agathobacter*, *Anaeroglobus*, *S5-A14a* (a genus in Anaerovoracaceae family) *Sneathia*, *Fannyhessea*, *Magasphaera*, *Mobiluncus*, *Porphyromonas*, *Ureaplasma*, *Ezakiella*, *Jonquetella*, and *Lawsonella*. These genera are strictly anaerobic, particularly fastidious, or lack a cell wall.

For other abundant anaerobes detected by either 16S sequencing or metagenomics sequencing, such as *Anaerococcus* and *Prevotella*, we had only a few strains in our cultured isolate collection (Additional file 1: Fig. S1, Additional file 2: Table S1). For example, the average relative abundance of *Anaerococcus* and *Prevotella* is about 2.8% and 3.2%, respectively, among all 16S rRNA gene sequenced samples, but our collection of 5619 isolates only contains 3 *Anaerococcus* and 5 *Prevotella* strains. Moreover, our sequenced genomes represented only a few of the *Anaerococcus* and *Prevotella* species detected by metagenomic sequencing (Additional file 2: Table S3).

In contrast, we were able to isolate and thus sequence taxa that were simply missed by 16S rRNA gene or metagenomics sequencing or were present only as rare taxa (<0.1% average relative abundances across all individuals); these included isolates from the genera *Acinetobacter*, *Alcaligenes*, *Bacillus*, *Bacteroides*, *Citricoccus*, *Citrobacter*, *Curtobacterium*, *Cytobacillus*, *Dermabacter*, *Dolosicoccus*,

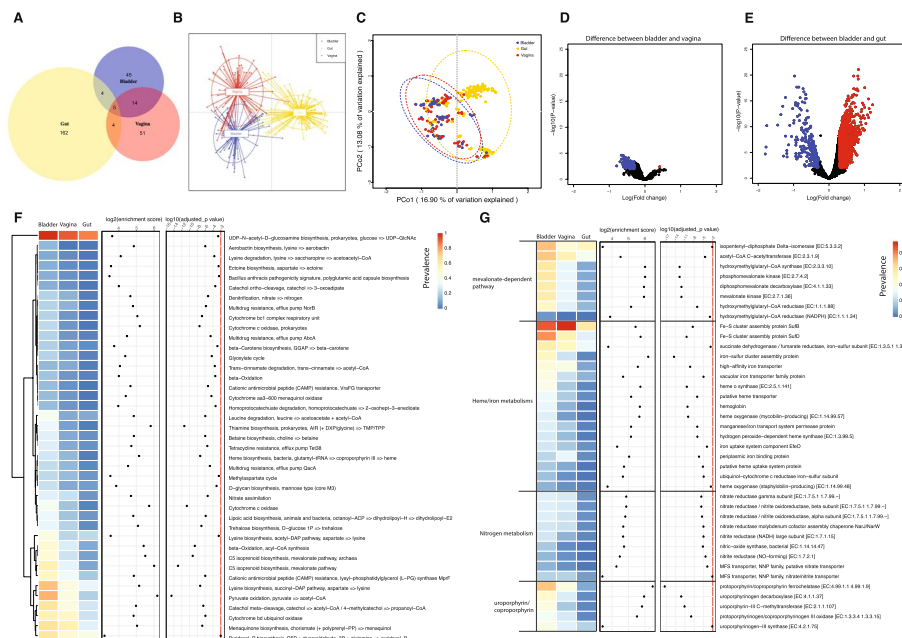
*Enterobacter*, *Gemella*, *Gleimia*, *Globicatella*, *Gordonia*, *Granulicatella*, *Haemophilus*, *Helcobacillus*, *Kocuria*, *Kytococcus*, *Ligilactobacillus*, *Metabacillus*, *Microbacterium*, *Moraxella\_A*, *Morganella*, *Micrococcus*, *Neisseria*, *Niallia*, *Nosocomiicoccus*, *Paenibacillus*, *Pauljensenia*, *Peribacillus*, *Priestia*, *Rothia*, *Serratia*, *Slackia*, *Sphingomonas*, *Trueperella*, *Weeksella*, and *Oligella* (Additional file 2: Tables S1, S2 and S3).

### **Bladder isolates are functionally and taxonomically more similar to vaginal isolates than gut isolates**

To assess the bladder microbiota relative to those of other well-studied anatomical sites, we compared the bladder strains that we had isolated from asymptomatic females to gastrointestinal and vaginal strains isolated by others from unrelated asymptomatic individuals. We compared the taxonomies of genomes of the 69 bladder bacterial species isolated from females to genomes of 74 vaginal bacterial species (including 4 species from our vaginal isolate collection) and genomes of 175 gut bacterial species, all cultivated from unrelated asymptomatic individuals (Additional file 2: Table S4). Only 5 species (*Bifidobacterium bifidum*, *Enterococcus faecalis*, *Escherichia coli*, *Lactocaseibacillus rhamnosus*, and *Proteus mirabilis*) were detected in all 3 niches, and only 9 species were found in both the bladder and gut microbiota. In contrast, 19 species were found in both the bladder and vaginal microbiota (Fig. 2A, Additional file 2: Table S4). Taken together, these results are consistent with the hypothesis that bacterial species from each niche are distinct and that the bladder shares more bacterial species with the vagina than the gut.

We next used the pangenomes to evaluate the functional differences in the microbiota by analyzing the protein functions encoded by the bacterial species in the three niches. Applying discriminant analysis of principal components (DAPC) and principal coordinates analysis (PCoA) using the Bray–Curtis Dissimilarity Index, we compared the Kyoto Encyclopedia of Genes and Genomes (KEGG, Fig. 2B, C, Additional file 2: Table S5) and Clusters of Orthologous Groups (COG, Additional file 1: Fig. S5A–B, Additional file 2: Table S6) annotations of the 69 bladder, 74 vaginal, and 175 gut bacterial species. Like the species analysis described above, this comparison identified more overlapping protein functions shared by the bladder and vaginal bacterial species that were largely distinct from protein functions found in the gut bacterial species. Volcano plots also revealed 321 KOfam functions differentially abundant between bladder and vaginal species ( $t$ -test, adjusted- $p < 0.01$ ,  $[\log\text{FD}] > 0.3$ , Fig. 2D) and 1778 KOfam functions differentially abundant between bladder and gut species ( $t$ -test, adjusted- $p < 0.01$ ,  $[\log\text{FD}] > 0.3$ , Fig. 2E). We observed similar findings for COG functions. A total of 115 COG functions were differentially abundant between the bladder and vaginal species. In contrast, 1611 COG functions were differentially abundant between the bladder and gut species ( $t$ -test, adjusted- $p < 0.01$ ,  $[\log\text{FD}] > 0.3$ , Additional file 1: Fig. S5C–D). These data indicate the existence of many functions shared by the bladder and vaginal bacterial species that are clearly distinct from those of gut bacterial species.

We identified Kegg Orthology (KO) modules that were significantly different between the three niches (Additional file 2: Table S7). Forty-one KO modules were enriched in the bladder (adjusted- $p < 0.01$  and  $> 10\%$  prevalence in all bladder genomes) compared with the gut and/or vagina (Fig. 2F). For example, the mevalonate-dependent pathway for



**Fig. 2** Comparison of bladder isolates with vaginal and gut isolates. **A** Venn diagram showing the number of bacterial species isolated from asymptomatic individuals shared among three different niches: bladder (blue;  $n = 68$ ), vagina (red;  $n = 74$ ), and gut (yellow;  $n = 175$ ). Reference genomes of the vagina and gut isolated from asymptomatic donors were obtained from previous studies. Bladder isolates from female asymptomatic controls were used in the comparison. **B, C** DAPC and PCoA of KEGG functional diversity of bacterial species isolated from the asymptomatic individuals of different niches. **D, E** Volcano plots of  $t$  tests corrected by the Benjamini and Hochberg method for changes in Kofam functions between the bladder and the vagina and between the bladder and the gut. An FDR cutoff of 0.01 and  $|\log_2FC| > 0.3$  were used. Data points highlighted in red indicate the Kofam functions that were significantly enriched in the gut or vagina, while data points highlighted in blue indicate the Kofam functions that were significantly enriched in the bladder. **F** KO modules that were enriched in the bladder compared with the gut and/or vagina. An FDR cutoff of 0.01 and more than 10% prevalence in all bladder genomes were used. **G** Selected Kofam functions that were enriched in the bladder as compared with the gut and/or vagina. Functions that had at least 10% prevalence in all bladder genomes with an FDR cutoff of 0.01 were used

isoprenoid biosynthesis was enriched in the bladder genomes (Fig. 2F, Additional file 2: Table S7). All Kofam functions associated with the mevalonate-dependent pathway were enriched in the bladder genomes (Fig. 2G and Additional file 2: Table S8), including acetyl-CoA C-acetyltransferase, hydroxymethylglutaryl-CoA synthase, hydroxymethylglutaryl-CoA reductase (NADPH), hydroxymethylglutaryl-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentenyl-diphosphate delta-isomerase. We also observed enriched KO modules in the bladder genomes related to acetyl-CoA, which is the precursor for the MEV pathway of isoprenoid biosynthesis. These acetyl-CoA-related KO modules include beta-oxidation, acyl-CoA synthesis, and pyruvate oxidation. Some bladder-enriched KO modules relate to lysine metabolism (i.e., lysine degradation and biosynthesis) or nitrogen metabolism (i.e., denitrification and nitrate assimilation), while others relate to iron utilization (e.g., aerobactin biosynthesis and heme biosynthesis) (Fig. 2F, Additional file 2: Table S7). Given the scarcity of iron in the bladder, we investigated specific functions involved in iron utilization/assimilation. Seventeen Kofam functions related to heme biosynthesis and degradation, as well as to iron utilization and transport, were enriched in the bladder

genomes, including iron-sulfur cluster assembly proteins, hemoglobin, heme o synthase, and iron uptake system component EfeO among others. Several bladder-enriched functions were related to protoporphyrin/coproporphyrin utilization, including protoporphyrin/coproporphyrin ferrochelatase, which uses ferrous iron as one of its substrates (Fig. 2G and Additional file 2: Table S8). Analyses of COG functions also supported the enrichment of the mevalonate-dependent pathway, enterochelin transport, and pyruvate oxidation in the bladder genomes (Additional file 1: Fig. S5E, Fig. S6, Additional file 2: Table S9). Enrichment of the mevalonate-dependent pathway over the more common methylerythritol 4-phosphate pathway and of functions involved in iron utilization and export may relate to iron availability within the bladder.

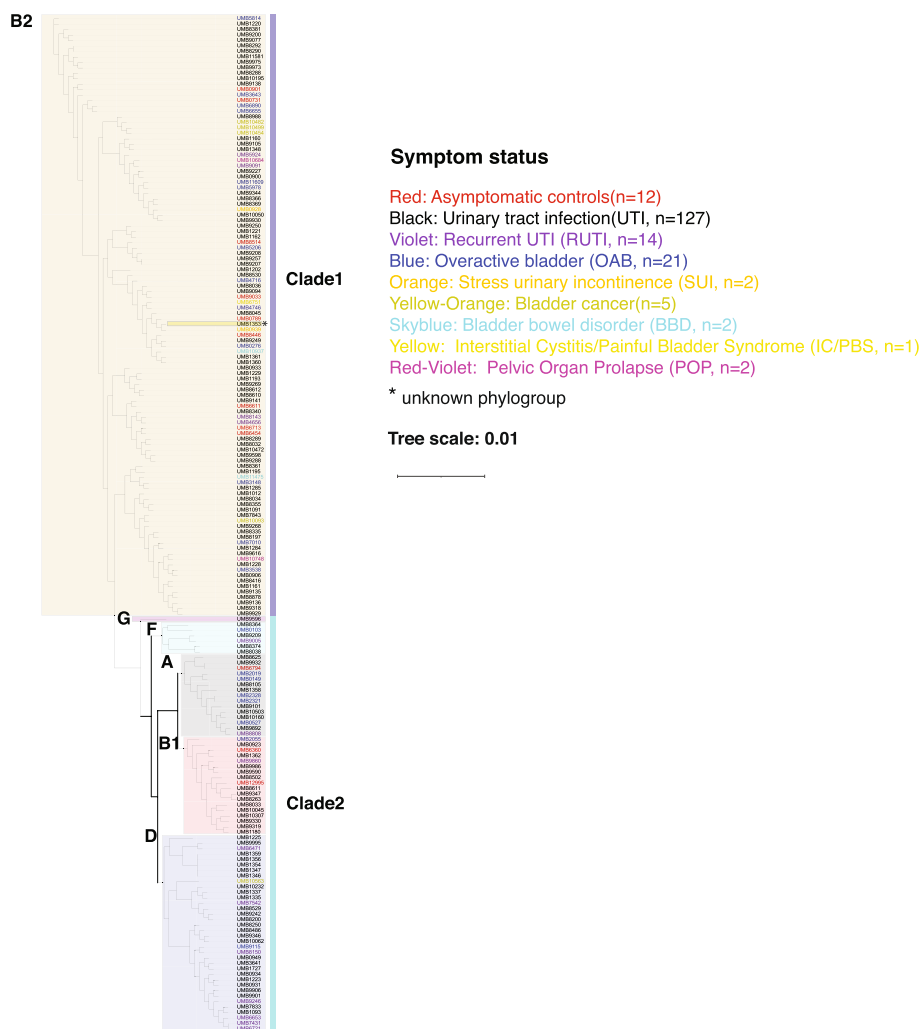
### Genomes of *Escherichia coli* strains isolated from the bladder differ from the gut

Given the significant taxonomic and functional differences between the bladder and gut genomes in asymptomatic individuals, we next sought to evaluate whether isolates of the same species differed in these 2 niches. As the gut is generally considered to be the source of uropathogenic *Escherichia coli* (UPEC) strains [41, 42], we compared our 186 bladder *E. coli* genomes with 387 publicly available gut *E. coli* isolates from unrelated healthy individuals [43].

From the genomes of these 186 bladder *E. coli* isolates, we constructed a phylogenomic tree based on 1084 single-copy core genes (Fig. 3). These genomes belonged to 2 distinct clades and 6 phylogroups. Phylogroup B2 and 1 unknown phylogroup belonged to clade 1, whereas phylogroups G, F, A, B1, and D belonged to clade 2. Phylogroup B2 (58.6%,  $n = 109$ ) predominated, followed by phylogroups D (19.4%,  $n = 36$ ), B1 (9.7%,  $n = 18$ ), A (8%,  $n = 15$ ), F (3.2%,  $n = 6$ ), G (0.5%,  $n = 1$ ), and an unknown phylogroup (0.5%,  $n = 1$ ) (Table 2, Additional file 2: Table S10). These genomes were from *E. coli* strains isolated from the bladders of 12 asymptomatic controls and 174 symptomatic individuals who were diagnosed with a urinary tract infection (UTI,  $n = 127$ ), recurrent UTI (RUTI,  $n = 14$ ), overactive bladder (OAB,  $n = 21$ ), stress urinary incontinence (SUI,  $n = 2$ ), bladder cancer ( $n = 5$ ), bladder bowel disorder (BBD,  $n = 2$ ), interstitial cystitis/painful bladder syndrome (IC/PBS,  $n = 1$ ), or pelvic organ prolapse (POP,  $n = 2$ ). However, symptom status was not associated with either clade or phylogroup (Fig. 3). The gut genomes also belonged to the same 6 phylogroups; however, the distribution differed. In contrast to the bladder isolates, only 3% of the gut genomes belonged to phylogroup B2 ( $n = 13$ ) while 33.85% belonged to phylogroup D ( $n = 131$ ), 37% to phylogroup F ( $n = 143$ ), 23.5% to A ( $n = 91$ ), 1.8% to B1 ( $n = 7$ ), 0.26% to phylogroup G ( $n = 1$ ), and 0.26% to an unknown phylogroup ( $n = 1$ ) (Table 2, Additional file 2: Table S10). Thus, phylogroup distribution differed dramatically between these 2 niches.

We next assessed the functions. Applying PCoA using the Bray–Curtis Dissimilarity Index, we compared the annotated KEGG functions of the bladder (Additional file 2: Table S11) and gut *E. coli* genomes (Additional file 2: Table S12) and found that they differed significantly (Fig. 4A,  $R^2 = 0.11598$ ,  $p < 0.001$ ). This difference may be because the functions of different phylogroups significantly differed (Fig. 4B,  $R^2 = 0.58919$ ,  $p < 0.001$ ), especially between phylogroup B2 and the other phylogroups and that the distribution of the phylogroups in bladder and gut differed (Table 2).



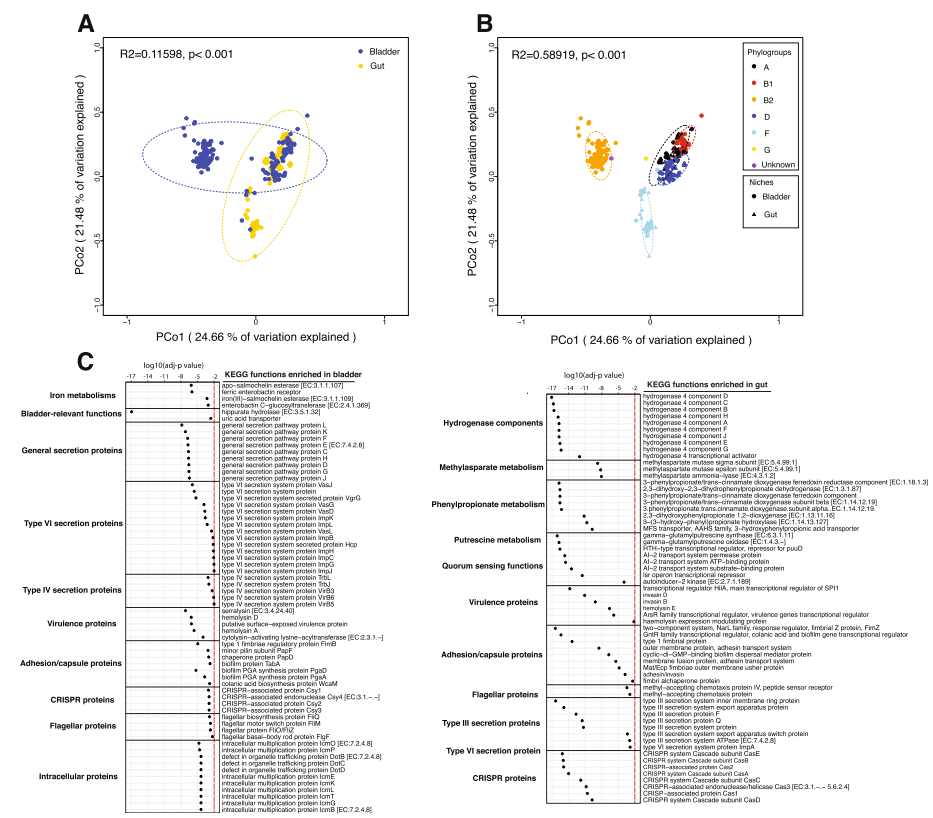


**Fig. 3** Phylogenetic analysis of 186 bladder-specific *Escherichia coli*-isolated genomes. Phylogenomic relationship of 186 *E. coli* isolates from the bladder derived using 201,624 single-copy core genes from their genomes. Phylogroups are indicated by shaded colors. The symptom status of the host for each isolate is indicated by the color of the isolate name

**Table 2** Number and percent of different *E. coli* phylogroups in bladder and gut genomes

	B2	A	B1	D	F	G	Unknown
Bladder	109 (58.6%)	15 (8%)	18 (9.7%)	36 (19.4%)	6 (3.2%)	1 (0.5%)	1 (0.5%)
Gut	13 (3.4%)	91 (23.5%)	7 (1.8%)	131(33.9%)	143 (37%)	1 (0.26%)	1 (0.26%)

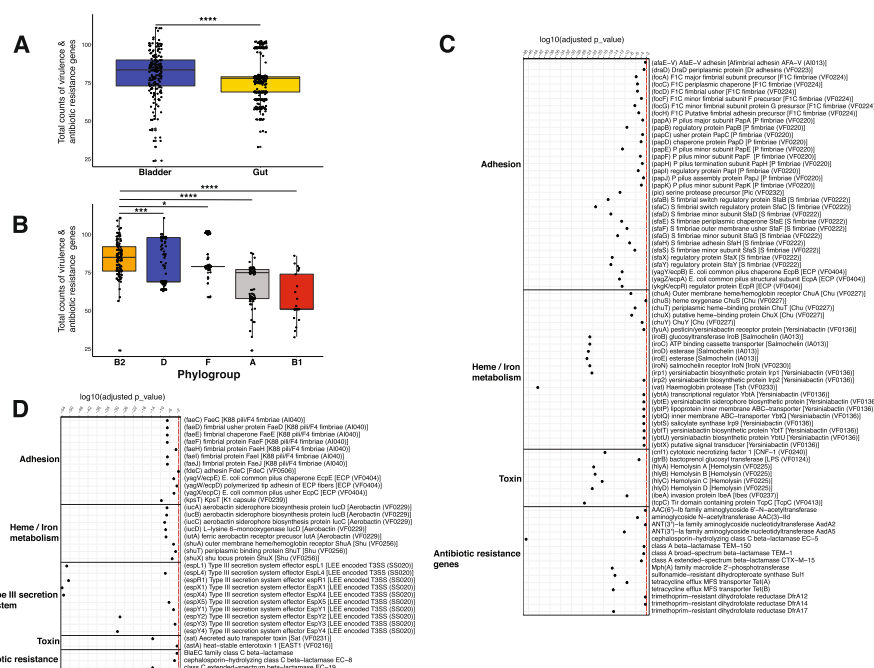
We next identified the functions that may facilitate *E. coli* adaptation and colonization to these two environmentally different niches by evaluating the predicted functions encoded by the bladder and gut genomes that were both abundant and differentially present (Fig. 4C, Additional file 2: Table S13). Bladder-enriched functions included those associated with iron metabolism (i.e., the siderophores salmochelin and enterobactin), bladder-relevant functions (i.e., hippuric hydrolase and a uric acid transporter), the general secretion pathway, both type VI and IV secretion systems, and intracellular



**Fig. 4** Bladder-specific *Escherichia coli* reference genomes are different from the gut reference genomes. **A** The KOfam functions of *E. coli* genomes annotated by KEGG are separated by niches (bladder and gut). **B** The KOfam functions of *E. coli* genomes annotated by KEGG are separated by phylogroups. **C** Overrepresented KOfam functions associated with bladder-specific or gut-specific genes

multiplication and trafficking proteins. In contrast, gut-enriched functions related to hydrogenase components, methylaspartate, phenylpropionate metabolism, putrescine metabolism, type III secretion system, and quorum sensing. Both sets of genomes harbored different enriched functions related to virulence, flagella, capsule, adhesion, and CRISPR.

To further evaluate the virulence and antibiotic susceptibility profiles of *E. coli* strains in these 2 niches, we mapped the *E. coli* genomes against the virulence factor database (VFDB) and Comprehensive Antibiotic Resistance Database (CARD). A total of 187 different virulence factors and 53 antibiotic resistance genes were detected in the 186 bladder genomes (Additional file 2: Tables S14 and S15), and 128 different virulence factors and 16 antibiotic resistance genes were detected in the 387 gut genomes (Additional file 2: Tables S16 and S17). The number of virulence and antibiotic-resistance genes was significantly greater in the bladder *E. coli* genomes (Fig. 5A). Again, this may result from the distribution of phylogroups, as phylogroup B2 (average counts per genome = 84.3) had slightly more virulence and antibiotic resistance genes than phylogroups D (average counts per genome = 79.6) and F (average counts per genome = 81.7) and much more than phylogroups A (average counts per genome = 67.65) and B1 (average counts per genome = 57.88) (Fig. 5B).



**Fig. 5** Comparison of virulence and antibiotic resistance genes between *Escherichia coli* reference genomes isolated from the bladder and gut. **A** Boxplots analysis showing the comparison of counts of virulence and antibiotic resistance genes in the bladder and gut *E. coli* genomes. Each data point represents the counts of virulence and antibiotic resistance genes in each *E. coli* genome. **B** Boxplot analysis showing the comparison of counts of virulence and antibiotic resistance genes in different *E. coli* phylogroups. Phylogroup G was not shown here due to the low sample size. \*, \*\*, \*\*\*, \*\*\*\* indicated FDR < 0.05, 0.01, 0.001, and 0.0001, respectively. **C** Virulence and antibiotic resistance genes that were enriched in bladder *E. coli* genomes. **D** Virulence and antibiotic resistance genes that were enriched in gut *E. coli* genomes. Virulence and antibiotic resistance genes with an FDR of < 0.01 were displayed here

Seventy-seven virulence and antibiotic resistance genes were enriched in the bladder genomes (Fig. 5C, Additional file 2: Table S18), whereas 35 were enriched in the gut genomes (Wilcoxon rank sum test, adj-*p* < 0.01, Fig. 5D, Additional file 2: Table S18). For example, the virulence factor hemoglobin protease (*vat*) and the antibiotic resistance gene cephalosporin-hydrolyzing class C beta-lactamase EC-5 (*blaEC-5*) were present in 44% and 55% of bladder genomes, respectively, but only in 0.5% and 0.3% of gut genomes, respectively. Other virulence genes enriched in the bladder genomes included those that are predicted to encode adhesins (e.g., *afaE-V*, *focA*, *C-D*, *F-H*, *papA-F*, *H*, and *sfaB-H*, *S*, *X-Y*), heme/iron-related virulence (e.g., *chuA*, *S-T*, *X-Y*, *iroB-E*, *N*, *irp1-2*, and *ybtA*, *E*, *P-Q*, *S-U*, *X*), toxins (e.g., *cnf1*, *hlyA-D*, *ibeA*, and *tcpC*), and resistance to several antibiotics (e.g., aminoglycosides, beta-lactamases, macrolides, tetracycline, and trimethoprim). In contrast, the gut strains were highly enriched for genes associated with type III secretion and enriched for multiple other virulence factors, including adhesion genes (e.g., *faeC-J*, *fdeC*, and *ecpC-E*), heme/iron-related metabolism (i.e., *iucA-D* and *shuA*, *T*, *X*), toxins (i.e., *sat* and *astA*), and 3 class C beta-lactamase genes. Taken together, these results suggest that bladder *E. coli* strains from asymptomatic individuals harbor more virulence and antibiotic resistance genes than commensal gut *E. coli* strains and that *E. coli* strains use different virulence genes to adapt to different niches.

## Discussion

Here, we have presented the largest collection of bladder-specific bacterial isolates paired with whole-genome sequences, providing a valuable resource for hypothesis-driven and data-driven bladder microbiome research. This extensive, genome-sequenced culture collection represents more than 70% of bacterial genera detected by previously reported 16S rRNA gene sequencing of bladder urine samples collected by transurethral catheterization [3, 5, 38–40]. The missing genera are primarily strict anaerobes that are not aerotolerant but also include a few that are particularly fastidious or lack a cell wall. For example, per 16S rRNA gene [4, 21, 38, 44–48] or shotgun metagenomic sequencing [14], the bacterial genera *Rothia*, *Dialister*, *Bacteroides*, *Peptoniphilus*, *Prevotella*, and *Anaerococcus* are abundant and prevalent in catheterized urine samples; however, we have few or no isolates from these genera in our collection. Efforts to obtain more aerotolerant anaerobes should be encouraged. Methods for the culture of non-aerotolerant anaerobes in urine samples also should be developed; this would require a method that does not expose the urine sample to air during collection.

The catalog reported here is a substantial extension of a previously published study [14]. In that previous study, we analyzed 149 bladder genomes from 78 different species. In the current study, we analyzed 1134 genomes from 196 species (Fig. 1E, Additional file 1: Fig. S1), including 7 species without representative genomes in the GTDB taxonomy database. With more comprehensive collections of genomes from the bladder, vagina, and gut, we found the functions and taxonomies of bladder isolates to be more like vaginal isolates than of gut isolates (Fig. 2), consistent with the previous report [14]. For example, both studies observed enrichment in the bladder of the mevalonate-dependent pathway for isoprenoid biosynthesis over the more common methylerythritol 4-phosphate pathway. This enrichment primarily results from the presence of certain bladder-specific species in the genera *Aerococcus*, *Actinotignum*, *Bifidobacterium*, *Corynebacterium*, *Enterococcus*, *Lacticaseibacillus*, *Lactobacillus*, *Limosilactobacillus*, *Staphylococcus*, and *Streptococcus*. Other bladder-enriched functions include several related to acetyl-CoA, the precursor to the mevalonate-dependent pathway, and several involved in iron utilization and export. The latter is likely related to the scarcity of iron in the bladder. The reason for the enrichment of functions related to lysine metabolism is less obvious.

Since the genomes of all the isolates from the bladder and gut differed substantially, we wondered whether the same might be true of a single species found in both niches. We chose *E. coli* for three reasons: (1) it is the most common cause of UTI, (2) it is the most fully characterized bacterial species, and (3) it was the species with the largest number of sequenced genomes from the bladder. Our analysis identified both phylogenetic and functional differences. For example, the bladder and gut genomes belonged to the same 6 phylogroups; however, their distribution differed with phylogroup B2 most common in the bladder, but A, D, and F were most common in the gut. In fact, phylogroup B2 was relatively rare (3%) in the asymptomatic gut samples examined here. Previous investigation of *E. coli* phylogroup diversity in the gut found the presence of B2 strains most prevalent in individuals with inflammatory bowel disease [49]. Prior studies suggested that most *E. coli* strains causing UTI and other extraintestinal infections belong to phylogroup B2, presumably due to its greater pathogenetic capacity relative to the other

phylogroups [50–53]. However, we did not find an association between symptom status and phylogroup, which concurs with prior observations in the literature [54]. It is important to note that most (74%) bladder *E. coli* isolates were collected from participants diagnosed with UTI or RUTI; future work isolating *E. coli* from the bladders of individuals with the other symptom statuses considered here is needed. Although most bladder *E. coli* isolates collected from participants diagnosed with UTI or RUTI belonged to phylogroup B2, this also was true for bladder isolates from asymptomatic controls. Further sequencing of *E. coli* strains from asymptomatic controls, however, is needed as our study only included 12 strains. Nevertheless, these results support the hypothesis that UTI symptoms most likely result from multiple factors, including the composition of the rest of the bladder microbiome and the host response [55–59].

The difference in phylogroup distribution suggests that only a subset of gut *E. coli* strains are adapted for life in the bladder. From enrichment analysis, it appears that *E. coli* uses different strategies to adapt to and colonize these two environmentally different niches (Fig. 4). Bladder-enriched functions included those associated with iron metabolism, hippuric acid hydrolase, and a uric acid transporter. Given the scarcity of iron and the presence of hippuric and uric acids in the bladder [60], this makes sense. In contrast, gut-enriched functions related to hydrogenase components, the methylaspartate cycle, phenylpropionate metabolism, putrescine metabolism, and quorum sensing. These are functions expected of strains adapted for survival in the highly crowded, anaerobic environment of the gut. The observation that the bladder genomes were enriched for type IV and VI secretion systems, as well as intracellular multiplication and trafficking proteins, whereas the gut genomes were enriched for type III secretion systems also suggests very different lifestyles.

### Strengths and limitations

This study has several strengths. First, we sequenced and analyzed only isolates from urine obtained by transurethral catheterization, which is known to sample urine from the bladder [10, 61]. In contrast, most studies attempting to inventory bacterial species of the urinary tract have been performed on voided urine. A prime example is a recent study by Dubourg and colleagues [62], who used their own enhanced culture method to detect microbes in voided urine. These authors claimed to have increased the microbial repertoire of the urinary tract and concluded that many urinary species originate from the gut; however, they should have restricted their conclusions to urine and not the urinary tract. This is because voided urine is often contaminated with post-urethral microbes, and thus, the origin of each bacterial isolate from a voided urine sample cannot be determined [56]. In contrast, because we sequenced only isolates from catheterized urine, our collection can be considered a bladder-specific bacterial genome reference catalog. Second, while our collection is limited by our ability to culture, EQUIC detects the vast majority of relatively prevalent bladder bacterial species detected by either 16S rRNA gene sequencing [3, 44] or shotgun metagenomic sequencing [14, 28]. Third, whole-genome sequencing provides a collection of high-quality genomes for the phylogenetic diversity found within the bladder, permitting exploration beyond what can be achieved by 16S rRNA gene sequencing. For example, we used some of these genomes to determine that the species *Aerococcus urinae* is actually a complex of at least

four species and two groups [20, 63]. Future metagenomic studies of the bladder microbiome will necessitate such a catalog of genomes, as most metagenomic analysis tools rely on reference databases. The availability of a bladder genome catalog, akin to the gut microbiome genome catalog, will enable future studies to use shotgun sequencing to capture complex communities.

We acknowledge that this study has limitations. First, the vast majority of the sequenced bacteria were obtained from adult female participants and thus cannot be considered to represent males or children. However, despite including only 19 males in this collection, we identified 12 (out of 196) bacterial species found exclusively in males. This indicates the unique composition of the male bladder microbiome and the necessity of including more male samples in our next collection. Second, the study population mostly consists of US participants. Notably, our previous observations revealed distinct differences in the bladder microbiomes between Chinese and US participants, indicating a possible geographic variation in the bladder microbiome [45]. An ideal collection should include isolates from different countries. Third, the bladder, vagina, and gut genomes were obtained from different participants. An ideal design would be to compare the genomes of isolates obtained from all 3 niches from the same cohort of participants. Alas, that collection does not yet exist. Finally, our comparisons between the bladder, vagina, and gut species are—with the exception of *E. coli*—limited to a single strain per species, which likely excludes some of the genetic diversity present in the species. Currently, however, many of these species have few sequenced genomes. The observations presented here can only be validated once more sequences are available from bladder and vaginal strains.

## Conclusions

Comparisons between genomes isolated from the bladder, vagina, and gut provide evidence that the genetic content of bacteria that inhabit the bladder is distinct, most notably by the functionalities they encode. With the genomes produced through this study, a more comprehensive catalog of bacteria species isolated from the bladder is now available, representative of >70% of bladder genera detected via high throughput 16S rRNA gene or shotgun metagenomic sequencing surveys. Most notably, the 1134 genomes examined here capture the genetic diversity of key bladder species associated with both the presence and absence of symptoms. The genomes, coupled with the participant metadata, provide a key reference for future hypothesis- and data-driven research into the bladder microbiota.

## Methods

### Study design and sample collection

Following Institutional Review Board approval from Loyola University Medical Center (LUMC), University of California San Diego (UCSD) Health, University of Iowa, University of California San Francisco (UCSF) Medical Center, Nationwide Children's Hospital (NCH), or University of Pittsburgh (Pitt) School of Medicine, participants gave verbal and written consent for chart abstraction and urine collection with analysis for research purposes. Urine samples were collected via transurethral catheter. Urine samples were placed into BD Vacutainer<sup>®</sup> plus C&S preservative tubes (Becton, Dickinson and Co.,

Franklin Lakes, NJ). All specimens from LUMC were transported within 4 h to the Wolfe Lab at Loyola University Chicago. Specimens from elsewhere were shipped overnight.

#### **Expanded quantitative urine culture (EQUC)**

EQUC was performed as previously described [7, 8]. Briefly, 100  $\mu$ L of catheterized urine was grown under five conditions with BD BBL-prepared plated media: (1) blood agar plate (BAP) in 5% CO<sub>2</sub> for 48 h, (2) chocolate agar (CHOC) in 5% CO<sub>2</sub> for 48 h, (3) colistin and nalidixic acid (CNA) agar in 5% CO<sub>2</sub> for 48 h, (4) CDC anaerobe BAP in an anaerobic jar (BD GasPak Anaerobe Sachets) or anaerobic chamber (Coylabs) for 48 h, and (5) BAP under aerobic conditions for 48 h, and MacConkey under aerobic or 5% CO<sub>2</sub> for 48 h, all at 36 °C. In one study (FUN), EQUC was modified to select for fungal microbes as well: BD BBL-prepared plated media (1) brain heart infusion agar, with 10% sheep blood, gentamicin, and chloramphenicol gentamicin-BHI in 5% CO<sub>2</sub>, (2) Hardy Chrome-Candida (Chrom-Candida), and (3) inhibitory mold agar (IMA) for 120 h, all at 36 °C. The detection level was 10 CFU/mL, represented by one colony of growth on any of the plates. In another study (PBSA), EQUC was modified to include BD BBL plate media and Thayer Martin agar for 48 h at 36 °C. Each morphologically distinct colony type was isolated on a different plate of the same medium to prepare a pure culture that was used for identification using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS).

#### **Choice of isolates for sequencing**

To assess genomes across as much of the phylogenomic spectrum as possible, we selected at least one member from each species identified by MALDI-TOF for which we had at least 1 isolate. For some of the species commonly found in either symptomatic or asymptomatic individuals, we chose more than 20 isolates. Given the limited phenotypic or genomic data for most species, the selection process was largely random. For *E. coli*, we chose 186 isolates to ensure that we captured several isolates from each phylogroup.

#### **DNA extraction and whole-genome sequencing**

A total of 342 strains from our collection were sequenced and reported previously, as described [14, 54, 64–97]. The remaining 708 strains were sequenced as part of this current effort. Their isolates were grown in their preferred medium and pelleted. To extract genomic DNA, cells were resuspended in 0.5 mL DNA extraction buffer (20 mM Tris–Cl, 2 mM EDTA, 1.2% Triton X-100, pH 8) followed by the addition of 50  $\mu$ L lysozyme (20 mg/mL) and 30  $\mu$ L mutanolysin (5 kU/mL, resulting in 0.15 kU/sample). After a 1-h incubation at 37 °C, 80  $\mu$ L 10% SDS and 20  $\mu$ L proteinase K were added, followed by a 2-h incubation at 55 °C. From this point, genomic DNA was either purified with phenol–chloroform or the MagMax DNA Multi-Sample Kit, according to the manufacturer's instructions. For phenol–chloroform extractions, 210  $\mu$ L of 6 M NaCl and 700  $\mu$ L phenol–chloroform were added. After a 1-h incubation with rotation, the solution was centrifuged at 13,500 rpm for 10 m, and the aqueous phase was extracted. An equivalent volume of isopropanol was added; after a 10-m incubation, the solution was centrifuged at 13,500 rpm for 10 m. The supernatant was decanted, and the DNA pellet precipitated using 600  $\mu$ L 70% ethanol. Following ethanol evaporation, the DNA pellet

was resuspended in nuclease-free H<sub>2</sub>O and stored at −20 °C. For both approaches, DNA purity and quality were spot-checked with a NanoDrop spectrophotometer. One percent agarose gel electrophoresis was performed in difficult-to-extract isolates to confirm genomic DNA isolation and assess degradation. DNA was quantified with the Qubit Fluorimeter Broad Range or with High Sensitivity Kits, depending on yield.

To sequence, the samples were normalized to a maximum of 16 ng/μL. Most were library-prepped with the Illumina Nextera Flex library prep kit with Nextera XT Indices, but 193 samples were prepared using the Qiagen FX library prep kit with Qiagen Indices. Libraries were quantified with qubit, size distribution assessed with Agilent Bioanalyzer HS kit, and were pooled. A quality control PE150 MiSeq flow cell was run on the pools. Successful pools were sent to the Northwestern University Core Sequencing Facility, where they were sequenced on a Novaseq 6000 yielding PE150 reads for an approximate target of 50 × coverage.

### Whole-genome sequence analysis and annotation

Raw reads of the 342 previously sequenced isolates were downloaded from the NCBI Sequence Read Archive ( $n=240$ ) and the European Nucleotide Archive ( $n=102$ ). With the addition of the 708 newly sequenced isolates for this study, a total of 1050 isolates with whole-genome sequences were used for analysis (Additional file 2: Table S1). The quality of the raw reads was assessed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Then, the raw reads were trimmed and filtered using BBDMap in BBTools (v38.94) (<https://jgi.doe.gov/data-and-tools/bbtools/>). Adaptors in samples were first removed using reference (adapters.fa) and trimmed using default settings ( $ktrim=r$ ,  $k=23$ ,  $mink=11$ ,  $hdist=1$ ,  $tpe$ ,  $tbo$ ). Bases with low-quality scores ( $qtrim=rl$ ,  $trimq=20$ ) and positions with high compositional bias ( $ftl=20$ ,  $ftr=135$ ) were removed from both ends, keeping only reads with lengths above 30 bp ( $minlength=30$ ), with no Ns ( $maxns=0$ ), and with an average quality above 20 ( $maq=20$ ). After quality control, all the clean paired-end reads were assembled using SPAdes (v3.14.1) [98] specifying the (−isolate) mode with a full  $k$ -mer size list (− $k$  21,33,55,77). To assess the completeness and contamination of the assembled genomes, CheckM (v1.0.12) [99] was performed using the “lineage\_wf” pipeline, and genomes were filtered at completeness  $\geq 90\%$  and contamination  $\leq 5\%$  to obtain high-quality genomes. MaxBin 2.0 (v2.2.7) [100] was used to filter samples with contamination above 5%, using the contigs’ coverage and the full marker gene set to obtain high-quality genome bins. Quast (v5.2.0) [101] was performed to examine the metrics of assemblies, such as N50, L50, total number of contigs, total length of the assembly (bp), and GC content (Additional file 2: Table S1). The clean reads from each sample were mapped to its high-quality assemblies to estimate the coverage of contigs in assembled genomes using BBDMap, and the mean coverage for each genome was calculated using samtools (<https://github.com/samtools/samtools>). The taxonomy of assembled genomes was classified by gtdbtk (v2.1.1) [102] with the database (release207\_v2) in the “classify\_wf” mode. Additional file 2: Table S1 lists the taxonomy for each of the sequenced strains, including the taxonomies from both the Genome Taxonomy Database (GTDB) and NCBI Taxonomy Database.



Based on statistical metrics of the assemblies, a single genome with relatively higher quality for each species identified by gtdbtk (v2.1.1) [102] was selected from the 1134 high-quality isolated genomes; representative genomes were selected based on assembly quality. In brief, we chose genomes with high completeness, low contamination, few contigs, and/or high coverage depth, prioritizing these metrics in this order. dRep (v2.2.3) [103] and fastANI (v1.32) [104] were used to compare and calculate the average nucleotide identity (ANI) of the representative genomes of the 196 species. Ninety-five percent ANI and 96% ANI cutoffs were used to cluster genomes into different groups (-pa 0.95 -sa 0.96 -S\_algorithm fastANI). Phylogenomic analysis of the representative genomes of bladder strains was conducted using anvi'o v7 [105]; assemblies were made into the anvi'o contig databases (anvi-gen-contigs-database) and populated with hmms (anvi-run-hmms). The multiple sequence alignment of the concatenated amino acid sequences for 71 universal single-copy marker genes was extracted via the anvi'o command `anvi-get-sequences-for-hmm-hits -get-aa-sequences -hmm-source Bacteria_71 -return_best-hit -concatenate`. Phylogenetic trees were constructed using FastTree v2 [106] with default settings (JTT + CAT model) and visualized in the iTOL v6 web browser [107]. Genomes were single annotated using the Database of Clusters of Orthologous Genes (COGs 2020) [108] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [109] in anvi'o v7 [105].

#### **Comparison of the bladder-, vaginal-, and gut-isolated genomes**

Since our bladder-isolated genomes were mostly collected from US participants, we only compared them to genomes of isolates from the vaginas or guts of asymptomatic individuals in Western and European populations. The reference collection of gut-isolated genomes was downloaded from the Broad Institute-OpenBiome Microbiome Library [43] (BIO-ML, bioproject PRJNA544527,  $n = 3423$ ) and a previously published study [110] (European Nucleotide Archive under accession number ERP012217,  $n = 152$ ). The vaginal genomes were mostly collected from the Human Microbiome Project [111] (including only vaginal genomes with metadata explicitly indicating collection from the vagina,  $n = 92$ ) and our laboratory's collection of vaginal isolates ( $n = 5$ ).

The statistical metrics of the assemblies were conducted in the same manner, as described above for bladder genomes. The GTDB-Tk (v2.1.1) was used for taxonomical identification of the reference genomes, classifying 175 species for gut genomes and 74 species for vaginal genomes, respectively. Due to the different number of isolated strains from the 3 niches, for the purpose of niche comparison, we selected a single genome with relatively high quality for each species as the representative genome, based on statistical metrics of the assemblies. The selection criteria were the same as described above for bladder genomes. We used the selected genomes for functional annotation and niche comparison (Additional file 2: Table S4). The pan-genomes of the bladder, vaginal, and gut representative genomes were generated using the parameter (anvi-pan-genome) with frags (-minbit 0.6 and -mcl-inflation 2). Using these pan-genomes, enrichment analysis for KEGG and COG functions was then conducted in anvi'o v7 with the parameter (anvi-compute-functional-enrichment) [112].

### Microbial 16S rRNA gene amplicon sequence analysis

16S rRNA gene sequence analysis of previously obtained 392 female bladder urine samples [3, 5, 38–40] was conducted, as previously described [44, 45]. Cutadapt (cutadapt.readthedocs.io) was used to quality trim the raw reads derived from the 16S rRNA V4 region by removing adaptors from both ends (-a ^GTGCCAGCMGCCGCGGTAA...ATTAGAWACCCBDGTAGTCC -A ^GGACTACHVGGGTWTCTAAT...TTACCGCGGCKGCTGGCAC) and discarding processed reads that were shorter than 200 bp or longer than 240 bp (-m 200 -M 240 -discard-untrimmed). To generate amplicon sequence variant (ASV) tables, DADA2 [113] was applied to process the trimmed reads, including quality control, dereplication, and chimera removal. A total of 392 bladder urine samples with > 1000 ASV counts were kept for downstream analysis. We then used BLCA v2.2 [114] with the NCBI 16S Microbial Database to obtain taxonomic identities at the genus level.

### Metagenomics sequence analysis

Raw sequencing reads of metagenomic sequences obtained from 42 urine samples collected using catheters in 2 previous studies [28, 31] were downloaded from NCBI (30 samples from BioProject PRJEB8104 and 12 samples from BioProject PRJNA700071). Quality control of metagenomes was performed similarly to whole-genome sequence analysis, as described above. An additional step was taken to eliminate human reads from the samples. This was achieved by mapping the reads against the human genome (GRCh38.p14, GCF\_000001405.40) and removing reads with a minimum identity of 95% (minid = 0.95) using BBmap (bbmap.sh). After quality control, all the clean paired-end reads were processed through MetaPhlan (v 4.0.5) [115] with the Bowtie database (mpa\_vJun23\_CHOCOPPhlanSGB\_202307) using default settings to identify taxonomy and estimate the relative abundances of discrete taxa. The Python script (sgb\_to\_gtdb\_profile.py) in MetaPhlan4 was utilized to transform the output of MetaPhlan4 (SGBs) to the corresponding GTDB taxonomy. Eight samples with no identified bacteria were subsequently removed. Results were parsed via Python and displayed in Additional file 2: Table S3.

### Phylogenetic grouping and virulence profiling of *E. coli* strains

We assembled and annotated publicly available gut *E. coli* isolates [43] following the protocol listed above for the bladder isolates. *E. coli* strains of bladder and gut with high coverage (>20×) were used for downstream pan-genome and phylogenetic analyses. The phylogroups of *E. coli* strains were determined using ClermonTyping [116]. anvio v7 was applied for the pan-genome analysis of *E. coli* strains using the parameter (anvi-pan-genome) with frags (-minbit 0.6 and -mcl-inflation 10). Phylogenetic analysis for bladder *E. coli* strains was conducted by extracting amino acid sequences of 1084 single-copy core clusters (201,624 genes); amino acid sequences were concatenated and aligned (-concatenate-gene-clusters) in anvio v7. The phylogenetic tree was constructed and visualized as mentioned above. The genomes were screened for virulence factors and antibiotic resistance genes via abricate V.1.0.1 (<https://github.com/tseemann/abricate>)

using the following databases: Virulence Factor Database [117] (VFDB, updated 2021-Mar-27) and Comprehensive Antibiotic Resistance Database [118] (CARD, updated 2021-Mar-27). The results were parsed via Python.

### Statistical analyses

All the statistical analyses were conducted using RStudio v3.6. The “ggplot2” R package was used for box plots, and “pheatmap” was used for heatmaps. Discriminant analysis of principal components (DAPC) was conducted using the “adegenet” package using the “dapc” function. Principal coordinates analysis (PCoA, function “capscale” with no constraints applied) was conducted using the “vegan” package and the `vegdist` function with method “bray” for Bray-Curtis dissimilarity analysis. Different groups in the ordination plot were tested using a PERMANOVA test (function “adonis”) and were clustered on the plot using the function “ordiellipse” with statistics (kind = “sd”, conf = 0.95).

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-024-03216-8>.

**Additional file 1: Fig. S1.** Phylogenomic tree of 1134 bladder isolated genomes. **Fig. S2.** Whole-genome ANI comparison of the 196 different GTDB-identified bacterial species in the bladder isolated genomes. **Fig. S3.** Phylogenomic tree of 169 bacterial species (assigned by GTDB) represented in the bladder genome collection from the symptomatic group. **Fig. S4.** Phylogenomic tree of 76 bacterial species (assigned by GTDB) represented in the bladder genome collection from the asymptomatic group. **Fig. S5.** COG function comparisons of bladder isolated genomes with vaginal and gut isolated genomes. **Fig. S6.** COG functions enriched in the bladder based on COG category.

**Additional file 2: Table S1.** Metadata for bladder-specific isolated genomes and hosts information. **Table S2.** Relative abundances of different genera in 16S sequencing samples. **Table S3.** Relative abundances (100%) of different GTDB-classified bacterial taxonomy in 34 metagenomics sequencing samples. **Table S4.** List of GTDB classified species isolated from gut, vaginal, and bladder samples. A single genome was selected for each species from the gut, vaginal, and bladder strains. **Table S5.** KEGG functions of bladder, gut, and vaginal genomes. **Table S6.** COG20 functions of bladder, gut, and vaginal genomes. **Table S7.** Statistical analyses of KO modules in bladder, vaginal, and gut genomes. KO module enrichment table was generated using `anvi'o` v7. **Table S8.** Statistical analyses of KOfam hits in bladder, vaginal, and gut genomes. KOfam enrichment table was generated using `anvi'o` v7. **Table S9.** Statistical analyses of COG20 functions in bladder, vaginal, and gut genomes. COG20 function enrichment table was generated using `anvi'o` v7. **Table S10.** Phylogroup distribution of bladder and gut *E. coli* strains. **Table S11.** KOfam profiles of bladder *E. coli* strains. **Table S12.** KOfam profiles of gut *E. coli* strains. **Table S13.** Differentially abundant KOfam profiles between bladder and gut *E. coli* isolates by Wilcoxon rank sum test. **Table S14.** Virulence factor genes identified in bladder *E. coli* isolate assemblies. Number listed is number of copies of the gene. **Table S15.** Antibiotic resistance genes identified in bladder *E. coli* isolate assemblies. Number listed is number of copies of the gene. **Table S16.** Virulence factor genes identified in gut *E. coli* isolate assemblies. Number listed is number of copies of the gene. **Table S17.** Antibiotic resistance genes identified in gut *E. coli* isolate assemblies. Number listed is number of copies of the gene. **Table S18.** Differentially abundant virulence and antibiotic resistance genes between bladder and gut *E. coli* isolates by Wilcoxon rank sum test.

**Additional file 3.** Review history.

### Acknowledgements

We wish to acknowledge the study participants who consented to donate urine and the clinicians who recruited those participants and collected their urine. We especially wish to acknowledge the clinical members of the Loyola Urinary Education and Research Collaborative (LUEREC), who obtained the vast majority of urines associated with this study. We thank all the members of the Wolfe laboratory who processed the urine samples to identify isolates and Genevieve Baddoo for the isolation of a few of the strains included in this study. Finally, we wish to thank Paul Schreckenberger (deceased) and Amanda Harrington, former and current directors of the Loyola University Clinical Microbiology Laboratory, respectively, for their willingness to share their MALDI-TOF instruments.

### Peer review information

Yanbin Yin and Kevin Pang were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 3.

### Authors' contributions

CP and AJW conceived of the study. TH, CP, and AJW designed the study. MK, TH, RL, LT, and BRH-K prepared the isolates and/or sequenced the genomes. JD, AE, CP, and AJW analyzed and interpreted the data. JD and MFN managed the sequencing data. JD took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

### Funding

This work was funded by multiple sources, including R01DK104718 from the National Institute of Health, a contract from Pathnostics, and a gift from an anonymous donor.

### Availability of data and materials

Raw data of whole-genome sequences for isolates and assemblies of isolates are available in the Sequence Read Archive under BioProject PRJNA970254 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA970254>) [119] and BioProject PRJNA316969 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA316969>) [120]. The 1134 bladder-specific isolated bacterial genomes and 387 gut-specific isolated *E. coli* genomes used in the study are also available in figshare ([https://figshare.com/articles/journal\\_contribution/Cataloging\\_the\\_phylogenetic\\_diversity\\_of\\_human\\_bladder\\_bacterial\\_isolates/25308346](https://figshare.com/articles/journal_contribution/Cataloging_the_phylogenetic_diversity_of_human_bladder_bacterial_isolates/25308346)) [121]. The bacterial isolates used in the article are also available from the corresponding authors upon reasonable request. Additional file 2: Table S1 includes taxonomy, assembly statistics, study information (participant number, study code, sample type), and participant information (age, gender, race, BMI, and symptom status).

### Declarations

#### Ethics approval and consent to participate

The isolates sequenced in this study were from an IRB-approved biorepository (LU215192). All experimental methods comply with the Helsinki Declaration.

#### Consent for publication

Not applicable.

#### Competing interests

AJW discloses advisory board membership for Pathnostics and Urobiome Therapeutics, funding from Pathnostics, the Craig Neilsen Foundation, and an anonymous donor. The other authors declare that they have no competing interests.

Received: 2 October 2023 Accepted: 14 March 2024

Published online: 21 March 2024

### References

1. Pearce MM, et al. The female urinary microbiome in urgency urinary incontinence. *Am J Obstet Gynecol.* 2015;213(347):e341–e311.
2. Brubaker L, et al. Urinary bacteria in adult women with urgency urinary incontinence. *Int Urogynecol J.* 2014;25:1179–84.
3. Pearce MM, et al. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *MBio.* 2014;5:e01283–e1214.
4. Karstens L, et al. Does the urinary microbiome play a role in urgency urinary incontinence and its severity? *Front Cell Infect Microbiol.* 2016;6:78.
5. Thomas-White KJ, et al. Incontinence medication response relates to the female urinary microbiota. *Int Urogynecol J.* 2016;27:723–33.
6. Nienhouse V, et al. Interplay between bladder microbiota and urinary antimicrobial peptides: mechanisms for human urinary tract infection risk and symptom severity. *PLoS One.* 2014;9:e114185.
7. Hilt EE, et al. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol.* 2014;52:871–6.
8. Price TK, et al. The clinical urine culture: enhanced techniques improve detection of clinically relevant microorganisms. *J Clin Microbiol.* 2016;54:1216–22.
9. Thomas-White KJ, et al. Evaluation of the urinary microbiota of women with uncomplicated stress urinary incontinence. *Am J Obstet Gynecol.* 2017;216(55):e51–55.e16.
10. Wolfe AJ, et al. Evidence of uncultivated bacteria in the adult female bladder. *J Clin Microbiol.* 2012;50:1376–83.
11. Coorevits L, Heytens S, Boelens J, Claeys G. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. *Eur J Clin Microbiol Infect Dis.* 2017;36:635–9.
12. Fouts DE, et al. Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic bacteriuria in neuropathic bladder associated with spinal cord injury. *J Transl Med.* 2012;10:174.
13. Khasriya R, et al. Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. *J Clin Microbiol.* 2013;51:2054–62.
14. Thomas-White K, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat Commun.* 2018;9:1557.
15. Khan Z, et al. Chronic urinary infection in overactive bladder syndrome: a prospective, blinded case control study. *Front Cell Infect Microbiol.* 2021;11:752275.
16. Torres-Puig S, et al. "Omics" technologies - what have they told us about uropathogenic *Escherichia coli* fitness and virulence during urinary tract infection? *Front Cell Infect Microbiol.* 2022;12: 824039.

17. Lo AW, et al. 'Omic' approaches to study uropathogenic *Escherichia coli* virulence. *Trends Microbiol.* 2017;25:729–40.
18. Rocha J, et al. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. *Int J Syst Evol Microbiol.* 2020;70:1522–7.
19. Vaneechoutte M, et al. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piovii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. *Int J Syst Evol Microbiol.* 2019;69:679–87.
20. Choi BI, et al. Taxonomic considerations on *Aerococcus urinae* with proposal of subdivision into *Aerococcus urinae*, *Aerococcus tenax* sp. nov., *Aerococcus mictus* sp. nov., and *Aerococcus loyolae* sp. nov. *Int J Syst Evol Microbiol.* 2023;73(9):006066.
21. Komesu YM, et al. Defining the relationship between vaginal and urinary microbiomes. *Am J Obstet Gynecol.* 2020;222(154):e151–154.e110.
22. Nardos R, et al. Network-based differences in the vaginal and bladder microbial communities between women with and without urgency urinary incontinence. *Front Cell Infect Microbiol.* 2022;12:759156.
23. Yamamoto S, et al. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. *J Urol.* 1997;157:1127–9.
24. Hooton TM. Recurrent urinary tract infection in women. *Int J Antimicrob Agents.* 2001;17:259–68.
25. Chen SL, et al. Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Sci Transl Med.* 2013;5:184ra160.
26. Thänert R, et al. Comparative genomics of antibiotic-resistant uropathogens implicates three routes for recurrence of urinary tract infections. *mBio.* 2019;10:e01977.
27. Moustafa A, et al. Microbial metagenome of urinary tract infection. *Sci Rep.* 2018;8:4333.
28. Garretto A, Thomas-White K, Wolfe AJ, Putonti C. Detecting viral genomes in the female urinary microbiome. *J Gen Virol.* 2018;99:1141–6.
29. Adebayo AS, et al. The urinary tract microbiome in older women exhibits host genetic and environmental influences. *Cell Host Microbe.* 2020;28:298–305.e293.
30. Janes VA, et al. Metagenomic DNA sequencing for semi-quantitative pathogen detection from urine: a prospective, laboratory-based, proof-of-concept study. *Lancet Microbe.* 2022;3:e588–97.
31. Adu-Oppong B, Thänert R, Wallace MA, Burnham CD, Dantas G. Substantial overlap between symptomatic and asymptomatic genitourinary microbiota states. *Microbiome.* 2022;10:6.
32. Thänert R, et al. Persisting uropathogenic *Escherichia coli* lineages show signatures of niche-specific within-host adaptation mediated by mobile genetic elements. *Cell Host Microbe.* 2022;30:1034–1047.e1036.
33. Wehedy E, Murugesan S, George CR, Shatat IF, Al Khodor S. Characterization of the urinary metagenome and virome in healthy children. *Biomedicines.* 2022;10:2412.
34. Neugent ML, et al. Recurrent urinary tract infection and estrogen shape the taxonomic ecology and function of the postmenopausal urogenital microbiome. *Cell Rep Med.* 2022;3:100753.
35. Lozano GL, et al. Introducing THOR, a model microbiome for genetic dissection of community behavior. *mBio.* 2019;10:e02846.
36. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 2009;106:19126–31.
37. Ciufo S, et al. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol.* 2018;68:2386–92.
38. Thomas-White KJ, et al. Urinary microbes and postoperative urinary tract infection risk in urogynecologic surgical patients. *Int Urogynecol J.* 2018;29:1797–805.
39. Fok CS, et al. Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. *Int Urogynecol J.* 2018;29:1765–71.
40. Price TK, et al. The urobiome of continent adult women: a cross-sectional study. *BJOG.* 2020;127:193–201.
41. Katouli M. Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. *Iran J Microbiol.* 2010;2:59–72.
42. Terlizzi ME, Gribaudo G, Maffei ME. UroPathogenic *Escherichia coli* (UPEC) infections: virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial strategies. *Front Microbiol.* 2017;8:1566.
43. Poyet M, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat Med.* 2019;25:1442–52.
44. Storm DW, et al. A child's urine is not sterile: a pilot study evaluating the pediatric urinary microbiome. *J Pediatr Urol.* 2022;18:383–92.
45. Liu F, et al. The bladder microbiome, metabolome, cytokines, and phenotypes in patients with systemic lupus erythematosus. *Microbiol Spectr.* 2022;10:e0021222.
46. Ugarcina Perovic S, et al. Urinary microbiome of reproductive-age asymptomatic European women. *Microbiol Spectr.* 2022;10:e0130822.
47. Lane G, et al. Distinguishing features of the urinary bacterial microbiome in patients with neurogenic lower urinary tract dysfunction. *J Urol.* 2022;207:627–34.
48. Bučević Popović V, et al. The urinary microbiome associated with bladder cancer. *Sci Rep.* 2018;8:12157.
49. Kotlowski R, Bernstein CN, Sepelri S, Krause DO. High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut.* 2007;56:669–75.
50. Nowrouzian FL, Wold AE, Adlerberth I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J Infect Dis.* 2005;191:1078–83.
51. Dadi BR, et al. Distribution of virulence genes and phylogenetics of uropathogenic *Escherichia coli* among urinary tract infection patients in Addis Ababa. Ethiopia *BMC Infect Dis.* 2020;20:108.
52. Rezaatofghi SE, Mirzarazi M, Salehi M. Virulence genes and phylogenetic groups of uropathogenic *Escherichia coli* isolates from patients with urinary tract infection and uninfected control subjects: a case-control study. *BMC Infect Dis.* 2021;21:361.

53. Halaji M, et al. Phylogenetic group distribution of uropathogenic *Escherichia coli* and related antimicrobial resistance pattern: a meta-analysis and systematic review. *Front Cell Infect Microbiol.* 2022;12:790184.
54. Garretto A, et al. Genomic survey of *E. coli* from the bladders of women with and without lower urinary tract symptoms. *Front Microbiol.* 2020;11:2094.
55. Garretto A, et al. Genomic survey of *E. coli* from the bladders of women with and without lower urinary tract symptoms. *Front Microbiol.* 2020;11:2094.
56. Brubaker L, Wolfe AJ. The female urinary microbiota, urinary health and common urinary disorders. *Ann Transl Med.* 2017;5:34.
57. Eto DS, Jones TA, Sundsbak JL, Mulvey MA. Integrin-mediated host cell invasion by type 1-piliated uropathogenic *Escherichia coli*. *PLoS Pathog.* 2007;3:e100.
58. Ragnarsdóttir B, Lutay N, Grönberg-Hernandez J, Köves B, Svanborg C. Genetics of innate immunity and UTI susceptibility. *Nat Rev Urol.* 2011;8:449–68.
59. Schreiber HL, et al. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. *Sci Transl Med.* 2017;9:eaaaf1283.
60. Subashchandrabose S, Mobley HL. Back to the metal age: battle for metals at the host-pathogen interface during urinary tract infection. *Metallomics.* 2015;7:935–42.
61. Brubaker L, et al. Forming consensus to advance urobiome research. *mSystems.* 2021;6:e0137120.
62. Dubourg G, et al. Deciphering the urinary microbiota repertoire by culturomics reveals mostly anaerobic bacteria from the gut. *Front Microbiol.* 2020;11:513305.
63. Gilbert NM, et al. A mouse model displays host and bacterial strain differences in *Aerococcus urinae* urinary tract infection. *Biol Open.* 2021;10:bio058931.
64. Putonti C, et al. Genome investigation of urinary *Gardnerella* strains and their relationship to isolates of the vaginal microbiota. *mSphere.* 2021;6:e00154.
65. Temelkova S, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Streptococcus anginosus* UMB1296, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00409.
66. Scott T, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Streptococcus anginosus* UMB0839, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00411.
67. Rivera I, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus jensenii* UMB0836, isolated from the female bladder. *Microbiol Resour Announc.* 2020;9:e00410.
68. Miller S, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Streptococcus anginosus* UMB7768, isolated from a woman with recurrent UTI symptoms. *Microbiol Resour Announc.* 2020;9:e00418.
69. Markovic T, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Actinomyces neuii* UMB1295, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00402.
70. Ali Khan A, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Corynebacterium coyleae* UMB8490, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00393.
71. Miller-Ensminger T, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus jensenii* strain UMB7766, isolated from the female bladder. *Microbiol Resour Announc.* 2020;9:e00392.
72. Allen O, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Proteus mirabilis* UMB1310, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00390.
73. Balce E, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Klebsiella pneumoniae* UMB7783, isolated from the female bladder. *Microbiol Resour Announc.* 2020;9:e00390.
74. Belmonte M, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Escherichia coli* UMB9246, isolated from the bladder of a woman with recurrent urinary tract infection. *Microbiol Resour Announc.* 2020;9:e00437.
75. Tsihere O, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus mulieris* UMB7784, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00403.
76. Ramotowski L, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus mulieris* UMB9245, isolated from the female bladder. *Microbiol Resour Announc.* 2020;9:e00408.
77. Khan F, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus crispatus* UMB1163, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00404.
78. Kemper L, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Staphylococcus epidermidis* UMB7765, isolated from the urobiome of a woman with recurrent urinary tract infection. *Microbiol Resour Announc.* 2020;9:e00417.
79. Kalski M, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Enterococcus faecalis* UMB7780, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00400.
80. Gondi S, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Klebsiella pneumoniae* UMB8492, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00405.
81. Gallian N, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Proteus mirabilis* strain UMB0038, isolated from the female bladder. *Microbiol Resour Announc.* 2020;9:e00401.
82. Eskandar S, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Corynebacterium aurimucosum* UMB7769, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00391.
83. Crawford T, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Escherichia coli* UMB1353, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00416.
84. Bhimalli P, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Klebsiella pneumoniae* UMB7779, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00396.
85. West-Pelak E, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Lactobacillus jensenii* UMB0847, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00395.
86. Schwartz A, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Enterococcus faecalis* UMB1309, isolated from catheterized urine. *Microbiol Resour Announc.* 2020;9:e00406.
87. Salgado LM, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Streptococcus agalactiae* UMB7782, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00423.

88. Truckenbrod A, Miller-Ensminger T, Voukadinova A, Wolfe AJ, Putonti C. Draft genome sequence of *Staphylococcus epidermidis* UMB8493, isolated from the female urinary tract. *Microbiol Resour Announc.* 2020;9:e00419.
89. Putonti C, et al. Draft genome sequences of six *Lactobacillus gasseri* and three *Lactobacillus paragasseri* strains isolated from the female bladder. *Microbiol Resour Announc.* 2019;8:e00973.
90. Putonti C, et al. Draft genome sequences of 11 *Lactobacillus jensenii* strains isolated from the female bladder. *Microbiol Resour Announc.* 2019;8:e00970.
91. Hilt EE, Price TK, Diebel K, Putonti C, Wolfe AJ. Draft genome sequence for a urinary isolate of *Nosocomiicoccus ampullae*. *Genome Announc.* 2016;4:e01248-01216.
92. Putonti C, et al. Draft genome sequence of *Staphylococcus epidermidis* (Winslow and Winslow) Evans (ATCC 14990). *Genome Announc.* 2017;5:e00619.
93. Putonti C, et al. Draft genome sequence of *Micrococcus luteus* (Schroeter) Cohn (ATCC 12698). *Genome Announc.* 2017;5:e00576.
94. Price TK, et al. Draft genome sequence of a urinary isolate of *Lactobacillus crispatus*. *Genome Announc.* 2016;4:e01278.
95. Price TK, et al. Genome sequences and annotation of two urinary isolates of *E. coli*. *Stand Genomic Sci.* 2016;11:79.
96. Malki K, et al. Seven bacteriophages isolated from the female urinary microbiota. *Genome Announc.* 2016;4:e01003.
97. Prasad A, et al. Comparative genomic study of *Streptococcus anginosus* reveals distinct group of urinary strains. *mSphere.* 2023;8:e0068722.
98. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics.* 2020;70:e102.
99. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
100. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics.* 2016;32:605–7.
101. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–5.
102. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics.* 2022;38:5315–6.
103. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 2017;11:2864–8.
104. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9:5114.
105. Eren AM, et al. Anvi'o: an advanced analysis and visualization platform for omics data. *PeerJ.* 2015;3:e1319.
106. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5:e9490.
107. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49:W293–6.
108. Galperin MY, et al. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 2021;49:D274–81.
109. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 2023;51:D587–92.
110. Forster SC, et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol.* 2019;37:186–92.
111. Consortium, I.H.I.R.N. The integrative human microbiome project. *Nature.* 2019;569:641–8.
112. Shaiber A, et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol.* 2020;21:292.
113. Callahan BJ, et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581–3.
114. Gao X, Lin H, Revanna K, Dong Q. A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics.* 2017;18:247.
115. Blanco-Míguez A, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol.* 2023;41:1633.
116. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom.* 2018;4:e000192.
117. Chen L, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005;33:D325–328.
118. Alcock BP, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2020;48:D517–25.
119. Du J, Khemmani M, Halverson T, Ene A, Limeira R, Tinawi L, Hochstedler-Kramer B, Noronha M, Putonti C, Wolfe A. Cataloging the phylogenetic diversity of human bladder bacterial isolates. *Datasets. NCBI SRA.* 2024. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA970254>.
120. Du J, Khemmani M, Halverson T, Ene A, Limeira R, Tinawi L, Hochstedler-Kramer B, Noronha M, Putonti C, Wolfe A. Cataloging the phylogenetic diversity of human bladder bacterial isolates. *Datasets. NCBI SRA.* 2024. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA316969>.
121. Du J, Khemmani M, Halverson T, Ene A, Limeira R, Tinawi L, Hochstedler-Kramer B, Noronha M, Putonti C, Wolfe A. Cataloging the phylogenetic diversity of human bladder bacterial isolates. *Datasets. figshare.* 2024. [https://figshare.com/articles/journal\\_contribution/Cataloging\\_the\\_phylogenetic\\_diversity\\_of\\_human\\_bladder\\_bacterial\\_isolates/25308346](https://figshare.com/articles/journal_contribution/Cataloging_the_phylogenetic_diversity_of_human_bladder_bacterial_isolates/25308346).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.