DOI:
10.1016/j.chom.2020.06.022

Document Version
Peer reviewed version

The urinary tract microbiome in older women exhibits host genetic and environmental influences

AS Adebayo[1], G Ackermann[2], RCE Bowyer[1], PM Wells[1], G Humphreys[2], R Knight[2], TD Spector[1], CJ Steves[1]*

[1]Department of Twin Research & Genetic Epidemiology, King's College, London;
[2]Knight Lab, Departments of Pediatrics and Computer Science & Engineering, University of California San Diego

*Corresponding author/Lead Contact: claire.j.steves@kcl.ac.uk

Summary

The urinary microbiome is a relatively unexplored niche that varies with gender. Urinary microbes, especially in ageing populations, are associated with morbidity. We present a large-scale study exploring factors defining urinary microbiome composition in community-dwelling older adult women without clinically active infection. Using 1600 twins, we estimate the contribution of genetic and environmental factors to microbiome variation. The urinary microbiome is distinct from nearby sites and unrelated to stool microbiome with more Actinobacteria, Fusobacteria and Proteobacteria, but fewer Bacteroidetes, Firmicutes and Verrumicrobia. A quarter of variants had heritability estimates greater than 10% with most heritable microbes having potential clinical relevance, including *Escherichia-Shigella* linked to urinary tract infections. Age, menopausal status, prior UTI and host genetics were top factors defining the urobiome with increased microbial diversity tending to associate with older age. These findings highlight the distinct composition of the urinary microbiome and significant contributions of host genetics.

Keywords: microbiome, genetics, urogenital tract, ageing

Introduction

The resident microbial community (microbiome) at different body sites continues to generate research interest, driven by evidence of a role in human physiology. The study of the urinary microbiome (urobiome) is much less established compared to the gut microbiome; perhaps due to the previous belief that the urine was sterile in the absence of a urinary tract infection. Recently, research has shown that this is not the case and that the urinary tract is in fact, another site with a microbiome, reflective of the microbes inhabiting the bladder and closely associated organs (Fouts et al., 2012; Wolfe et al., 2012; Whiteside et al.; 2015). A urinary microbiome has been found in in catheterized/aspirated urine (bladder microbiome) and voided urine, using both

enhanced quantitative cultures and DNA-based identification (16S marker studies/metagenomics), and samples from different populations (e.g Khasriya et al., 2013 ; Hilt et al., 2014; Wu et al., 2017; Adebayo et al., 2017; Kramer et al., 2018).

Studies to date have largely focused on differences in the urobiome in relation to urinary tract conditions (Sihra et al., 2018; Wolfe & Brubaker, 2019) including urinary infections (UTI). There is evidence for differences in the male and female urinary microbiomes (Bajic et al. 2018; Pearce et al., 2014). Women are much more likely to develop UTI, with a lifetime risk of up to 50% (Franco, 2005) compared to 12% for men (Lee & Neild, 2007). UTI is also the most common reason for antibiotic treatment in adult women, which has implications for urinary and other microbiomes and antimicrobial resistance. Early work has indicated that the non-infected state microbiome may influence resilience to infection (Pearce et al., 2015; Thomas-White et al., 2018). The present study is focused on understanding the major factors defining the urobiome in community dwelling women without active infection, who are not seeking clinical assessment.

Previous studies involving urinary/bladder microbiomes have involved relatively small sample sizes (dozens or few hundreds of people) in hospital or clinic attending patients. For instance, results from our literature search (Jan 2015 to September 2018) included incontinence (Pearce et al., 2015, n=182); case-control studies on elderly/non-elderly patients (Liu et al., 2017; n=100); urinary tract infections (Moustafa et al., 2018; n=112,Price et al., 2016, n=150); cancer (Wang et al., 2017; n=65); overactive bladder (Wu et al., 2017; Fok et al., 2018; n=55-126); chronic kidney disease (Kramer et al, 2018; n=41); surgical transplant patients (Rani et al., 2018, n=20); surgery (Thomas-White et al., 2018,n=104); and menopause (Curtiss et al., 2018; n=78). Reinforcing this, a recent review (covering studies up to 2016) carried out by Aragon et al. (2018) reported that the sample sizes in urinary microbiome studies varied between 8 to 60 for healthy controls and 10-197 for cases. Their report shows that many studies are commissioned on incontinence, bladder-related and gynaecologic patients. Recently, Price et al (2019) have studied data from 224 patients who were free from known urinary tract conditions, bladder problems or surgery.

Studies to date have not investigated the contribution of host factors, like age or genetics to the "normal" urinary tract microbiome. Genetic factors have been shown to influence the gut microbiome (Goodrich et al., 2014, Luca et al., 2018), although environmental factors remain predominant (Rothchilds et al.,2018). The urine microbiome has an additional complexity, that is that many samples may be below the threshold for a detectable microbiome. Few studies to date have included study of sequence negative samples; notably Pearce et al. (2015) reported that sequence-negative samples were comparable in many characteristics to sequence-positive ones in incontinent individuals.

We aimed to characterize the host influence on the urinary tract microbiome in women who are well. We analysed midstream urine samples from 1600 older females in the TwinsUK cohort recruited from the community for research, who had no apparent infection, and were not undergoing hospital treatment. We hypothesized that, in an unselected average population, (1) the inherent core urinary bacterial community could be defined (2) that the urobiome is influenced by host-specific genetic and environmental factors, (3) that some host-specific factors may relate to the microbial biomass in the urine.

Results

Urinary tract microbiome was distinct from proximal body sites and similar to other urine samples.

Initially, we compared the overall composition of the urinary microbiome to other body sites , applying a similar analytical pipeline to 4 datasets of women older than 45 years from published studies (Goodrich et al., 2014; Thomas-White et al., 2017; Pearce et al., 2014; Yatsunenko et al. 2012) (Methods, Data S1). Species diversity (Shannon index) was comparable in urine and the vaginal datasets and reduced relative to the stool(Fig 1A). Stool samples in the vast majority ordinated separately from urine samples, and the current study (urine3) was the most dissimilar to gut or vagina samples of the urine studies sampled (Fig 1B, Data S1). Repeating these diversity analyses with 100 randomly chosen samples each available for 3 datasets showed similar results (SFig1A,B). There was no clear correlation in stool and urine

microbiome dissimilarity for paired stool and urine samples from TwinsUK, even when obtained on the same day) (Mantel's $r \leq 0.02$, $p > 0.1$) (Fig 1C-D, Data S1,SFig1C). The urine studies also had some taxa differences (SFig1D).

General description of TwinsUK urinary tract microbiome

Urine samples from 1600 mainly postmenopausal women (mean age= 66.4) in the TwinsUK cohort were analysed, revealing 10955 taxa variants from filtered 16S data. Participant characteristics are shown in Table 1. There was a high level of variability in particular taxa present in an individual, with only 245 (2.2%) variants occurring in at least 5% of samples. The use of a compositionally-sensitive analysis improved the ranking of some abundant taxa as compared to common non-compositional analysis (SFig2A). To highlight potential intra-microbiome relationships, clusters of frequent (present in >20%), co-occuring proportionally-balanced species were predicted, resulting in 61clusters of common variants (hereafter referred to as the core taxa) (Fig2A). There were more Actinobacteria, Fusobacteria and Proteobacteria, but fewer Bacteroidetes, Firmicutes and Verrumicrobia in urine compared to gut microbiome (SFig 2B).

Low read count (no reliably-detected microbiome (<2000 reads post-filtering)) (Data S2) associated with slightly younger age and lower level of health deficit; specifically, a ~20% increase in the chance of a detectable microbiome for unit increase in standardized age ($p=0.0048$, OR=1.21, CI=1.07 - 1.39) and ~14% increase for a unit increase in standardized frailty index (OR=1.144,CI=1.01-1.30,$p=0.0359$). There was no association between low read status and the number of previous Urinary Tract Infections (UTIs), recent antibiotics usage, surgery episodes or number of childbirth episodes (parity). Amplicon concentrations associated with parity (β =1.89,$p=0.0035$) but not other demographics (Data S2).

Host genetics' influences variation of urine microbiome

Various measurements can be useful to detect host genetics effect. First, we used heritability, a quantitative measure of the contribution of additive genetics to variability in a phenotype, in this case the microbiome data. This showed considerable and significant genetic contribution to variance in the first principal coordinate of Bray-Curtis dissimilarities (PCo) (A=0.147) and 2nd PCo (A=0.356). When

accounting for phylogenetic closeness and dominant taxa using weighted unifrac dissimilarities (which captured as much as 57% of the variation), heritability of the first PC was estimated at 18% (A= 0.179, CI=0.05-0.415, p=0.003351; C=0.0049, E=0.8164, n=760). Significant heritability was maintained when controlling for age, history of UTI, menopause status and cohabitation (Data S3). Some clusters of frequent, co-occuring balanced microbial species showed particularly high heritability (Fig 2A).

Second, we used a form of family segregation by applying constrained principal coordinates analysis on the Bray-Curtis dissimilarity, with the family identity as a factor, and then compared dissimilarities in identical and non-identical twins. The dissimilarity was lower for monozygotic twins (Fig 2B) (p=0.0022). The difference in the dispersion within a twin-pair (Euclidean distances to the median) in the unconstrained analysis of Bray-Curtis, was also lower for monozygotic pairs (SFig3A) (p=0.027) (Data S3). The first PCo was also associated with family identity (Kruskal-Wallis p=0.043). Third, we compared ancestral origin of participants. The study population was primarily of British ancestry, with microbiome data available for 1141 British, 27 non British white, 19 SouthEast Asian, and 9 others, and therefore findings would need to be confirmed in other studies. We minimized imbalanced groups' effect by subsampling (n=98-118) using 20-fold partition with resampling of non-British groups, and bootstrapping Kruskal-Wallis statistic. The second PCo of the microbiome diversity (Bray-Curtis) differed according to the 4 major ancestry present (1st PC; p=0.156; 2nd PC p=0.000143, bootstrap averaged p= 0.0081, percentileCI =(0.00003-0.1223),1000 replicates), as was the permanova test of Bray-Curtis dissimilarity between the ancestry groups (Fig. 2C , Data S3).

Finally we analysed the heritability of the core microbiome. The relative abundance of these common variants as a group were influenced by host genetics (heritability 23%(CI=8.77-33.7, C=1.66E-12,E=0.76). Almost a quarter (59 of 245) of variants found in at least 5% of the participants had heritability estimates greater than 10% (STable1, SFig3B), though in five of these confidence intervals span zero. Some of the most heritable variants some sequence similarity (SFig4). One of such heritable variants, *Lactobacillus iners* AB-1 showed phylogenetic relatedness with Christenellaceae variant previously reported in the gut (SFig4). Another,

Escherichia_Shigella (A=0.165) is a potential culprit in urinary tract infection. Based on this finding, we also tested the heritability of occurrence of prior urinary tract infections, finding prior UTI to be significantly heritable (A=0.273(0.399 for high recurrence), 95%CI=0.178 – 0.368, Data S3).
Taken together, these four different sets of analyses support the hypothesis that host genetic factors influence the urinary tract microbiome.

Host-related/environmental factors in urinary tract microbiome, especially age, have important effects

The potential effects of multiple factors were assessed (Fig 3). Age, diet, recent antibiotic usage and overall health deficit were assessed in relation to the urobiome as they are known 'host-specific' influencers of gut microbiome variation. Parity (previous number of births) and surgical history (had previous surgery or not) were assessed as host-related "environmental" factors as they could potentially alter structures in or proximal to the urinary tract. Previous history of UTI was also assessed.

With increasing age, there is overall increase in alpha diversity (Shannon) (Table 1), which was robust to uneven sample sizes or exclusion of small number of participants aged <50 ($0.10 \geq \beta \leq 0.22$, $0.00027 \leq p \geq 0.0045$). Age differed with beta diversity estimates ($p<0.001$), and was a main influencer of the ordination patterns of samples (Fig 3B). The core taxa and one-third (22) of the subclusters, differed with age ($1.92E\text{-}30 \leq FDR \leq 0.046$).

Diet (Healthy Eating Index), health deficit (frailty index) and antibiotics usage did not produce significant associations in alpha diversity (Shannon) but borderline associations were found with changes in beta diversity (Bray-Curtis) (diet, p=0.052, n=1004; recent antibiotics usage, p=0.041,n=992; health deficit, p=0.031, n=1139). Parity trended toward an association with reduced alpha diversity (Shannon) (p=0.058,n=1047), and was significantly associated with beta diversity (Bray-Curtis) (p=0.026,n=1047); surgical history did not differ with Bray-Curtis or Shannon metrics (n=540). Occurrence of UTI differed with alpha diversity (p=0.0027) and beta diversity (p=0.001). Similar results were obtained using unweighted unifrac sample distances or controlling for other factors.

The contribution to variance that could be attributed to all factors, including host genetics was then examined (Fig 3A). For individuals with available data for genetic-based kinship, microbiome data, and the phenotypes in the preceeding paragraph (n=545), unique contribution was obtained from R^2 decomposition on microbiome Bray-Curtis beta diversity estimates, in permanova models (1000 permutations) controlling for other factors. The average for each factor was used after randomly rearranging all factors 20 times. Age was the top contributor, followed by menopause status, history of prior UTI and host genetics (Fig3A, Data S3).

Metagenomes confirm overall 16S microbiome data variation

In microbiome studies, metagenomes not only provide taxonomic information comparable to 16S analyses, they also offer deep insights into metabolic pathways and better species resolution. Using shotgun metagenome data for a subset of 178 individuals, we also examined how closely the overall patterns of the 16S data are replicated in the metagenome data. The classified metagenome reads were 99.64% Bacteria (Data S4) and a greater number of urine metagenomes (total and per individual) were obtained than earlier reported in literature. Sample-sample variation or inter-sample distances in the microbiome data were highly correlated from metagenome and 16S data (Bray dissimilarities, Mantel's r=0.799,p=0.001). Sixteen of the top 20 abundant taxa using 16S are also within the top 20 of the metagenome data. The core microbiome found in 16S data was largely recapitulated in the metagenomics analysis; 27 of the 31 genera (87%) forming the core taxa using 16S data were also replicated in the metagenome data. From this core, the total number of species identifiable approximately doubled (125 vs 61 in total, 94 vs 53 in the replicated genera) most likely to due to better species assignment. Given the choice of the subset for shotgun sequencing (Methods), heritability values were inflated (STable 2). Considered together, the metagenomes largely mirror 16S data and consolidate results on heritability.

Discussion

In this study, we used a relatively larger, unselected community-based study population of women and sensitive approaches (amplicon sequence variants,

compositional clusters and environmental effect control in twin-pairs) to explore the influence of host factors on the urinary tract microbiome. These approaches strengthen deductions made here, for instance that age is associated with increasing urinary tract alpha diversity, contrary to previous studies (e.g. Curtiss et al., 2018; Kramer et al., 2018; Liu et al., 2017;Wang et al., 2017).

Urine and other body sites

The ordination patterns of the urine microbiomes support current thinking that the urobiome is a distinct site, similar to the observations that bladder microbiome (urine obtained directly by catheter) differ from vaginal or stool microbiome (Thomas-White et al., 2018; Wolfe & Brubaker, 2019). Here, the more divergent of the urine studies (Urine1 cohort) shared more vaginal taxa, involved patients with incontinence and collection was wholly catheterized, though had smaller sample size. In a very small minority of individuals where urine microbiome taxa appear closer to stool, this is most likely due to phylogenetic or genome similarity in species (as no such closeness occur with non-phylogenetic measures), rather than common demographics (Data S1). In all, the current study show clear dissimilarities in stool and midstream urine for the average unselected population.

Host-related factors and host genetics' contribution in urinary tract microbiome

Parity (childbirth episodes), previous UTI occurrence, recent antibiotics usage and diet showed changes with urine microbiome diversity. Using heritability analysis, the current study showed a considerable genetic influence in the microbiome of ageing women, reaching almost a third of the variation, with the remainder of contribution largely due to variance unique to individuals. Some clinically important genera such as *Escherichia* (A=0.165) had variants with high heritability estimates, In addition, *Lactobacillus iners* (A=0.177), a commonly found vaginal and bladder microbe, was phylogenetically close to the heritable gut microbe Christenellaceae and heritable in urine.

Previously, Rothschild and colleagues (2018) reported that environmental factors such as sharing household eclipse genetic influence in gut microbiome composition, while Goodrich and colleagues (2014) showed host genetics played roles in gut microbiome patterns of twin-pairs. The current study, indicates significant contributions of genetics to the pattern of urine microbial composition; and

controlling for cohabitation (participants asked if they live together or close with their sibling) and other known factors in urine microbial variation, did not alter the estimated the significant contributions to the pattern. Other parameters from this study bolster the observation of genetic influence: (1) samples of a member in a twin-pair were not extracted or sequenced in the same batch as the other member, (2) there was lower intra-twin difference distance among monozygotic pairs, and (3) the second PCo which had higher heritability than the first was the same PCo which differed along the lines of ethnic ancestry (though the proportion of white British was dominant). Thus we conclude that host genetics influenced variation in urinary microbiome composition in this population of women.

Relative to other factors, only age, menopause status and prior history of current UTI were greater than the influence of genetics. Incidental to our main purpose, we also report here that history of urinary tract infections itself has a significant heritability in humans as suggested in Scholes et al. (2000) using family records and in Norris et al. (2000) using dogs. The results here also show a projected shift in microbiome structure after five UTI episodes (Fig3E).

Heritable urinary tract microbes

While *Corynebacterium* variants were frequent among top core taxa and clusters with high heritability, the patterns detected for *Lactobacillus and Escherichia* variants deserve mention. Our study showed the Escherichia-Shigella taxon, renamed as such to reflect the extreme sequence similarity of Escherichia coli and Shigella, was part of the urinary tract microbiota in older women. In absence of diagnosed infections, the current study shows that presence of this taxon is influenced by (1) host genetic make up (its proportions had one of the top heritability estimates (A=0.17,CI=0.11-0.29) of all frequent urine microbial variants); and (2) age. Price et al. (2019) also recently reported that *Escherichia* urotype were more likely in older asymptomatic patients. These findings may have implications in the mixed success of *E. coli* vaccine trials (Huttner et al., 2017) and in diagnostics.

The current study has limitations. Questionnaire data, which is subject to accurate recall and self-report by participants, was part of measures used in deriving variables such as UTI, diet and frailty. Another limitation may be the use of a single midstream urine sample set from an individual, and as such, prior microbiome stability

information is unknown. Clearly, further research is needed to confirm if the findings also relate to the male urinary microbiome.

To conclude, we report on the factors influencing composition of the urinary tract microbiome in unselected community-dwelling adult women. The urinary microbiome was distinct and apparently unrelated to stool microbiome. It shows a significant contribution of host genetics. Key species known to be clinically relevant were among the most heritable microbes. Age and menopausal status were the factors with greatest influence on the urinary microbiome in women.

Author Contributions

Conceptualization: C.J.S, T.S. and A.S.A; Investigation: C.J.S., G.H., G.A., R.B., P.W. and R.K.; Methodology: C.J.S. and A.S.A.; Formal Analysis: C.J.S. and A.S.A; Writing: A.S.A, C.J.S, G.H., T.S. and R.K.; Funding Acquisition: C.J.S. and T.S; Supervision: C.J.S., T.S. and R.K.

Acknowledgement

We thank Prof Alan Wolfe and Roberto Limeira of Health Sciences Division, Loyola University Chicago, United States for providing access to raw sequence data from two urine studies; the phenotype data team at TwinsUK; laboratory team at TwinsUK for sample handling; and Rachel Horsfall, Marina Mora Ortiz, Mary NiLochlainn and Max Freydin for discussions and comments on the manuscript. CS received research funding through the Chronic Disease Research Foundation which receives funds from the Denise Coates Foundation. We also thank all participants in TwinsUK who altruistically donated their time and samples for this research (www.twinsuk.ac.uk).

Declaration of Interests

The authors declare no competing interests

Figure Legends

Fig 1. Urinary tract microbiome in older women is mostly distinct from proximal body sites and unrelated to stool microbiome. Alpha diversity plots were based on Shannon index and beta diversity based on unweighted unifrac distances. **(A) Alpha**

diversity of urine microbiomes and other body sites. star symbol indicates significance compared to TwinsUK midstream urine. **(B) Dissimilarities in urine microbiomes and other body sites. (C) Paired alpha diversity analysis of stool and urine collected at same time point (D) Differences in paired stool and urine microbiome from the same time point.**

Fig 2. Host genetics considerably influences variation of urine microbiome. (A) Heritability and interaction of core urinary tract microbes. Size of circles at each subcluster and intensity of rectangular bars at the tips represent increasing heritability of taxa. Neighbouring variants in a clade show co-abundance and clustering is not phylogenetic. Taxa are annotated to indicate different variants. **(B) Microbiome dissimilarities within family of twin pairs.** MZ-monozygotic; DZ-Dizygotic **(C) Microbiome principal coordinates with ancestral origin.** Ellipses represent 95% confidence interval. White British constitute>90% of individuals, and bootstrap/permanova testing were used due to imbalanced sizes.

Fig 3. Top contributors to urinary microbiome variation. (A) Relative contributions to urinary microbiome. Bars represents average R^2 for each variable, controlled for the presence of other factors. Microbial variation was measured using Bray-Curtis dissimilarities. Genetic PC was derived from principal components of SNP-based genetic kinship. **(B) Trends in individual Shannon diversity with age and prior number of UTI. (C-E) Microbiome dissimilarities with top factors: (C) age (D) menopause (E) prior number of UTI**.

Tables

Table 1. Summary of participants in TwinsUK urinary microbiome study

Phenotype category	Subcategory	α-D index (mean±SD)	Ave. no of unique taxa(mean±SD)	No. of samples	Age (mean±SD)
Participants		2.01±1.05	65.7±48.8	1600	66.7±8.3
Previous UTI occurrence[s]	0 times	2.14±1.0	66.1±43.1	393	67.6±8.2[s]
	1-4 times	2.02±1.04	67.5±51.0	719	65.9±7.8
	5-9 times	1.98±1.03	65.4±45.2	208	66.3±8.3
	10times >	1.79±1.17	60.0±53.9	201	65.7±8.3

Age[s]	<50-54	1.56±0.76	45.9±32.2	117	-
	55-59	1.86±1.13	61.7±49.7	210	-
	60-64	2.00±0.98	63.5±44.8	327	-
	65-69	2.04±1.03	66.0±49.8	409	-
	70-74	2.16±0.97	71.5±50.6	276	-
	75-79	2.26±1.12	74.5±50.1	170	-
	80-84	2.02±1.12	63.7±41.9	68	-
	85-	1.73±1.42	71.7±62.3	23	-
RecentAntibiotic usage:3mths	Yes	1.97±1.20[ns]	70.0±53.0[ns]	47	68.3±8.0[ns]
	No	2.03±1.06	66.0±49.0	945	66.6±8.3
Frailty	<0.15	2.05±1.01[ns]	67.0+49.0[ns]	511	65.9±7.5[s]
	0.15-0.29	1.99±1.05	64.8±49.0	834	66.1±8.0
	0.3-0.44	2.04±1.15	67.5±48.0	227	68.4±8.9
	>0.45	1.75±1.17	62.0±47.0	28	68.5±8.2

Legend. α-D: Shannon H index of alpha diversity; No. of taxa refers to number of unique sequence variant per sample i.e. no of potential species. Diversity measures were calculated after subsampling to 2000. S/NS indicates statistical significance or not for tests of a phenotype as a continuous variable. Post-hoc pairwise comparisons showed no difference in alpha diversity for individuals aged 75years and older.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, Claire Steves (claire.j.steves@kcl.ac.uk).

Materials Availability

This study did not generate new unique reagents

Data and Code Availability

Raw sequence data is available from EBI's European Nucleotide Archives with accession number ERP119822. Phenotype data is available on request from TwinsUK data access committee at http://twinsuk.ac.uk/resources-for-researchers/access-our-data.html. Scripts and codes used are available at github.com/waleadebayo/urobiome-host-genetics

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cohort and Phenotypes

The TwinsUK cohort has been described elsewhere (Verdi et al. 2019). It comprises over 14,000 volunteers in total over more than two decades, predominantly female (>80%) and middle-aged (mean age 59). Data were collected with visits to the Department of Twin Research and Genetic Epidemiology, King's College London, resulting in biochemical, behavioral, dietary and socioeconomic cohort characterization. Participants in the cohort are community dwelling twin pairs,

recruited without any specific clinical phenotype. The current study included 1600 individuals, and various demographics were examined. Medical history questionnaires were used to define age (from birth date), history of urinary tract infections (UTIs), cohabitation (live together or close neighbourhood), antibiotic usage, previous hysterectomy, previous oophorectomy, caesarian section and menopause status. The frailty index, calculated from clinical, physiological and mental domains (Livshits et al., 2017) was used as a measure of health deficit, and the Healthy Eating Index (Bowyer et al. 2018) based on food frequency questionnaires used to assess diet.

METHOD DETAILS

16S Microbiome Sequencing and Analysis

Twin-pair samples were separated for processing. Extraction and Sequencing of samples along with 128 negative controls was performed at the Knight Lab, University of California San Diego using Earth Microbiome Project (EMP) protocols (http://www.earthmicrobiome.org/protocols-and-standards/16s) with the Qiagen MagAttract PowerSoil DNA kit. Amplicon PCR was performed on the V4 region of the 16S rRNA gene using the primer pair 515f to 806r with Golay error-correcting barcodes on the reverse primer. The barcoded amplicon pool was purified with the MO BIO UltraClean PCR cleanup kit and sequenced on the Illumina MiSeq platform. Sequence data were demultiplexed using the QIIME2 (Bolyen et al. 2019). Multilevel quality filtering procedures and data analysis were applied to remove potential contaminants (see below). Amplicon sequence variants (ASVs), were generated with DADA2(Callahan et al, 2016), filtered and analysed as individual taxa. They were also analysed as balances (Morton et al.,2017), essentially by forming clusters from highly frequent variants (presence in >20% of individuals) which were transformed compositionally, correlated and linked in an hierarchical fashion. ASVs are error-corrected sequences and offer better sequence resolution in taxonomy assignment, which was done with Silva (v132) (Yilmaz et al., 2014). The current data was also compared to those of previous microbiome studies with similar age-range of participants after re-analysis of such data to produce ASVs (see below). Diversity analysis was carried out with Shannon index, unifrac and Bray-Curtis metrics, and permutational multivariate analysis of variance was used to test inter-sample

differences. Taxa counts were centred-log-ratio transformed after adding a pseudocount (van de Boogart et al. 2019).

Filtering and removal of possible contaminants

Initially, common QC processes were followed such as artifact removal, chimera checking, read-length trimming, and short-read discard. As low biomass microbiome tend to be influenced more by contaminants and cross-talk than high biomass sites such as gut, more steps were utilised.

(a)Blank controls (n=105) were sequenced along with normal samples.

(b) Sequence variants (or potential taxa) were removed if the counts attributed to it in the blanks was more than 5% of the total counts for that taxa variant; OR if the number of blanks in which a sequence variant occur is more than 10% of the total number (blanks + actual samples).

(c) Sequence variants were removed if they significantly exhibit a pattern such that its abundance was prevalent in blank controls that were sequenced along with normal samples (e.g. the variant occurs in 50% of blanks but only in 10% of normal samples) or a strong negative correlation ($p<0.1$) exist between the amplicon library concentration and the number of reads generated for a sequence variant, as implemented in Davis et al. (2018). Step (b) above was used to complement step c which could not deal with this.

Subsequently, a sample with reads higher than 2000 was deemed to be reliably detected. Setting cut-off at 2000 reads is based on the fact that

(1) It covers about 99.6% of diversity (Shannon) in rarefaction plots and 99.4% coverage (Good's statistic)

(2) it was much higher than any number reads still present in any blanks after QC and further filtering. i.e. after all QC steps, 30 of 105 blanks sequenced still contained some reads, 90% of these 30 blanks had less than 335 reads, the mean was 152. Because the QC was rather rigorous, these reads are probably due to cross talk in sequencer rather than contaminants.

We also briefly examined potential biological explanations for the occurrence of extremely-low DNA urine sample, apart from efficacy of technical protocols.

Comparison of microbiome studies of similar age

Raw sequence data used in Pearce et al. 2014 (Urine 1), Thomas-White et al. (2017) (Urine 2), Puerto Rico and Plantanal study described as part of Yatsunenko et al. 2012 (Vaginal), Goodrich et al. 2014 (Gut), were obtained on request from authors or from

the EBI's ENA database. These studies, also using 16S V4 region, generally sampled by requesting participants from the general population but some participants in the urine studies were recruited based on specific phenotype of interest. Each sequence set was analysed using the same bioinformatics pipeline described for the current study, and each dataset was subsetted to include only women aged 45 years and above to match current study. Re- analysis of these published datasets helped to avoid some data-induced differences in alpha diversity and create a uniform platform for comparison. Also, to minimise multi-study protocol variations and include as many sample as possible, ASV counts were subsampled to 1000 reads in all studies and also subsampled randomly to 100 subjects in each of two replicate sets (except Urine1 and vaginal with smaller participants, n=57 and 11,respectively).

Metagenome Analysis

Shotgun metagenomic sequencing was carried out for 178 of the participants with additional 14 blanks for quality control. The protocol involved 5ng DNA per reaction quantified using a PicoGreen fluorescence assay. After fragmentation, end repair and A-tailing, sequencing adapters and barcode indices are added following the iTru adapter protocol. Unique error-correcting i7 and i5 indices were used after purification, and indexed libraries were then purified again, quantified and normalized, prior to sequencing on the Illumina HiSeq4000 platform. The approach involved shallow shotgun methods (SHOGUN) (Hillman et al., 2018). This subset of participants included equal numbers of dizygotic pairs and monozygotic twin pairs, as well as equal numbers of twin pairs showing discordance and concordance in 16S microbial diversity (pair-to-pair difference in diversity (Shannon index) greater than 3SD or lower than 1SD). This was expected to inflate heritability estimates. After quality control filtering (with average q ≥30), and mapped human reads' removal (based on hg19) one sample was excluded, and the final data included 177 samples. Potential contaminants, eleven species, were removed for presence in blanks and constituting >2% (between 6% and 100%) of the abundance of that species. These 'contaminant' species included *Mycobacterium_iranicum, Gordonia_paraffinivorans, Staphylococcus_saprophyticus, Delftia_acidovorans,Corynebacterium_matruchotii,Staphylococcus_capitis, Acinetobacter_harbinensis, Corynebacterium_singulare, Cutibacterium_granulosum, Acinetobacter_towneri, and Cutibacterium_acnes.*

Host genetics analyses

Heritability was calculated using an ACE model in which the component of phenotypes explained by genetics in twin pairs was estimated. Samples from co-twin were separated into different batches for sample preparation and sequencing to remove the shared technical environment related to batching. This further solidified the deductions made from the analysis of the genetic effects. Discordance analysis, quantitative differences in microbiome for pairs of monozygotic and dizygotic twins, was approached using constrained principal coordinates as well as dispersions in microbiome variance. Where constrained principal coordinates analysis was used, microbiome data was ordinated with the family ID tested as a predictor, then the dissimilarity within a family was then extracted to compare twin types. Analysis on ethnic origin of participants was based on information obtained from questionnaires, and to reduce the impact of the large difference in the group sizes, partitions were created in which non-British groups were repeatedly sampled, before bootstrapped Kruskal-Wallis statistic were estimated. Also, as a confirmation, permutation-based testing were used for the original undivided data. To represent host genetic variation, first principal component from genome-based kinship matrix data were obtained. These analyses were carried out with plink1.9b, R base and R packages: vegan, mets, car, phyloseq(see Resources Table).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details and tests can be found in Results and Method Details sections following the contexts in which they were used. n represents number of individuals, SD represents one standard deviation, confidence intervals were set at 95%, and significance threshold was set at alpha less than 0.05. Throughout analysis, technical covariates, including extraction kit lots, mastermix kit lot, batch, extraction and sequencing processors, and depth/library sizes (sequence reads post-QC filtering) were controlled for. Visualizations were generated using ade4, ggplot2, graphlan2, decontam, FastTree and ggtree (see Resources Table).

STable1. Heritability of midstream urine microbiome abundance in paired twins. Related to Figure 2.

STable2. Heritability of urine metagenomes from diversity-discordant and diversity-concordant pairs of twins. Related to Figure 2.

References

Adebayo, A.S., Survayanshi, M., Bhute, S., Agunloye, A.M., Isokpehi, R.D., Anumudu, C.I., and Shouche, Y.S. (2017). The microbiome in urogenital schistosomiasis and induced bladder pathologies. PLoS Negl. Trop. Dis. *11*.

Alanee, S., El-Zawahry, A., Dynda, D., McVary, K., Karr, M., and Braundmeier-Fleming, A. (2019). Prospective examination of the changes in the urinary microbiome induced by transrectal biopsy of the prostate using 16S rRNA gene analysis. Prostate Cancer Prostatic Dis.

Aragón, I.M., Herrera-Imbroda, B., Queipo-Ortuño, M.I., Castillo, E., Del Moral, J.S.G., Gómez-Millán, J., Yucel, G., and Lara, M.F. (2018). The Urinary Tract Microbiome in Health and Disease. Eur. Urol. Focus *4*, 128–138.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. *3*:e1029.

Bajic P, Van Kuiken ME, Burge BK, Kirshenbaum EJ, Joyce CJ, Wolfe AJ, Branch JD, Bresler L, Farooq AV. (2008). Eur Urol Focus. 2018 Aug 21. pii: S2405-4569(18)30220-7

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A. et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology *37*, 852–857

Bowyer, R.C.E., Jackson, M.A., Pallister, T., Skinner, J., Spector, T.D., Welch, A.A., and Steves, C.J. (2018). Use of dietary indices to control for diet in human gut microbiota studies. Microbiome *6*, 77

Brown, K., Church, D., Lynch, T., and Gregson, D. (2014). Bloodstream infections due to Peptoniphilus spp.: Report of 15 cases. Clin. Microbiol. Infect. *20*, O857–O860.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. and Holmes, S.P. (2016). Dada2: high-resolution sample inference from illumina amplicon data. Nature methods, *13*(7):581.

Curtiss, N., Balachandran, A., Krska, L., Peppiatt-Wildman, C., Wildman, S., and Duckett, J. (2018). Age, menopausal status and the bladder microbiome. Eur. J. Obstet. Gynecol. Reprod. Biol. *228*, 126–129.

Davis, N.M., Proctor, D., Holmes, S.P., Relman, D.A., Callahan, B.J. (2017). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. bioRxiv. *221499*.

Dray, S., Dufour, A. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists." Journal of Statistical Software *22*(4), 1–20.

Eline Slagboom, P., van den Berg, N., and Deelen, J. (2018). Phenome and genome based studies into human ageing and longevity: An overview. Biochim. Biophys. Acta - Mol. Basis Dis. *1864*, 2742–2751.

El-Zawahry, A., Dynda, D., McVary, K., Karr, M., and Braundmeier-Fleming, A. (2019). Prospective examination of the changes in the urinary microbiome induced by transrectal biopsy of the prostate using 16S rRNA gene analysis. Prostate Cancer Prostatic Dis.

Felice, V.D., and O'Mahony, S.M. (2017). The microbiome and disorders of the central nervous system. Pharmacol. Biochem. Behav. *160*, 1–13.

Fok, C.S., Gao, X., Lin, H., Thomas-White, K.J., Mueller, E.R., Wolfe, A.J., Dong, Q., and Brubaker, L. (2018). Urinary symptoms are associated with certain urinary microbes in urogynecologic surgical patients. Int. Urogynecol. J. *29*, 1765–1771.

Fouts, D.E., Pieper, R., Szpakowski, S., Pohl, H., Knoblach, S., Suh, M.J., Huang,S.T., Ljungberg, I., Sprague, B.M., Lucas, S.K., Torralba, M., Nelson, K.E., Groah, S.L. (2012). 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 *10*:174.

Franco, A.V. (2005). Recurrent urinary tract infections. Best Practice & Research Clinical Obstetrics and Gynaecology. *19*, 861–873.

Goodman, B., and Gardner, H. (2018). The microbiome and cancer. J. Pathol. *244*, 667–676.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Hillmann, B., Al-ghalith, G.A., Shields-cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., and Knights, D. (2018). crossm Metagenomics. MSystems *3*, 1–12.

Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. (2014). J Clin Microbiol. *52* (3), 871-6.

Huttner, A. et al. (2017). Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. Lancet Infect. Dis. *17*, 528–537.

Karstens, L., Asquith, M., Caruso, V., Rosenbaum, J.T., Fair, D.A., Braun, J., Gregory, W.T., Nardos, R., and McWeeney, S.K. (2018). Community profiling of the urinary microbiota: considerations for low-biomass samples. Nat. Rev. Urol. *15*, 735–749.

Kline, K.A., and Lewis, A.L. (2016). Gram-Positive Uropathogens, Polymicrobial Urinary Tract Infection, and the Emerging Microbiota of the Urinary Tract. Microbiol. Spectr. *4*.

Könönen, E., and Wade, W.G. (2015). Actinomyces and related organisms in human infections. Clin. Microbiol. Rev. *28*, 419–442.

Khasriya R, Sathiananthamoorthy S, Ismail S, Kelsey M, Wilson M, Rohn JL, Malone-Lee J. (2013). J Clin Microbiol. *51* (7), 2054-62.

Kramer, H., Kuffel, G., Thomas-White, K., Wolfe, A.J., Vellanki, K., Leehey, D.J., Bansal, V.K., Brubaker, L., Flanigan, R., Koval, J., et al. (2018). Diversity of the midstream urine microbiome in adults with chronic kidney disease. Int. Urol. Nephrol. *50*, 1123–1130.

Lee, J.B.L., Neild, G,H. (2007). Urinary tract infection. Medicine *35*, 423–8

Liu, F., Ling, Z., Xiao, Y., Yang, Q., Zheng, L., Jiang, P., Li, L., and Wang, W. (2017). Characterization of the urinary microbiota of elderly women and the effects of type 2 diabetes and urinary tract infections on the microbiota. Oncotarget *8*, 100678–100690.

Livshits, G., Lochlainn, M.N., Malkin, I., Bowyer, R., Verdi, S., Steves, C.J., and Williams, F.M.K. (2018). Shared genetic influence on frailty and chronic widespread pain: A study from TwinsUK. Age Ageing *47*, 119–125.

Long, T., Hicks, M., Yu, H. et al. (2017). Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet *49*, 568–578

Luca F., Kupfer S.S., Knights D, Khoruts A and Blekhman R. (2018). Functional Genomics of Host–Microbiome Interactions in Humans. Trends in Genetics *34*,30-40.

Ma, Z. (Sam), Li, L., and Gotelli, N.J. (2019). Diversity-disease relationships and shared species analyses for human microbiome-associated diseases. ISME J. 1.

McMurdie, P.J., Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE, *8*(4), e61217

Moayyeri, A., Hammond, C.J., Hart, D.J., and Spector, T.D. (2013). The UK adult twin registry (twinsUK resource). Twin Res. Hum. Genet. *16*, 144–149.

Morton, J.T., Sanders, J., Quinn, R.A., McDonald, D., Gonzalez, A., Vázquez-Baeza, Y., Navas-Molina, J.A., Song, S.J., Metcalf, J.L., Hyde, E.R., et al. (2017). Balance Trees Reveal Microbial Niche Differentiation. MSystems *2*, 1–11.

Moustafa, A., Li, W., Singh, H., Moncera, K.J., Torralba, M.G., Yu, Y., Manuel, O., Biggs, W., Venter, J.C., Nelson, K.E., et al. (2018). Microbial metagenome of urinary tract infection. Sci. Rep. *8*, 1–12.

Oksanen, J., Guillaume Blanchet, J., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. et al (2019). vegan: Community Ecology Package. R package version2.5-3.

Pearce, M.M., Hilt, E.E., Rosenfeld, A.B., Zilliox, M.J., Thomas-White, K., Fok, C., Kliethermes, S., Schreckenberger, P.C., Brubaker, L., Gai, X., et al. (2014). The female urinary microbiome: A comparison of women with and without urgency urinary incontinence (2014). MBio *5*. 5 (e01283-14)

Pearce, M.M., Zilliox, M.J., Rosenfeld, A.B.,Thomas-White, K.J. Richter,H.E., .Nager,C.E. Visco,A.G.Nygaard,I.E.Barber,M.D. et al. (2015). The female urinary microbiome in urgency urinary incontinence. Am J Obstet Gynecol *213*, 347.e1-11

Price, M.N., Dehal, P.S., Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE *5*(3): e9490.

Rani, A., Ranjan, R., McGee, H.S., Andropolis, K.E., Panchal, D. V., Hajjiri, Z., Brennan, D.C., Finn, P.W., and Perkins, D.L. (2017). Urinary microbiome of kidney transplant patients reveals dysbiosis with potential for antibiotic resistance. Transl. Res. *181*, 59–70.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Scheike, T.H., Holst, K.K., Hjelmborg, J.B. (2014). Estimating heritability for cause specific mortality based on twin studies. Lifetime Data Analysis, *20*(2), 210-233.

Shrestha, E., White, J.R., Yu, S.H., Kulac, I., Ertunc, O., De Marzo, A.M., Yegnasubramanian, S., Mangold, L.A., Partin, A.W., and Sfanos, K.S. (2018). Profiling the Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. J. Urol. *199*, 161–171.

Sihra, N., Goodman, A., Zakri, R., Sahai, A., and Malde, S. (2018). Nonantibiotic prevention and management of recurrent urinary tract infection. Nat. Rev. Urol. *15*, 750–776.

Thomas-White, K.J., Hilt, E.E., Fok, C., Pearce, M.M., Mueller, E.R., Kliethermes, S., Jacobs, K., Zilliox, M.J., Brincat, C., Price, T.K., et al. (2016). Incontinence medication response relates to the female urinary microbiota. Int. Urogynecol. J. *27*, 723–733.

Thomas-White, K.J., Kliethermes, S., Rickey, L., Lukacz, E.S., Richter, H.E., Moalli, P., Zimmern, P., Norton, P., Kusek, J.W., Wolfe, A.J., et al. (2017). Evaluation of the urinary microbiota of women with uncomplicated stress urinary incontinence. Am. J. Obstet. Gynecol. *216*, 55.e1-55.e16.

Thomas-White K.J., Gao X, Lin H, Fok CS, Ghanayem K, Mueller ER, Dong Q, Brubaker L, Wolfe AJ. (2018). Int Urogynecol J. *29* (12),1797-1805.

van den Boogaart, K.G., Tolosana-Delgado, R., Bren, M. (2019). compositions:Compositional Data Analysis. R package version 1.40-3.

Verdi, S., Abbasian, G., Bowyer, R., Lachance, G., Yarand, D., Christofidou, P., et al. (2019). TwinsUK: The UK Adult Twin Registry Update. Twin Research and Human Genetics, *22*(6), 523-529.

Verma, R., Morrad, S., and Wirtz, J.J. (2017). Peptoniphilus asaccharolyticus -associated septic arthritis and osteomyelitis in a woman with osteoarthritis and diabetes mellitus. BMJ Case Rep. *2017*.

Wang, H., Altemus, J., Niazi, F., Green, H., Calhoun, B.C., Sturgis, C., Grobmyer, S.R., and Eng, C. (2017). Breast tissue, oral and urinary microbiomes in breast cancer. Oncotarget *8*, 88122–88138.

Whiteside, S.A., Razvi, H., Dave, S., Reid, G., and Burton, J.P. (2015). The microbiome of the urinary tract - A role beyond infection. Nat. Rev. Urol. *12*, 81–90.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4

Wolfe, A.J., and Brubaker, L. (2019). Urobiome updates: advances in urinary microbiome research. Nat. Rev. Urol. *16*, 73–74.

Wolfe, A.J., Toh, E., Shibata, N., Rong, R., Kenton, K., FitzGerald, M.P., Mueller, E.R., Schreckenberger, P., Dong, Q., Nelson, D.E., et al. (2012). Evidence of uncultivated bacteria in the adult female bladder. J. Clin. Microbiol. *50*, 1376–1383.

Wu, P., Chen, Y., Zhao, J., Zhang, G., Chen, J., Wang, J., and Zhang, H. (2017). Urinary microbiome and psychological factors in women with Overactive bladder. Front. Cell. Infect. Microbiol. *7*.

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O. (2014). The SILVA and All-species Living Tree Project (LTP) taxonomic frameworks. Nucl. Acids Res. *42*:D643-D648

673 Yu, G., Lam, T.T., Zhu, H., Guan, Y. (2018). Two methods for mapping and visualizing associated
674 data on phylogeny using ggtree. Molecular Biology and Evolution. *35*, 3041-3043.
675
676
677
678
679
680
681
682
683
684