Urobiome: State of Knowledge, and Evaluation of Methods for Enabling Genome-Resolved Metagenomics of Urinary Tract Microbiota

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of the Ohio State University

Zachary Jordan Lewis, BA

Comparative Biomedical Sciences Graduate Program

The Ohio State University

2024

Thesis Committee

Vanessa L. Hale, MAT, DVM, PhD

Virginia I. Rich, PhD

Brian D. Husbands, DVM, DACVIM

Abstract

The urinary tract contains a distinct and diverse set of microbiota (urobiome). Important work has established links between the composition of the urobiome and various diseases of the urinary tract, including bladder cancer, urolithiasis, incontinence, and recurrent urinary tract infection. However, efforts to characterize the urobiome and assess its functional potential have been limited due to technical challenges including low microbial biomass and high host cell shedding in urine. To begin addressing these challenges, we evaluate urine sample volume (100 µl – 5 mL), and host DNA depletion methods and their effects on urobiome profiles in healthy dogs, which are a robust large animal model for the human urobiome. We collected urine from seven dogs and fractionated samples into multiple aliquots. One set of samples was additionally spiked with host (canine) cells to model a biologically relevant host cell burden in urine. Samples then underwent DNA extraction followed by 16S and shotgun metagenomic sequencing. We tested six methods of DNA extraction: QIAamp BiOstic **Bacteremia** (no host depletion), QIAamp **DNA Microbiome**, **Molzym** MolYsis, **NEBNext** Microbiome DNA Enrichment, **Zymo HostZERO**, and **Propidium Monoazide**. Sequences were processed and analyzed using QIIME2 and MetaPhlAn4. Metagenome assembled genomes (MAGs) were generated using MEGAHIT and MetaWRAP pipelines. Statistical analysis were performed in R. In relation to urine sample volume, ≥ 3.0 mL resulted in the most consistent urobiome profiling. In relation to host depletion, individual (dog) but not extraction method drove overall differences in microbial composition. DNA Microbiome yielded the greatest microbial diversity in 16S ($p=0.0025$) and shotgun metagenomic data ($p=0.01$), and maximized MAG recovery while effectively depleting host DNA ($p=0.0039$) in host-spiked urine samples. As proof-of-principle, we then mined MAGs for core metabolic functions and environmental chemical metabolism. We identified long chain

alkane utilization in two of the urine MAGs. Long chain alkanes are common pollutants that result from industrial combustion processes and end up in urine. This is the first study, to our knowledge, to demonstrate environmental chemical degradation potential in urine microbes through genome-resolved metagenomics. These findings lay the foundation for future evaluation of urobiome function in relation to health and disease.

Acknowledgements

Many thanks to the Hale Lab: Angela Scott, for her assiduous management of the laboratory, as well as assistance at the benchtop and with data analysis; Dr. Christopher Madden, for benchtop and methodological assistance; Dr. Tessa Cannon, Danielle Curcio, Dr. Andrew McGlynn, Dr. Emily Coffey, Jordan Mason, and Madison Griffin, for feedback and support.

To my committee members, Dr. Virginia Rich and Dr. Brian Husbands, for support and knowledge throughout my degree.

To the Center of Microbiome Science and Infectious Disease Institute Genomics and Microbiology Solutions Laboratory, including Dr. Dean Vik, Shashanka Murthy, and Dr. Michael Sovic for significant metagenomics support.

To my advisor, Dr. Vanessa Hale, for her indispensable guidance, teaching, and support, without whom I could not succeed as I have.

To Dad, Mom, Samuel, Nathaniel, and Abigail, for everything.

Funding for this research was provided through the Ohio State University College of Veterinary Medicine Canine Funds and the Duncan Alexander Advisory Fund.

Vita

2018-2022, BA, Biology (biochemistry minor), Carleton College, Northfield, MN

Magna Cum Laude, Distinction in major and thesis

2022-2024, Graduate Research Associate, The Ohio State University, Columbus, OH

2022-2024, MS, Comparative Biomedical Sciences, The Ohio State University, Columbus, OH

Fields of Study

Major Field: Comparative Biomedical Sciences

Table of Contents

List of Figures

List of Tables

Chapter 1 – Literature Review

Section 1 – Urobiome in Health & Disease

Urobiome – Introduction, Misunderstandings, and Potential

Over the last two decades, scientists have made extensive effort to study, understand, and engineer host-associated microbiota, beginning with the Human Microbiome Project (HMP) and spurred by advances in molecular sequencing and bioinformatic technology.[1] Causal relationships between the host-associated microbiota and various health outcomes and diseases have been established, including obesity, inflammatory bowel disease, and cancer.[2] The importance of the microbiota across body sites, including the urinary tract, is clear; however, clinical dogma that the urine of healthy individuals is sterile[4] led to the exclusion of the urinary tract from the first phase of the HMP. Indeed, clinical testing of urinary tract microbes is primarily focused on identifying and treating urinary tract infection (UTI), with commensurate methods: simple aerobic culture, microscopy, and cytology.[6,7] This limited understanding of the urobiome has real effects; clinicians may treat asymptomatic bacteriuria (the presence of bacteria $>10^5$ colony-forming-units/mL (CFU/mL) in urine) with antibiotics, which may increase the likelihood of progression to symptomatic or recurrent UTI (rUTI) and promote antimicrobial resistance (AMR).[6,8,9] Similarly, urolithiasis (bladder/kidney stones) was thought to be of spontaneous, dietary, or UTI-related origin, but new work has revealed the potential for contributions of the urobiome to stone formation.[10,11] Researchers' ability to understand possible functions of the urobiome, such as nutrient utilization, metabolite production, xenobiotic degradation, and virulence, is likewise limited.

Though a vaginal microbiome, which may be considered part of a broader "urogenital microbiome," was included in the HMP,[1] the first notion of a urobiome came in 2012 with the molecular characterization of microbes in the urine of adolescent male humans by Nelson et al.[12] Four notable genera of the newly-characterized community, identified via 16s rRNA amplicon sequencing, include *Corynebacteria, Staphylococcus, Lactobacillus,* and *Anaerococcus*, which are frequently detected in urine samples across studies and species.[13,14] Because voided urine may contain bacterial cells from any part of the urinary tract (kidney, ureter, bladder, urethra) as well as from the skin near the urethral opening, Wolfe et al. (2012) used transurethral catheterization and suprapubic aspiration to collect urine directly from the bladders of healthy women; 16S rRNA Sanger and pyrosequencing revealed the presence of bacteria that had not been cultured from urine using conventional methods.[14] Hilt et al. (2014) developed an expanded quantitative culturing method (EQUC) that included longer incubation, aerobic, anaerobic, and facultative conditions, and higher-sensitivity plating, to show that many of the bacteria whose genomes could be detected from bladder samples were indeed viable.[4]

Following these landmark results, it is now well established that the urinary tract microbiota (kidney, ureter, bladder, urethra, "urobiome") are viable and diverse,[4,11,13,15–24] but the urobiome remains understudied compared to other host-associated ecosystems, such as the gut microbiome. Experimental methods and models allowing the manipulation of the gut microbiota, such as fecal microbiota transplantation and germ-free mice, have allowed researchers to perform controlled, hypothesis-driven experiments to establish mechanistic links between the gut microbiome, health, and disease. Efforts are underway to reach this level of rigor in urobiome study.

What do we know about the healthy urobiome in humans and dogs?

The normal microbiota of the gut are essential for many basic physiological processes in the host, including nutrient extraction and biosynthesis, innate and adaptive immunity, mucosal-barrier homeostasis, and colonization resistance to pathogens.[2] Undesirable changes to the gut microbial community (dysbiosis) can cause disruptions to these critical functions that lead to disease in some settings.

The urinary tract lumen, like other mucosal body sites, comprises specialized epithelial cells (urothelium) beneath a mucosal layer.[25] Though urinary lumen is biochemically and structurally different from other mucosal sites, there is an intense interest in understanding how the urinary tract microbiota may interact with the mucosa and underlying cells in the setting of various urinary tract diseases, including urinary incontinence, UTI, urolithiasis, and bladder cancer, similar to how such functions have been discovered at mucosal sites like the gut.[11,20,26–28]

However, urobiome study has largely been limited to case-control, hypothesis-generating work comparing the microbiota of healthy individuals to those in cases of various diseases. Meta-analyses of such studies are difficult because research groups vary in their approaches to urine collection, DNA extraction, primer choices, sequencing platform, and bioinformatic pipelines, all of which can influence urobiome results.[23,29–31]

Still, limited analyses of the healthy control groups across studies, as well as a few health-focused urobiome studies hint at a "core" healthy urobiome. In a study of 1600 postmenopausal women, Adebayo et al. (2020) identified 61 taxa that were present in >20% of samples, with notable genera including *Corynebacterium, Arcanobacterium, Anaerococcus, Actinomyces, Escherichia*, *Prevotella, Fusobacterium, Bacteroides, Enterococcus, Gardnerella,* and *Lactobacillus*.[32] In the same year, Pohl et al. reported female urobiota dominated by

Lactobacillus and *Prevotella,* while male urobiome samples had higher proportions of *Streptococcus*, *Staphylococcus, Veillonella*, and *Gardnerella.*[31] Other studies have reported high proportions of *Prevotella* and *Lactobacillus* in male urobiota, as well as *Staphylococcus, Streptococcus, and Veillonella*, but do not report *Gardnerella* in males.[33,34]

Sex differences in urobiota may reflect differences in anatomy, different "sources" of urobiota, and differential risks for disease; *Gardnerella* and *Prevotella spp.* are associated with the vaginal microbiome and can be causative agents of bacterial vaginosis, while *Lactobacillus spp.*, which are common urine and vaginal microbiota, appear to be protective against both vaginosis and UTI.[19,35–37] Thomas-White et al. (2018) show overlap between vaginal and bladder isolates in adult human women, suggesting that the vagina may be a source of microbes for the bladder.[19] Komesu et al. (2020) confirm 60% overlap between vaginal and bladder microbiota by comparing 16S profiles between vaginal swabs and catheterized urine samples from the same patient.[38] In the context of rUTI dynamics, where patients undergo periods of infection latency and reemergence, vaginal colonization followed by ascension of the urethra to the bladder of pathogens is a proposed route of infection/colonization.[39] For example, sexual activity and bacterial vaginosis are both associated with UTI.[9] Further, intravaginal administration of *L. crispatus* following antibiotic administration for acute UTI episodes in women with a history of rUTI reduces recurrence.[40] These urobiome, vaginosis, and UTI data collectively implicate the vagina as a source of urinary tract microbes. Once in the bladder, some uropathogens, named uropathogenic *E. coli* strains (UPEC) are able to establish quiescent, intracellular colonies in transitional cells of the bladder epithelium that can be a source of future infection.[41] In studies utilizing a 3D model of human bladder epithelia (7-8 cell layers including all sublayers, as well as the thin glycosaminoglycan layer at the lumen surface), Flores et al. (2023) were able to show

that both UPEC and commensal *E. coli* were able to adhere extracellularly to urothelial cells at multiple layers of epithelium (in addition to canonical, intracellular invasion of urothelial cells), causing variable amounts of cytotoxicity and exfoliation according to pathogenicity.[42] Commensal *E. coli* also demonstrated aggregate formation on the *underside of actively exfoliating urothelial cells*, a novel niche that may protect commensals from shear stress during bladder filling and emptying and allow persistence in the bladder.[42]

In males, the source of urinary tract microbes is less clear, though a study by Dubourg et al. (2020) comparing the gut and urine microbiota via culturomics and Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI) suggested that the gut may be a major source of urine anaerobes in both sexes. In this study, 64% of isolated urobiota species had previously been isolated from the gut, while 32% had been previously isolated from the vagina.[43] Worby et al. (2022), in a longitudinal study of the gut microbiota of women experiencing rUTI, showed that most episodes of infection were preceded by a bloom of the culprit *E. coli* strain in the gut, strongly suggesting the gut as a source of urinary tract microbes, but the actual route of translocation (i.e. ascension, hematogenous spread) was not established.[44] Some researchers have proposed that a more permeable gut epithelium (as results from inflammation in the gut) may allow bacteria to translocate from gut to bladder via the blood or viscera, and indeed, there are associations between gastrointestinal permeability and urinary tract infection, but direct evidence for such a process has not been demonstrated.[45]

Whether or not the urobacteria consistently associated with health are depleted or missing in the setting of specific diseases, except in the case of *Lactobacillus*, remains an open question. Many normal members of the urobiome are also common opportunistic uropathogens in humans and dogs, including *Enterococcus spp.* and *E. coli*,[46] and variability in what is normal

makes it difficult to define an abnormal: several studies of the human female urobiome report "urotypes," urobiome community composition that are dominated by one or two of the most common taxa, generally *Lactobacillus*, *Gardnerella*, or *Enterobacteriaceae*, with urotype sometimes associating with clinical factors like body mass index, previous UTI, and age.[20,47,48] Notably, a similar concept was proposed for gut microbiota variation – "enterotypes" – but has been subject to debate given the vast diversity of gut microbiota and the observation of gradient-like, rather than discrete, variation in composition.[49] Additionally, strain-level variation in urine isolates (and microbiome consituentsconstituents generally) can be ecologically important: some *E. coli* are pathogenic, while others can be pathogen-competitive.[21,50,51] It is likely that, as more urobiomes are sequenced, researchers will appreciate greater diversity in composition, though stratification of urobiomes by urotype may still be clinically or theoretically useful.

The human urobiome literature landscape is generally biased towards two demographics: female sex and older age. Both biases are attributable to the clinical landscape; women are much more likely than men to suffer from infectious urinary tract disease,[52] and declining bladder function and structural changes to the urinary tract with aging are well-documented.[53] However, urinary tract neoplasms are more common in human males than females, and this is reflected in the bladder-cancer-specific urobiome literature.[29] The age bias is particularly important, as some urobiome work suggests that urobiome diversity and composition are related to aging,[32,47] and such changes could be related to the changing immune state in aging patients, i.e., "inflammageing," and affect health outcomes, particularly in the context of age-associated diseases including neoplasms.[28] The relationship between age, inflammation, and increased urobiome diversity is suggestive also of potential hematogenous spread of microbes as a result of increased epithelial permeability.

There is a relative paucity of data on the canine urobiome, though available data suggest some similarities to the patterns seen in human samples. Burton et al. (2017) collected urine via cystocentesis from 24 dogs and submitted the samples for 16s rRNA sequencing.[46] Only five taxa were present in any sample at >1% abundance: *Pseudomonas*, *Sphingobium*, *Acinetobacter*, and unclassified Bradyrhizobiaceae and Xanthomonadaceae. However, other notable members detected with high prevalence include *Staphylococcus* and *Bacteroides,* two common human urobiota.[32] This study had neither negative controls nor methods to bioinformatically decontaminate samples, which may account for the dominance of common reagent contaminants like Bradyrhizobiaceae and Xanthomonadaceae; without such controls, it is impossible to know the true source of such taxa, which are often contested.[54–56] The authors noted no sex differences in canine urobiome.

Melgarejo et al. (2021) performed a similar study with 50 animals, and the six most prevalent genera were *Streptococcus, Sphingomonas, Pseudomonas, Corynebacterium, Staphyloccocus, and Actinomyces*, which represents considerable overlap with human urobiota. This study employed negative controls and adjusted true samples by subtracting absolute read counts of sequence variants in negatives from true samples. This increases the confidence that leftover variants are truly host-associated.[57] A small number of case-control and methodological studies with healthy controls have also been performed in companion dogs.[11,13,22,23] Mrofchak et al. (2023), in a study comparing the urobiota of healthy dogs and dogs with urothelial carcinoma, note high abundances of *Staphylococcus, Streptococcus, Corynebacterium,* and *Sphingomonadaceae.*[13]

Dogs as models for the human urinary tract: microbiome, urinary tract disease

Dogs have been proposed as a suitable model for the human microbiome[58,59] and several urinary tract diseases, including induced stress urinary incontinence[60] and spontaneous urothelial carcinoma.[61] The canine gut microbiota are generally more similar, in both taxonomic and gene composition, to those of humans than are the microbiota of pigs or rodent models.[58] This similarity is thought to arise from the shared environment between companion dogs and their human families. Early urobiome work as well as UTI strain-tracking suggests that this similarity may extend to a shared urobiome between dogs and humans as well. In a three-year study of a six-member family (5 humans and a dog), clones of UTI-causing *E. coli* circulated between the feces and urine of all members, resulting in UTI in two members, including the dog.[62] Mrofchak et al. (2023) compared, via Weighted UniFrac, the overall composition of human and canine urobiome samples using publicly available data; though human and canine urobiome compositions were statistically different, there was considerable, visualizable, phylogenetic overlap between communities.[13]

Dogs are increasingly seen as an important model for muscle invasive urothelial carcinoma.[61] The disease occurs spontaneously in immunocompetent animals and proceeds through similar histopathology and genetic subtypes as human disease[61,63] It is associated with similar exposures in both dogs and humans, including pesticides and cigarette smoke.[64,65] Rodent models of bladder cancer, in contrast, must be immunocompromised, have tumors chemically induced or implanted, and do not share passive exposures or cohabitate with humans.[66] There is a breed bias for bladder cancer in dogs, with Scottish Terriers and Shetland Sheepdogs being "high-risk," and this breed disparity represents an opportunity to parse genomic, microbiome, and environmental contributions.[61,64] Given the similarities between dogs and humans in bladder

cancer, microbiome, and environment, companion dogs are primed to be valuable models for this disease. Additionally, translational research performed using spontaneous animal models benefits both the model animal and humans.

Microbiome and cancer

An important driver for urobiome research is the potential to unearth interactions between bladder tumors and urobiota,[29] because the discovery important relationships between the microbiota of other body sites and various cancers continues.[67] Such relationships can exist on the level of specific microbial strains: Rubinstein et al. (2013) showed that *Fusobacterium nucleatum* adheres directly to colorectal cancer cells via FadA, producing an intratumoral inflammatory state that results in tumorigenesis, driving disease.[68] Whole microbial communities and specific microbes can impact the response to cancer treatment: 30% of non-responders to immune checkpoint blockade therapy for metastatic melanoma became responsive after receiving a transplant of fecal microbiota from responsive donors,[69] while in another study on metastatic renal cell carcinoma, a single-agent *Clostridium* probiotic extended progression free survival by twelve months when combined with immune blockade.[70]

A host of studies demonstrate extensive microbial involvement in therapy response, chemotherapy pharmacokinetics, antitumor immune response, tumor microenvironment, and cancer therapy side effects.[67] Even the skin microbiome has proven important in cancer treatment – presenting tumor antigens in an engineered *S. epidermidis* inoculated onto the skin of tumor-bearing mice induced a strong and enduring T-cell response to the tumors.[71] The gut microbiome may indeed be related to bladder cancer as well: in a multicenter observational study of neoadjuvant chemotherapy for muscle-invasive bladder cancer, distinct clustering of the gut

microbiota was observed between bladder cancer patients and healthy controls, and machine learning models trained on microbiota were better at predicting response outcomes than when trained on clinical variables.[72] Given these strong associations between the gut microbiota and cancer, especially those involving direct tumor-microbe interactions, the importance of understanding the urobiome in the context of bladder tumors is clear.

Urobiome and bladder cancer in dogs and humans – What do we know?

Several studies have investigated the urobiome or bladder tissue microbiome in human bladder cancer patients, and a single study, from our lab, has investigated the urobiome in canine bladder cancer patients.[13,16,34,73–84] These studies vary significantly in methodology, making direct comparisons of results difficult, though all utilized 16S rRNA sequencing with various primers. Indeed, there are few identifiable trends. Chorbińska et al. (2023) compared the urine and gut microbiota of 25 cancer patients (four women) to those of 7 controls (2 women), and found no differences in alpha or beta diversity according cancer status; nor were any individual taxa differentially abundant according to group.[84] The highest-quality evidence for bladder cancer-associated changes to the urobiota is contained in a global meta-analysis by Bukavina et al. (2023),[16] which combines the raw sequencing data from three other studies, in addition to a new cohort of patients and controls from the United States: Bučević Popović et al. (2018, Croatia),[74] Zeng et al. (2020, China),[80] and Mansour et al. (2020, Hungary).[78] All included studies amplified and sequenced the V1-V3 region of the 16s rRNA gene, but utilized a mix of extraction and sequencing methods (Illumina or Ion Torrent sequencing). The combined dataset compares the urobiota of 129 patients with bladder cancer, mostly male, to 60 healthy controls, using a mix of catheter-collected and free catch urine samples. Despite strong effects of

geography and collection method, there were identifiable discrepancies between cancer urobiota and healthy urobiota. Alpha diversity did not differ according to health studies across all cohorts, but US bladder cancer patients had reduced richness and Shannon diversity compared to US healthy controls. This seems to be in line with some studies not included in the meta-analysis,[73,77] but contrasts with others that report increased diversity in cancer-associated urobiota[82,85] and others that report no differences in diversity.[76,79]

However, most compelling from this meta-analysis were the differentially abundant taxa between bladder cancer and healthy urobiota. Bladder cancer urobiota, across cohorts, were enriched for *Enterococcus, Acinetobacter, Pseudomonas, Micrococcus, Ralstonia, Escherichia, Streptococcus,* and *Gardnerella*. These genera tend to agree with studies not including in the meta-analysis, which have also identified *Micrococcus Pseudomonas* as bladder cancer biomarkers in the urobiome.[83,85] And this is a broadly interesting cadre: Bukavina et al. note that the first five genera all have constituent species capable of degrading polycyclic aromatic hydrocarbons (PAHs), which are carcinogenic byproducts of combustion associated with bladder cancer and found in cigarette smoke and occupational exposures.[16,65,86,87] Such a finding suggests that the urobiome could have an impact on how PAHs interact with bladder cancer – *in vitro* work has shown that microbial metabolites of PAHs such as benzo[a]pyrene can be tumorigenic.[88–90] Using a gut simulator, Van de Wiele showed that consortia of human colon microbiota could transform benzo[a]pyrene and other PAHs into estrogenic metabolites.[90] A detailed discussion of the estrogen receptor and its contribution to bladder cancer is beyond the scope of this review; briefly, *in vitro* research with bladder cancer cell lines has shown the estrogen receptor to be tumorigenic and inductive of chemoresistance,[91] a pathway observed in multiple cancers.[92] Additionally, hydroxylated metabolites of PAHs can be carcinogenic in their

own right, typically forming mutagenic DNA adducts.[93] Because Bukavina et al.'s work was based in 16S amplicon sequencing, the strain and functional resolution of these putative PAH-degrading microbes are unconfirmed, and improvements to metagenomic urobiome sequencing that allow metabolic reconstruction are necessary to establish mechanistic links between microbial metabolism and cancer.

Enterococcus, Escherichia, and *Gardnerella* are all associated with rUTI, either as causative agents or as biomarkers of potential bladder dysbiosis.[51,94,95] Their concurrent association with bladder cancer in this study suggests a possible relationship between UTI and bladder cancer; this meta-analysis did not assess UTI history, but there is a well-documented link between bladder cancer and history of UTI or rUTI: in a cohort of 1809 bladder cancer cases and 4370 controls, bladder cancer patients were more likely to have a history of rUTI (odds ratio = 6.6 males, 2.7 females).[96]

One study, from our lab, assessed differences in urobiota and fecal microbiota according to bladder cancer status in companion dogs using 16S rRNA sequencing. The study cohort was small (n=7 dogs with bladder cancer, n=7 age-, sex-, breed-matched controls) but used extremely strict exclusion criteria (recent antibiotics, history of chemotherapy, low read depth) to ensure high-quality samples and data.[13] In line with some human studies,[16] urobiome alpha diversity was decreased in dogs with cancer. The urobiota were also different in composition between healthy and bladder cancer samples according to the unweighted UniFrac beta-diversity metric, suggesting that the cancer-affected urinary tract has a urobiome with altered phylogenetic makeup compared to healthy urinary tracts. Most interestingly, the genus *Fusobacterium,* constituent strains of which are known tumor promoters,[68] was present in urine samples from cancer-affected dogs but not those from healthy dogs.

Three studies have assessed bladder *tissue* microbiota, as opposed to urine samples that are surmised to contain tissue-colonizing microbes. Pederzoli et al. (2020) found, via 16S sequencing of the tissues, no difference between cancer and healthy tissue.[79] Liu et al. (2019) report an increased abundance of *Acinetobacter, Escherichia,* and *Sphingomonas* in cancer tissue, as well as a depletion of health-associated *Lactobacillus*.[77] Mansour et al. (2020) compared tumor microbiota to urine microbiota on a within-patient basis, using a small (n=4) cohort and found little overlap between the two sample types, a finding that runs directly counter to supplementary analyses from Pederzoli et al., which found 80% overlap of ASVs within patients across sample types.[78,79] Tumor-associated microbiota may be deeply involved with the immune microenvironment of the tumor, and may vary according to tumor type,[67,97] so this line of work warrants further study with larger cohorts and optimized protocol.

The state of urobiome research

Despite extremely important advances in our understanding of the urobiome, the state of the research can be broadly delineated into two methodological disciplines, each with overlapping but unique limitations. There are mechanistic, *in vitro* studies that reveal and recapitulate important processes but are limited in terms of the diversity they can capture, which may rely on cultures of urine isolates or use in vitro models;[19,41,42] and there are sequencing based studies, which, with few exceptions, are largely 16S rRNA amplicon studies of DNA extracted from urine samples or whole genome sequencing of cultured isolates, stratified by some clinical factor of interest.[18,21,39,94] These studies are generally limited to correlational, rather than mechanistic, conclusions, though sequencing of isolates does provide functional insight into disease-associated genomes.[94] Shotgun metagenomic sequencing and metabolic reconstruction of urinary

microbes can begin bridging the gap between microbes correlated with urinary tract states and the *functions* of those microbes and can better guide *in vitro* studies on microbes of interest based on their functions. Efforts to optimize shotgun metagenomic sequencing of urine are limited by characteristics of the urinary tract.

Section 2 – Urobiome Challenges

Technical Difficulties: Urobiota are elusive

While it is true that urobiome research has lagged that of the microbiota of other body sites because of clinical dogma, work characterizing the urobiome is also hampered by aspects of the urinary tract that affect microbial DNA extraction, sequencing, and bioinformatic processing. The urinary tract has a distinct and diverse microbiota, but microbial biomass is low overall and may often be overwhelmed by biomass from shed host cells.

Low microbial biomass and associated challenges

A urine sample from a healthy individual may contain fewer than 100 CFU and up to 10,000 CFU per milliliter of urine.[39] In comparison, a gram of feces may contain 10-100 billion CFU,[98] a difference of at least five orders of magnitude. This low relative microbial biomass requires special considerations for the identification of contaminant sequences, sample collection, and DNA extraction.

Contaminants are sequences – metagenomic or amplicon – in a given dataset that were sequenced with a sample but are not truly from the source community.[57] Such contaminants can be stray DNA or bacterial cells present in laboratory reagents, surfaces, and supplies that are

introduced to a sample during any step of sample collection and processing.[56] Contaminants may also include mis-assigned nonmicrobial sequences resulting from mis-mapping to references or the inclusion of contaminant sequences in reference databases.[99,100] In a fecal sample containing the DNA of 100 billion bacterial cells, the DNA content of a stray reagent contaminant is relatively negligible and unlikely to influence the resultant profile post-sequencing, though it may lead to false conclusions about constituent microbes.[57] However, in low microbial biomass environments, such as the urinary tract, vagina, breast milk, and respiratory passages, such contaminant sequences may constitute a large fraction of the sequenced DNA and lead to completely spurious results.[99] To characterize this problem, Salter et al. (2014), noting that their sequenced negative controls often contain bacterial diversity despite having no template, used 16S rRNA amplicon sequencing on pure *Salmonella bongori* cultures.[56] They sequenced serial dilutions of culture (10^8-10^3 cells/input) at multiple sequencing sites; across sites, non-*S. bongori* sequences (contaminants) increased in abundance and diversity with each successive dilution. Contaminating sequences became dominant at 10^3 cells of input DNA, and contaminant sequences common across sequencing sites were presumed to be from DNA extraction reagents and supplies; site-specific contaminants were presumed to be introduced from library preparation and sequencing centers, highlighting the multiple sources of contaminants. They further extracted pure samples using different production lots of the same supplies and different commercially available extraction kits, and contaminants were not consistent or predictable across these variables.[56]

The identification and treatment of contaminants has been subject to significant debate among researchers. In 2014, Aagard et al. identified a placental microbiome, albeit low biomass, and associated compositional differences with pre-term birth rates.[101] Follow up studies

comparing negative controls to placental samples found extreme biases introduced by extraction kit and batch, and could not distinguish placental microbiota from negatives, suggesting that Aagard et al. identified solely contaminant microbiota.[102,103] And there is ongoing debate about the presence of blood microbiome, and whether blood microbiota may be used to identify cancer signatures, or whether these signals are simply contamination.[104,105] Notably, many of these signals depend on how researchers bioinformatically handle putative contaminating sequences. In the case of microbial biomarkers of cancer in the blood, criticism of the original paper establishing those biomarkers identified several cases where microbial reads associated with cancer came from putative extremophiles, which are unlikely to be present in human tumors but are somewhat common as contaminants.[105]

Several strategies for the identification and filtering of contaminating sequences have been proposed, relying on a mix of statistical methods and biological inference. Some researchers simply subtract reads sequenced in negative controls from all other samples,[106] while others remove reads that are assigned to taxa commonly found in reagents.[104] However, a number of software tools based on statistical patterns of contaminating sequences have been released and variably utilized. Karstens et al. (2019)[107] compared four different strategies for handling contaminants using 16S sequencing of mock communities that were serially diluted to progressively increase the proportion of contaminants: directly filtering those sequences found in negative controls; filtering based on an abundance threshold; filtering using *Decontam,*[57] which identifies contaminants statistically based on the inverse relationship between sample DNA concentration and the relative abundance of putative contaminants as well as the prevalence of contaminants in negative controls; and *SourceTracker*,[108] software that predicts the proportion of a sample arising from a known contaminant source. *Decontam* removed the greatest proportion

of contaminants, and when combined with abundance filtering, was most effective at restoring the expected mock community by removing contaminant sequences,[107] though the authors note that *SourceTracker* may work more effectively when working in well-defined microbial communities. *Decontam* remains field standard and is used in urobiome studies from our lab. In urine samples, common soil microbes and kit contaminants are often identified as putative urine microbes, such as *Sphingomonas* and *Ralstonia*.[13,16] However, these microbes often also have clinical relevance as opportunistic pathogens,[109] may or may not appear in negative controls, and could plausibly be present in the urine or on skin. Microbiome research with low biomass substrates, therefore, requires careful consideration of the substrate environment as well as thorough negative controls.

Contamination aside, one might also attempt to address the low microbial biomass in two ways: 1) by optimizing DNA extraction to maximize bacterial DNA yield,[22,110] or, 2) by collecting more sample. Multiple commercial DNA extraction kits are available that are designed to allow convenient processing of multiple small samples at once, and theoretically can be used for multiple sample types. The basic steps of microbial DNA extraction are as follows: 1) removal of extracellular DNA and PCR inhibitors (proteins, minerals, fatty acids, etc.) from the samples, 2) lysis of bacterial cells, 3) purification of microbial nucleic acids.[111] Protocols may vary in how they approach each step: lysis may be achieved chemically or mechanically, or both, and purification can be achieved using magnetic beads, silica-based columns, or via ethanol precipitation.[111] The choice of extraction method introduces bias; methods may variably preferentially lyse certain clades, introduce variable amounts of contaminants, or perform differently on different substrates.[112] The Microbiome Quality Control Project assessed this by extracting replicate samples of stool, simulated stool, chemostat, and negative controls via

multiple commercial kits at multiple sites before submitting to amplicon sequencing with a variety of sequencing platforms and primer protocols.[112] Sample handling (i.e., extraction) introduced the largest variation between replicates, while bioinformatic protocols and sequencing platform were less important, though significant. This variation reduces inter-study comparability and necessitates the field moving to standardized extraction protocols. We expect that this inter-study variability is amplified in low microbial biomass substrates like urine.

Commercial DNA extraction kits are often designed specifically for high or low biomass substrates, or even for specific microbial environments (e.g. Qiagen PowerFecal specifically for stool samples).[113–116] However, they perform variably well within a given substrate, and differently between substrates. Knudsen et al. (2016) extracted DNA from human feces, pig feces, and sewage (all extremely high biomass) using seven different commercially available kits. Extraction method impacted the amount of DNA extracted, the alpha diversity of the samples, and the overall composition of samples within and between sample types. For example, one procedure (Easy-DNA) preferentially lysed gram-positive bacteria in human waste and sewage, but not in pig feces.[113] Similar results were shown in a study optimizing DNA extraction from samples of first-pass meconium, a low microbial biomass environment.[114] It is therefore clear that DNA extraction protocols must be optimized on a per-substrate, per-situation basis, and the field has not yet met this standard.

Initial work to do so for urobiome research is ongoing in our lab and others. Mrofchak et al. (2021) extracted DNA from 3mL replicate aliquots of canine urine using five different commercial methods, including the four most commonly used in published urobiome research, designed for substrates of different levels of biomass.[22] The Qiagen QIAamp BiOstic Bacteremia Kit (Bacteremia), which uses chemical, mechanical, and thermal lysis and isolates DNA using a

silica-based column, is designed for DNA extraction from potentially low microbial biomass substrates like blood and urine.[117] This kit maximized the total bacterial DNA that could be extracted from urine based on a 16S gene qPCR,[118] as well as the number of samples overall that had quantifiable bacterial DNA and the number of 16S reads per sample. There was no difference in diversity or composition according to extraction method; the authors recommend Bacteremia for urine samples but note that the choice of kit did not broadly impact the 16S profiles generated.

Karstens et al. (2021) performed a nearly identical experiment using human urine, and included two of the kits (Bacteremia, Blood & Tissue) assessed in Mrofchak et al.[110] This study found that Blood & Tissue maximized total DNA yield, but did not assess bacterial DNA specifically, and Blood & Tissue samples did not have significantly different DNA concentrations from Bacteremia samples in pairwise comparisons. They found no difference in any other variable (sequencing depth, diversity, composition), suggesting that Bacteremia performs at least as well as other extraction methods in both human and canine urine for 16S rRNA profiling of the urobiome.

Simple reasoning suggests that urine samples of greater volume will contain greater absolute counts of bacterial cells, and that extracting DNA from larger volume samples might result in higher bacterial DNA yield. Given the sparsity of cells in urine (100-10,000CFU/mL), it is possible too that there exists a threshold volume of urine at which a true replicate of the overall community is represented. Collecting and sequencing a volume below such a threshold would subject the resulting profile to stochasticity as low abundance microbes are excluded or included randomly (**Fig 1**). In a high biomass environment like feces, this threshold is likely to be so small as to be unimportant in the overall sampling process. However, in urobiome studies, there is high

variability in the volume of urine that is collected and pelleted for DNA extraction: some studies use as little as 0.5mL[17] of initial sample, and some use as much as 50mL.[119] Karstens et al. (2018) report different rates of samples having nonzero sequencing depths according to sample volume.[55] Additionally, observations that diversity and nonzero sequencing rates may be decreased in samples collected via suprapubic aspiration or transurethral catheterization compared to free catch samples suggest that sample volume threshold may be different according to collection method.[16] However, the relationship between sample volume, sequencing depth, contamination, and community stochasticity has not been thoroughly investigated. In Chapter 2, new data addressing these relationships using canine urine collected via free catch are presented.

An additional ripple: potential for high host biomass in urine samples and the quest for genomes

In 16S amplicon studies, host biomass is relatively unimportant; primers targeting bacterial sequences with high specificity and well-characterized off-target effects (such as amplification of mtDNA) make host nucleic acids generally unproblematic.[30,39] However, shotgun metagenomic strategies sequence all of the DNA in a sample in an untargeted manner.[30] In high host biomass, low microbial biomass substrates like urine, tissue, breast milk, and respiratory passages, host biomass represents an extremely high barrier to reconstruction and analysis of microbial genomes.[120–127] For example, metagenomic sequencing of lung sputum samples resulted in 99.2% of reads assigning to the host (human),[120] and 95.9% host in vaginal samples.[127] In one of two studies using shotgun metagenomics to characterize the urobiome, human reads were reported to comprise 1.3%-99.9% of the sequencing depth in samples from patients with UTI;[27] in this scenario, the expected proportion microbial biomass would vary: microbial biomass is presumed to increase due to overgrowth of a single pathogen or

polymicrobial overgrowth, but increased host shedding is also expected due to upregulation of urothelial exfoliation during UTI.[9,128,129] Normal urine may contain 0-100,000 host cells/mL,[130] and shedding is expected to be increased in certain health states including incontinence and bladder cancer.[128]

Metagenome-assembled genomes (MAGs) cannot be assembled and analyzed with confidence without adequate microbial sequencing depth, and even read- or gene-level analysis of microbial metagenomes are limited by sequencing depth.[131] Construction of MAGs is critical for tying specific microbiome constituents to specific functions *in silico*, and high quality MAGs have high degrees of coverage (multiple sequenced copies of the genome).[131] If a metagenome of 10 million reads is 99.9% host, only 10,000 reads of 150-250 base pairs may be microbial, spread across multiple taxa; MAG assembly in this scenario is impossible, and read-based profiling, which relies on marker genes, is likely to be extremely limited, representing a wasted sequencing effort if microbial sequence information was the target.[132]

Inferring metabolic potential from sequence information is useful for characterizing the putative functional abilities of metagenomes and is achieved through read- or contig- level annotation. Lloyd-Price et al. (2019) used HUMAnN2 to functionally profile fecal metagenomes in patients with and without inflammatory bowel disease and were able to link functional enrichments to changes in the abundances of specific taxa and metabolites using multi-omics.[133]

However, MAG-resolved metagenomics allows critical insight into microbiome functions and strain variation and the impacts they can have on host health. Genome-resolved metagenomics revealed that the gut microbiota of humans exposed to high levels of environmental contaminants (xenobiotics) were enriched not only for specific taxa, but also for specific xenobiotic-degrading *genes*.[134] Because microbial sequences had been assembled into

MAGs, the xenobiotic functions could be traced to specific genomes – directly linking taxonomic enrichment to putative functional ability to degrade the contaminants. Such links can guide future *in vitro* work that can experimentally verify and characterize microbial functions, and generate druggable targets, identify toxifying or detoxifying pathways, and elucidate mechanisms by which xenobiotics might promote disease.[67,134] Urobiome researchers are especially interested in reconstructing such microbial mechanisms, given early indications that bladder cancer associated microbiota may be enriched for PAH degraders.[16] Additionally, these metabolisms have high translational relevance when considering companion dogs as a model system. As members of the household, dogs share similar exposure and cancer-risk profiles to their owners, including cigarette smoke (secondhand) and diesel exhaust, and may have higher exposure rates to other cancer-associated contaminants like pesticides, with which they come into greater contact than humans (dogs roll in, lick, and step on many more substances than most humans).[64] However, resolution of the host biomass problem is necessary to achieve the sequencing depth for MAG assembly and metabolic profiling. One small study by Kochroo et al. (2022) attempted to assess functional urobiome differences between kidney stone human patients and healthy controls via shotgun sequencing.[135] They assembled 17 draft microbial genomes with median completeness of 36.7% and performed gene-level metabolic profiling by annotating contigs, rather than the draft genomes. This allowed them to partially attribute an enrichment for oxalate-degrading genes in healthy controls to the uroprotective *L. crispatus*; however, this analysis is limited by the relatively low completion of the MAGs, which may stem from high proportions of host sequences (and therefore low microbial sequencing depth) and prevented a top-down approach of annotating MAGs directly to identify full pathways.

Deeper sequencing can improve microbial sequencing depth through "brute force" without changing the proportion of host to microbial reads, but this method is costly and suffers from diminishing returns in microbial depth.[132] Instead, focus has shifted to benchtop methods for depleting host biomass and nucleic acids prior to extracting and sequencing microbial DNA. A variety of commercial and laboratory methods for depleting host biomass are available and have variable efficacy in various substrates.[120,122,124,126,136] Similar to basic DNA extraction, there is evidence that host depletion steps may amplify community bias prior to sequencing.[137] There are two available "types" of host depletion. Some methods take advantage of the fact that mammalian cells are more vulnerable to osmotic stress than prokaryotic cells, and preferentially lyse mammalian cells in the sample using hypotonic or chaotropic buffers.[124] The resulting free host DNA can be degraded using Benzonase (a DNase) or light-activated propidium monoazide (PMA, which covalently bonds and cleaves DNA).[124,138] The resulting partial-lysate can then be processed through basic DNA extraction and cleaved DNA is removed along with other PCR inhibitors. This type of host depletion has been reported to introduce artificial enrichment for gram positive bacteria,[137] whose cell walls render them less susceptible to osmotic lysis than gram negatives, though at least one study has reported the opposite effect.[127]

The other "type" of host depletion targets CpG methylation, a pattern seen in mammalian, but not prokaryotic DNA. The NEBNext Microbiome DNA Enrichment Kit uses methylation-binding protein bound to magnetic beads to remove mammalian DNA and leave prokaryotic DNA in solution.[139]

There is broad variability in which host depletion method appears optimal according to substrate. In bovine milk, CpG targeting methods appear to work moderately well for maximizing bacterial DNA yield, and alpha diversity, and can be combined with pre-extraction

host depletion methods, though a host depletion kit (via preferential lysis) from Molzym (Germany) appears to work just as well. However, the NEBNext kit appears to be the only readily available CpG-methylated-DNA targeting method and is not frequently used across studies. In human respiratory tract samples, the methods using DNase to cleave extracellular DNA appears to be more effective methods at reducing the proportion of host reads than PMA-based methods.[120,140] In a study comparing host depletion methods (two lysis with DNase, and one lysis with PMA) in skin, vaginal, and oral swabs, osmotic lysis with DNase treatment was extremely effective at depleting host reads in vaginal samples.[127] However, the authors cautioned against host depletion in skin and saliva samples, which showed significant depletions of gram negative microbes. However, Marotz et al. (2018) designed a protocol that used PMA cleavage of extracellular DNA and found it was more effective at reducing host read proportions than commercially available kits in saliva samples.[124]

No study has assessed the efficacy or impact of host depletion methods in studies of the urobiome, though it is clear that methods must be tested on a per-substrate basis. In Chapter 2, we present data on the impact of host depletion of six different extraction methods on both 16S and shotgun metagenomic sequencing of the canine urobiome.

Next steps for the urobiome

Study of the urobiome will reveal insights on how microbes in the urinary tract interact with host health in a variety of urinary tract diseases of high burden, including UTI and bladder cancer. The functions of urobiome constituents are especially important to characterize to begin investigating mechanistic links between these constituents and health. The urobiome field must, therefore, answer critical questions to enable such characterization: 1) How much urine sample

should be collected for accurate, replicable, and low-contamination urobiome samples? 2) How does host depletion bias microbial communities pre-sequencing? 3) Are genome-resolved metagenomics and functional analyses of the urobiota possible? In Chapter 2, we present new data addressing and extending these three questions.

Chapter 2 - Evaluating urine volume and host depletion methods to enable genome-resolved metagenomics of the urobiome

Abstract

The gut microbiome has emerged as a clear player in health and disease, in part by mediating host response to environment and lifestyle. The urobiome (microbiota of the urinary tract) likely functions similarly. However, efforts to characterize the urobiome and assess its functional potential have been limited due to technical challenges including low microbial biomass and high host cell shedding in urine. Here, to begin addressing these challenges, we evaluate urine sample volume (100 µl – 5 mL), and host DNA depletion methods and their effects on urobiome profiles in healthy dogs, which are a robust large animal model for the human urobiome. We collected urine from seven dogs and fractionated samples into multiple aliquots. One set of samples was additionally spiked with host (canine) cells to model a biologically relevant host cell burden in urine. Samples then underwent DNA extraction followed by 16S rRNA gene amplicon and shotgun metagenomic sequencing. We tested six methods of DNA extraction: QIAamp BiOstic **Bacteremia** (no host depletion), QIAamp **DNA Microbiome**, **Molzym** MolYsis, **NEBNext** Microbiome DNA Enrichment, **Zymo HostZERO**, and **Propidium Monoazide**. In relation to urine sample volume, ≥ 3.0 mL resulted in the most consistent urobiome profiling. In relation to host depletion, individual (dog) but not extraction method drove overall differences in microbial composition. DNA Microbiome yielded the greatest microbial diversity in 16S rRNA gene sequencing data ($p=0.0025$) and shotgun metagenomic sequencing data ($p=0.01$), and maximized MAG recovery while effectively depleting host DNA ($p=0.0039$) in host-spiked urine samples. As proof-of-principle, we then mined MAGs for core metabolic functions and environmental chemical metabolism. We identified long chain alkane

utilization in two of the urine MAGs. Long chain alkanes are common pollutants that result from industrial combustion processes and end up in urine. This is the first study, to our knowledge, to demonstrate environmental chemical degradation potential in urine microbes through genome-resolved metagenomics. These findings lay the foundation for future evaluation of urobiome function in relation to health and disease.

Introduction

Alterations in the urobiome (microbiota of the urinary tract) have been associated with bladder cancer,[13,16] incontinence,[20] urinary tract infection,[18,44] and urolithiasis,[11,141] but the study of the urobiome remains fraught with technical challenges. Historicaly, culturing was employed to identify microbes and microbial functions (e.g., antimicrobial resistance) present in urobiome. However, standard urine culture captures very few members of the urobiota.[142] Expanded quantitative urine culturing methods (EQUC) have improved culture resolution to > 80% of taxa identified via 16S sequencing, but many urobiota remain uncultured, highlighting the need for effective culture-independent methods to profile urine microbes.[142] Thus, more recent research has focused on molecular techniques for higher resolution and more comprehensive evaluation of the urobiome.

Molecular methods for analyzing the urobiome remain challenging and have critical need for foundational technical research in the urobiome to overcome specific problems. First, urine generally contains low microbial biomass,[55] making urine samples vulnerable to contamination by microbes or microbial DNA introduced during extraction or sequencing.[54,107] Additionally, there are no evidence-based guidelines on minimum urine volumes for microbiome research, and studies on the urobiome range from using 0.5 mL[17] to 50 mL of urine.[119] Critically, there are

conditions (e.g., urinary tract inflammation), populations (e.g., pediatric), and model species (e.g, dogs, rodents) for which collecting 10 mL of urine or more in a single void may be infeasible. Finally, urine can contain a high burden of host cells, especially in diseased states such as urinary tract infection or bladder cancer,[42,52,128] which can complicate DNA extraction, introduce noise in 16S rRNA profiling,[143] and overwhelm shotgun sequencing attempts with host reads rather than microbial reads. This then limits our ability to understand the functional potential of the urobiome and how these functions drive health and disease.

Commercial DNA extraction methods and pubished protocols that include host cell and DNA depletion are available, but these methods have not been comparatively evaluated in urine. In this study, we assess four commercially available DNA extraction kits with host cell depletion (MolYsis Complete5; NEBNext Microbiome DNA Enrichment Kit; QIAamp DNA Microbiome Kit; and Zymo HostZERO) as well as a protocol using light-activated propidium monoazide, and compare them to a method with no host depletion (BiOstic Bacteremia). Host depletion has been successful in other low-microbial-biomass, high-host-biomass substrates including breast milk, oral, respiratory tract, and tumor samples.[110,122,124,136,144] For example, in saliva samples, two host depletion methods reduced the host read proportion from 95% to < 30%, thereby improving the microbial resolution of shotgun metagenomics.[124] Host depletion methods offer promise for improving characterization of urobiome structure and function, but require evaluation for efficacy in urine samples.

The urobiome has been characterized via culture, whole genome sequencing of urine isolates, 16S rRNA sequencing, and shotgun metagenomic sequencing. However, few studies have reported metagenome-assembled microbial genomes (MAGs) and genome-resolved community analyses of the urobiota.[21,27,94,135] Bioinformatic construction of MAGs from urine

would allow for more thorough functional reconstruction of the urobiome, including rare and unculturable taxa, revealing potentially important mechanistic links between the urobiome and disease in a genome-resolved fashion.[30,145]

In this study, we tested several approaches for studying the urobiome using urine from healthy dogs. Dogs are a robust translational model for the human urobiome[13,62,146,147] and for urinary tract diseases, including bladder cancer[61] and urinary tract infection.[62,146,147] We specifically set out to **i)** assess the impact of urine sample volume on urine microbial community profiles (**Fig. S1**), **ii)** determine how DNA extraction methods that include host depletion affect urobiome profiles (16S and shotgun metagenomics) (**Fig. S2**), **iii)** determine if we could sufficiently reconstruct metagenome assembled genomes (MAGs) of urine microbes from shotgun metagenomic data to then mine them for relevant microbial functions, and **iv)** assess if and how urine microbes metabolize environmental chemicals linked with urinary tract diseseases like bladder cancer.

Methods

Urine Volume Experiment

The goal of this first experiment was to determine if microbial community profiles and the presence/abundance of microbial contaminants (e.g. from reagents, kits, etc.) differed by urine sample volume (Experimental Design: **Fig. S1**).

Subject Recruitment

Healthy dogs were recruited through the Ohio State University Veterinary Medical Center (IACUC: 2020A00000050). Each dog underwent a comprehensive, physical exam, blood work

(serum chemistry, complete blood count), urinalysis, and urine culture. All dogs were between one and ten years of age, weighed at least 20 lb with a body condition score of 4 or 5 (out of 9) and normal muscle condition. Dogs with a history, physical examination findings, clinical signs, or laboratory abnormalities consistent with urinary tract, liver, kidney, or gastrointestinal disease were excluded. Dogs with any history of antibiotic use, chemotherapy, or radiation in the past three months were also excluded (**Table S1**).

Urine Sample Collection & Preparation

Midstream, free-catch urine was collected and stored from 5 healthy dogs as described previously.[148] Urine samples were fractionated into 0.1, 0.2, 0.5, 1.0, 3.0, and 5.0 mL aliquots prior to DNA extraction. Samples were centrifuged at 4°C and 20,000g for 30 minutes. Following centrifugation, supernatant was discarded, and the pellet was saved. The pellets were then used for DNA extractions.

DNA Extraction & Quantification

DNA was extracted using the QIAamp BiOstic Bacteremia DNA Kit (Bacteremia; Qiagen, Hilden, Germany), as described previously.[22] This kit does not include host depletion steps. Briefly: pellets were resuspended in a lysis buffer and underwent two rounds of bead beating at 6m/s for 60s in an MP FastPrep-24 5G (MP Biomedicals, Solon, OH). Following bead beating, samples were cleaned using the kit's inhibitor removal solution and processed according to manufacturer protocol. All centrifugation steps were conducted at 13,000 x g and, in the final step, samples were eluted twice through the silica membrane to maximize DNA yield. DNA

concentrations were quantified using a Qubit Fluorometer (ThermoFisher Scientific, Waltham, MA).

16S rRNA Gene Amplicon Library Preparation and Sequencing

DNA then underwent library preparation and sequencing at Argonne National Laboratory (Lemont, IL), as described previously.[22] Briefly: we used primers 515F and 806R to amplify the V4 region of the 16S rRNA gene, followed by paired-end amplicon sequencing via Illumina Miseq (2x250). Sequences are available at NCBI Bioproject PRJNA1109516.

16S rRNA Gene Amplicon Sequence Processing and Statistical Analyses

Raw sequences were processed using QIIME2 v.2023-5. Reads were denoised and clustered into amplicon sequence variants (ASVs) using DADA2[149] with the following parameters: 5 base pairs (bp) were trimmed from the 5' end of each read and forward reads were truncated at 225 bp while reverse reads were truncated at 220 bp. Putative contaminant reads were identified and removed using the R package *decontam*[57] with prevalence-based filtering (threshold = 0.5) (**Table S2**). Microbial contaminants are microbes or microbial sequences that get introduced during the extraction, library preparation, or sequencing process. These contaminants are putatively identified based on their tendency to be more prevalent or abundant in negative (n=6) control samples. Contaminant read counts were exported into a new table for analysis. Contaminant abundances were calculated by dividing each count by the total number of 16S reads in each sample, and contaminant abundances between groups were statistically compared using the Friedman test. Taxonomy was assigned using the Silva 138 99% OTU 515F/806R classifier. Sequences assigning to mitochondria or eukaryotes and unassigned

sequences were removed. A total of 37 samples were sequenced, and sequencing depth ranged from 1-30,408 reads. Samples with fewer than 4,125 reads were excluded from analyses, and remaining samples were sampled to even depth. This excluded all negative controls and 6 true samples which were largely low volume samples (3 samples from dog ArB (0.1 ml, 0.5 ml, 3 ml), 1 sample from dog MS (0.1 ml), and two samples from dog FC (0.1 ml, 0.5 ml).

For all analyses, statistical significance was set at $p<0.05$. Microbial diversity (Shannon Index, Observed Features, and Faith's Phylogenetic Diversity) and distance metrics (Bray Curtis, Jaccard, and UniFrac) were calculated and tested using QIIME2 and the R packages *phyloseq* and *vegan*. Differences in bacterial diversity were assessed via t-test, Friedman test, or Kruskal-Wallis test depending on the normality and pairing of the data, and pairwise comparisons were conducted using the Benjamini, Krieger, and Yekutieli procedure for controlling the false discovery rate (FDR) at $Q=0.05$.[150] Differences in microbial composition were assessed via PERMANOVA, with $Q=0.05$ for FDR adjustments in pairwise comparisons.

Host Depletion – 16S

The goal of this experiment was to evaluate how DNA extraction methods that include host depletion steps affected bacterial DNA recovery and microbial community profiles (Experimental Design: **Fig. S2**). Subject recruitment occurred as described above.

Urine Sample Collection, Host Cell Spiking, & DNA Extraction

We collected mid-stream free catch urine from seven healthy dogs (**Table S1**). Urine was then aliquoted into two batches: one batch was spiked with canine cells (canine thyroid adenocarcinoma cells[151] – CTAC) to a concentration of 75,000 cells/mL, to model a biologically

relevant host cell concentration in urine from healthy dogs.[130] The other batch remained unspiked. Urine samples were then pelleted as described above. All urine samples then underwent DNA extraction using six different extraction methods: QIAamp BiOstic Bacteremia DNA Kit (Bacteremia; Qiagen, Hilden, Germany); MolYsis Complete5 (Molzym, Bremen, Germany); NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, MA); QIAamp DNA Microbiome Kit (Qiagen, Hilden, Germany); HostZERO Microbial DNA Kit (Zymo Research, Irvine, CA); and a protocol using light-activated propidium monoazide described in Marotz et al., 2018.[124] All of these methods except QIAamp Bacteremia included host depletion steps. In addition to urine samples, we also included a positive control sample (ZymoBIOMICS Gut Microbiome Standard, Zymo Research, Irvine, CA, **Table S3**) that we extracted with each method. The ZymboBIOMICS gut microbiome standard contains 21 microbial strains, including 18 bacteria, 2 microbial eukaryotes, and one archaeaon. Samples were extracted according to the respective manufacturers' protocol, with modifications described below. Each extraction run included a negative control that was sequenced along all samples (n=5). All extracted DNA was stored at -80°C until library preparation and sequencing. Unspiked samples underwent 16S rRNA gene amplicon sequencing; spiked samples underwent shotgun metagenomic sequencing (described under Host Depletion – Shotgun Metagenomics). Because 16S rRNA gene amplicon sequencing targets bacterial sequences, host spiking would not reveal additional information in that setting.

DNA Extraction Methods

QIAamp BiOstic Bacteremia (Qiagen)

No host depletion is included in this protocol. Protocol is detailed above under Urine Volume Experiment. Prior to extraction, the ZymoBIOMICS Gut Microbiome Standard was centrifuged at 20,000g and the supernatant was saved. To maximize DNA recovery per recommendations from Zymo, the pellet was processed through the kit, and the supernatant was added to the MB spin column and centrifuged at 13,000g for 1 minute. The flow-through was discarded, and lysate from the pellet was added per the manufacturer's protocol.

Molzym MolYsis Complete 5 (Molzym MoYsis)

This method uses a chaotropic buffer to selectively lyse host cells then removes host DNA using a DNAase prior to extracting microbial DNA. Samples were extracted following the manufacturer's protocol.

NEBNext Microbiome DNA Enrichment Kit (NEBNext)

This method uses nonselective lysis followed by selective binding and depletion of CpG-methylated host DNA in order to enrich microbial DNA recovery. Samples were first extracted using the QIAamp BiOstic Bacteremia DNA Kit and frozen at -80°C. Samples were defrosted and further extraction was performed according to the NEBNext manufacturer's protocol. For samples that did not have detectable DNA after the initial extraction, a threshold of 0.05 ng/µL was used to calculate the MBD2-Fc Protein to Protein A magnetic bead value. A solution of MBD2-FC Protein and Protein A magnetic beads was prepared and aliquoted into each sample accordingly. To avoid DNA loss, purification was not performed at the end of the protocol (neither option A nor B).

QIAamp DNA Microbiome Kit (DNA Microbiome)

This method uses selective osmotic lysis and Benzonase to degrade host cells and digest host DNA prior to extraction of microbial DNA. A Thermomixer at 600 rpm was used instead of end-over-end rotation. Prior to extraction, the ZymoBIOMICS Gut Microbiome Standard was centrifuged at 20,000g and the supernatant was saved. To maximize DNA recovery per recommendations from Zymo, the pellet was processed through the kit, and the supernatant was added to the MB spin column and centrifuged at 13,000g for 1 minute. The flow-through was discarded, and lysate from the pellet was added per the manufacturer's protocol at step 12.

Zymo HostZERO Microbial DNA Kit (Zymo)

This method uses selective osmotic lysis followed by enzymatic degradation of DNA to degrade host cells and host DNA prior to extraction of microbial DNA. A FastPrep-24 5G bead beater was used for optimized lysis (Appendix D of manufacturer's protocol). Extraction proceeded following the manufacturer's protocol. Samples were eluted with 20-26 uL ZymoBIOMICs DNase RNase-Free Water.

Propidium Monoazide (PMA)

This method uses PMA to intercalate the DNA of membrane-disrupted host cells, and light activation triggers covalent bonding between dsDNA and PMA, fragmenting the DNA. Samples were pretreated with 10uM PMA as described in Marotz et al. (2018), beginning with resuspending urine pellets in 200uL sterile water. After PMA treatment, samples were stored at -20°C and then extracted using the Qiagen QIAamp BiOstic Bacteremia kit.

DNA Quantification and 16S rRNA Gene Amplicon Sequencing

In both spiked and unspiked samples, we quantified total DNA via Qubit fluorometer and bacterial DNA via qPCR using universal 16S bacterial primers as described previously.[22,118] Bacterial concentrations were compared between groups using either Friedman tests or Kruskal-Wallis tests. Finally, we analyzed microbial community profiles (16S rRNA sequencing) in each sample. Library preparation, sequencing, decontamination (**Table S4**), and analysis were conducted as described above in the Urine Volume Experiment with the following DADA2 parameters: 5 bp were trimmed from the 5' end of each read and forward reads were were truncated at 250 bp while reverse reads were truncated at 231 bp. One urine sample (Dog SJ, Extraction Method: Molzym Molysis) appeared to be cross-contaminated with DNA from the ZymoBIOMICS Gut Microbiome Standard and was excluded from analysis (**Fig S3**). Statistical analyses were performed as described above (see Urine Volume Experiment) to assess differences by extraction method.

16S rRNA Gene Amplicon Sequence Processing and Statistical Analyses

16S sequencing processing and statistical analyses were performed as described above (see Urine Volume Experiment) to assess differences in microbial community diversity and composistion by extraction method.

Host Depletion – Shotgun Metagenomics

The goal of this experiment was to assess the efficacy of host depletion by extraction method and the viability of performing genome-resolved metagenomics on low biomass urine

samples. To do this, we used the same urine samples and the ZymoBIOMICs Gut Microbiome Standard positive control from the Host Depletion – 16S experiment described above and spiked them with host (canine) cells (Experimental Design: **Fig. S2**). DNA was extracted using the same 6 methods as described above under Host Depletion -16S.

Shotgun Metagenomic Library Preparation and Sequencing

Samples underwent shotgun metagenomic sequencing at the Ohio State University Infectious Diseases Institute – Genomics and Microbiology Solutions (IDI-GEMS) Laboratory. Metagenomic libraries were prepared following the Illumina (San Diego, CA) DNA Library Prep protocol with the following modifications: 1) Illumina's (M) beads were substituted with (L) beads to obtain larger insert sizes, 2) 9 or 12 PCR amplification cycles were used based on sample DNA concentration (Qubit) (**Fig. S4**), and 3) library purification was performed using a 1:1 sample to bead ratio. Samples were barcoded using IDT for Illumina UD Indexes. Tagmentation-based library construction has been validated and adopted as a standard operating procedure within the IDI-GEMS Laboratory to characterize the presence of microbes in samples and was recently shown to be an effective repeatable method for microbiome analysis of the human gut.[152] Metagenomic libraries were sequenced targeting a minimum of 50 million 2x150 base pair paired-end reads using an Illumina NextSeq2000. Negative extraction (n=5) and sequencing (n=2) controls were sequenced along with samples. Sequences were processed using the Ohio Supercomputer.[153] Sequences are available at NCBI Bioproject PRJNA1123238.

Metagenomic Sequence Processing and Statistical Analyses

Raw reads from the Illumina sequencer were quality filtered and trimmed of adapters using Trimmomatic.[154] Host reads were quantified by mapping to a concatenated canine and feline genome with CoverM.[155] Reads not assigned host were assumed microbial. Read counts were compared across extraction methods using the Friedman test. Taxonomy and abundance tables for microbial community profiling of metagenomes were generated using MetaPhlAn4.0[156] and SingleM[157] and SingleM condense. Metagenomes were *de novo* assembled into contigs using MEGAHIT[158] and quality assessed with QUAST.[159] Contigs were binned into metagenome-assembled genomes (MAGs) using portions of the MetaWRAP[160] pipeline, which combines the binning methods MetaBat2,[161] MaxBin2,[162] and CONCOCT,[163] and chooses the highest-quality representative of each bin from across these automated methods. dRep[164] was used to dereplicate MAGs at 99% average nucleotide identity, and CheckM[165] was used to evaluate MAGs for completeness and contamination. Only medium (>70% completion and <10% contamination) and high (>95% completion and <5% contamination) quality MAGs were retained for analysis. GTDB-Tk[166] was used to assign taxonomy to MAGs according to the Genome Taxonomy Database. Abundance tables of MetaPhlAn, SingleM, and MAG profiles were processed using *decontam* to identify putative contaminants. Because MetaPhlAn generates species-level taxonomic assignments, genera were also manually filtered: taxa commonly identified as kit contaminant genera[56] present in at least one negative control sample were bioinformatically removed, even if they were not filtered by *decontam* (**Table S5**). Additionally, reads assigned to taxa from the Zymo Gut Microbiome Standard in urine samples profiled with MetaPhlAn or GTDB-Tk were considered putative cross-contaminants and were removed from those samples. Diversity and community composition metrics from metagenomic data as well as read-level statistics were analyzed using the R packages *phyloseq,*[167] *vegan,*[168] and *tidyverse*.[169]

Alpha diversity was compared between kits using Friedman tests, and comparisons between dogs were performed using Kurskal-Wallis. Pairwise comparisons were conducted using the Benjamini, Krieger, and Yekutieli procedure for controlling the false discovery rate (FDR) at Q=0.05. Differences in microbial composition were assessed via PERMANOVA, with Q=0.05 for FDR adjustments in pairwise comparisons. Genes in MAGs were annotated using DRAM.[145]

Hydrocarbon Degradation Profiing

As a proof-of-principal test, we then mined the MAGs for microbial functions of interest including urea utilization and environmental chemical degradation. These functions are relevant as urine is a urea rich environment, and environmental chemicals, such as polycyclic aromatic hydrocarbons have been associated with urinary tract diseases like bladder cancer. Urea utilization was identified by searching within the DRAM output. To idenitfy putative hydrocarbon degrading genes, we queried custom, curated, published Hidden Markov Model (HMM) profile databases: aerobic degradation of polycyclic aromatic hydrocarbon pathways (PAHp),[170] and markers for the activation of various hydrocarbons (CANT-HYD).[171] Coding genes called by DRAM were queried against these databases using the hmmsearch function of HMMER (version 3.3)[172] and filtered to a maximum expect-value (e-value) of 1e-10. The full scores were compared to the score cutoffs specific to each gene in the database, i.e., gather cutoffs for PAHp and noise or trusted cutoffs implemented by CANT-HYD. Given the potential for high stringency in profiles generated largely from a few well-characterized model organisms, these cutoffs were relaxed to a minimum of 80% of the gather cutoff and 90% of the noise and trusted cutoffs for the respective databases.

Results

Urine Volume Influences Contamination, Diversity, and Composition

Current urobiome studies vary widely in the volume of urine used for profiling microbial communities. Moreover, low biomass samples, like urine, are highly susceptible to contamination by microbes or microbial DNA (hereafter referred to as "contaminants") that can be introduced during the DNA extraction and sequencing process. As such, in this experiment, we first assessed the relationship between urine sample volume and microbial contaminant load. Contaminants, as identified by *decontam* (**Table S2**), were at significantly lower relative abundances in urine samples of greater volume (**Fig. 1A**, **Table S6**, $p = 0.026$, Friedman).

We then evaluated bacterial diversity and composition by urine sample volume. Microbial richness, or the total number of unique ASVs in each sample, increased significantly with sample volume (**Fig. 1B, S5, Table S7,** $p = 0.015$, Friedman). Sequencing reads also increased with urine sample volume; although, this difference was not significant (**Fig. 1C**, $p=0.075$, Friedman).

Bacterial composition, however, did not differ significantly by urine sample volume but did differ significantly between dogs (**Fig. 2A, S5,** between dogs: $p = 0.001$, by urine sample volume: 0.98, Bray-Curtis, PERMANOVA), indicating that inter-dog differences overwhelmed differences based on sample volume. We next evaluated within-dog microbial composition by sample volume. Within each dog, the 3 mL and 5 mL samples were more consistent in microbial composition, while the 0.1, 0.2, 0.5 and 1 mL samples were more variable (**Fig. 2B, S6**). Based on this pattern, we grouped 3 mL and 5 mL urine samples into a "High" volume group, and the remaining urine volumes into a "Low" volume group. There was no significant difference in microbial composition between the High and Low groups ($p=0.6$, PERMANOVA); however,

High volume samples had significantly less variable microbial communities than Low volume samples, indicating that Low volume samples are more subject to stochasticity (**Fig. 2C, S6**; p = 0.0017, PERMDISP). Based on these results, we proceeded to use 3 mL urine samples for subsequent experiments.

Host Depletion – 16S

Healthy urine contains shed host epithelial cells at a relatively low abundance. However, in the presence of urinary tract disease (e.g., urinary tract infection, bladder cancer, bladder stones), host cell shedding can dramatically increase. There are multiple DNA extraction methods that incorporate host cell / host DNA depletion steps to facilitate microbial DNA recovery. In this experiment, we evaluated how six different extraction methods affected DNA concentrations and microbial community profiles. Extraction methods included: QIAamp BiOstic Bacteremia DNA Kit (Bacteremia); MolYsis Complete5 (Molzym); NEBNext Microbiome DNA Enrichment Kit; QIAamp DNA Microbiome Kit (DNA Microbiome) HostZERO Microbial DNA Kit (Zymo HostZERO); and a protocol using light-activated propidium monoazide described in Marotz et al., 2018.[124] All methods except Bacteremia included host depletion steps. The Bacteremia extraction method was included for reference here because this method has already been validated as an optimal method for profiling canine urine microbial communitites,[22] and it has been applied across multiple urobiome studies in humans and animals.[13,18] However, it has not been tested against extraction methods that include host depletion steps, which we did here.

We first compared how each extraction method impacted total and bacterial DNA concentrations derived from urine samples. We also compared DNA concentrations in urine

samples that were unspiked versus those spiked with host (canine) cells. While healthy mid-stream free-catch urine contains a low abundance of host cells, we opted to spike additional canine cells into urine at biologically relevant concentrations to best assess the host depletion capabilities of each extraction method. In unspiked samples, Bacteremia and NEBNext recovered the greatest total DNA concentrations (host + microbial); although, this result was not significant ($p=0.62$, Friedman, **Fig. 3A).** Bacteremia, DNA Microbiome, and Molzym MolYsis demonstrated significantly greater *bacterial* DNA recovery than propidium monazide, Zymo HostZERO, and NebNEXT; although no pairwise comparisons were significant (overall $p=0.014$, Friedman) (**Fig 3B).** In spiked urine samples, Bacteremia and NebNEXT recovered significantly greater total DNA than all other extraction methods (**Fig. 3C,** overall $p<0.0001$, Friedman, pairwise p between Bacteremia or NebNEXT and all other methods <0.05), while DNA Microbiome recovered the most bacterial DNA; although, overall differences in bacterial DNA concentrations by extraction method were only marginally significant (**Fig. 3D**, overall $p = 0.051$, Friedman). There was no significant difference in total or bacterial DNA recovery by dog in unspiked or spiked samples (**Fig. S7)**

We next assessed urine microbial diversity (16S) of unspiked urine samples by extraction method. Sequencing data from all samples extracted using NEBNext did not pass quality control steps[149] and, as such, were excluded from analysis. Urine microbial diversity varied significantly by extraction method (**Fig. 4, S8, Table S8,** Microbial richness $p = 0.0018$, Shannon Entropy $p = 0.0091$, Friedman). Specifically, urine samples extracted using Bacteremia and DNA Microbiome contained the greatest micobial richness (unique ASVs) and significantly greater microbial richness than samples extracted using Zymo HostZERO (**Fig. 4A, Table S8**, overall $p=0.0018$, pairwise $p=0.0041$). Samples extracted via Bacteremia, DNA Microbiome, or

propidium monoazide also exhibited the greatest microbial diversity (Shannon Entropy), all three showing significantly greater microbial diversity than samples extracted via Molzym MolYsis (**Fig 4B**, **Table S8,** pairwise p=0.025, 0.028, and 0.017, respectively).

Finally, we assessed urine microbial composition (16S) of unspiked urine samples by extraction method. Microbial composition (Bray-Curtis) differed significantly by dog but not by extraction method (**Fig 4C, S8,** Bray-Curtis, p=0.001, PERMANOVA). When composition was weighted by phylogeny (relatedness of microbes between samples; Unweighted UniFrac), composition differed significantly by both extraction method and by dog (**Fig 4D, S8,** Unweighted UniFrac, extraction method p=0.002, dog p=0.001, PERMANOVA). Urine samples extracted using Bacteremia, DNA Microbiome, and Molzym MolYsis exhibited more similar microbial composition as compared to samples extracted with propidium monoazide or Zymo HostZERO (**Fig. 4E, F, G, Table S9,** Bray-Curtis p=0.037, Jaccard p=0.034, Unweighted UniFrac p=0.0071, Friedman).

Host Depletion – Shotgun Metagenomics

We next assessed host depletion efficacy of each extraction method using shotgun metagenomic sequencing performed on urine samples spiked with host (canine) cells. Samples averaged 28.2 million paired-end reads per sample (range: 1399-80 million reads, SD: 16.7 million reads). There was no significant difference in the total number of reads obtained per sample by extraction method (**Fig. 5A** p=0.12, Friedman). However, the total number of *microbial* reads did vary significantly by extraction method (**Fig. 5B,** p=0.0039, Friedman), with DNA Microbiome, Molzym MolYsis, and Zymo HostZERO yielding a significantly greater number of microbial reads compared to Bacteremia, which includes no host depletion steps (all

pairwise p=0.01). The *proportion* of total microbial reads also varied significantly by extraction method with Molzym MolYsis and ZymoHostZERO yielding the greatest proportion of microbial reads (**Fig. 5C**, overall p<0.0001, pairwise p<0.02, Friedman). In terms of host reads, each method yielded the following (on average): Bacteremia, 82% host reads; DNA Microbiome, 78%; Molzym MolYsis, 29%; PMA, 81%; Zymo HostZERO, 30%. Finally, we quantified the abundance of contaminant reads by extraction method and found that DNA Microbiome samples contained the lowest abundance of contaminant reads (**Fig 5D,** overall p=0.014, Friedman), although contaminant data ranges from zero to 100 percent.

To determine whether efficacy in host depletion translated to improved capture of the urobiome, we employed MetaPhlaAn4 and SingleM - computational tools used for profiling microbial communities from marker genes found in metagenomes. Urine microbial diversity varied significantly by extraction method (**Fig. 6A, B,** MetaPhlAn, Observed Species p=0.011, Shannon entropy p=0.002, Friedman), with DNA Microbiome yielding the greatest number of observed microbial species and significantly more species than all other extraction methods (all pairwise p=0.015) except Molzym MolYsis. Urine microbial composition did not differ significantly by extraction method but did differ significantly by dog (**Fig. 6C, D,** MetaPhlAn4, By extraction method: Jaccard p=0.67, Bray-Curtis p=0.96; By dog: Jaccard p=0.001, Bray Curtis p=0.001, PERMANOVA), indicating that interindividual variation overwhelmed microbial community differences due to extraction method. SingleM largely recapitulated the MetaPhlAn results (**Fig. S9**).

We then assessed the viability of performing genome-resolved metagenomics on low biomass urine samples. To do this, we assembled metagenome-assembled genomes (MAGs) within each sample (Assembly metrics for each sample: **Fig. S10**). We generated a total of 26

unique MAGs: 11 were bacteria found in the ZymoBIOMICs Gut Microbiome Standard (**Table S3**), and five were derived from urine samples **(Fig. 7);** 10 were probable contaminants (**Table S5**). The five *E. coli* strains present in the standard assembled into a single MAG. The greatest number of urine-derived MAGs (n=4) were identified in DNA Microbiome samples while three or fewer MAGs were identified in all other extraction methods. The total number of MAGs did not vary by extraction method (**Fig. S11**, p=0.3, Friedman); although, fewer contaminant MAGs arose from DNA Microbiome samples as compared to other extraction methods (**Fig. S11,** overall p=0.018, Friedman, no pairwise significant).

Next, we compared the microbial taxonomic profiles generated by 16S rRNA sequencing, shotgun metagenomic sequencing (MetaPhlAn4), and genome-resolved metagenomics (MAGs) (**Fig. 7**). Each method is fundamentally different and employs different reference databases for taxonomy assignment. However, all five urine-derived MAGs also appeared in the top twenty most abundant taxa in the shotgun metagenomics and 16S datasets. Notably, *Arcanobacterium* is not present in the MetaPhlAn4 reference database, but was identified in the shotgun metagenomic data through the SingleM reference database (**Fig S9**). Additional top 20 genera common between the metagenomics and 16S datasets include: *Peptacetobacter/Peptoclostridium spp.* and *Blautia spp.*

Finally, we compared our capture of the ZymoBIOMICs Gut Microbiome Standard community across extraction, sequencing, and bioinformatic methods. The Standard contained 21 microbial taxa including 18 bacterial strains, 1 Archaea, and 2 microbial eukaroyotes at differing and biologically relevant abundances. Amongst the bacterial strains, there were 5 closesly related strains of *E. coli*. In the 16S dataset, we were able to detect a total of 12/21 taxa, all of which were present at ≥0.1% abundance in the Standard. Expectedly, we did not detect the

2 microbial eukaryotes (which do not encode a 16S rRNA gene). We were also unable to differentiate the 5 *E. coli* strains in the Standard as this is not feasible with amplicon sequencing. We also did not detect the 4 taxa found at ≤0.01% abundance in the Standard (*Methanobrevibacter smithii, Salmonella enterica, Enterococcus faeecalis, Clostrdium perfringens*). In the shotgun metagenomic data profiled using MetaPhlAn4, we detected a total of 14/21 taxa in the Standard including the 2 microbial eukaryotes. As with with 16S sequencing, we were able to detect all taxa present at ≥0.1% abundance in the Standard and not able to detect the 4 taxa found at ≤0.01% abundance in the Standard. MetaPhlan4 did not distinguish the 5 *E. coli* strains. We were further able to assemble a total of 11 MAGs from the shotgun metagenomic data. This included all taxa at ≥1.5% abundance, excluding the eukaryote *Candida albicans,* which was found at 1.5% abundance but for which we were not able to assemble a MAG. We assembled a single *E. coli* MAG (rather than the expected 5 unique *E. coli* strains). The threshold we employed for MAG dereplication (99% ANI) did not allow us to distinguish between the 5 *E. coli* strains; therefore, as with our 16S data, we only detected "one" *E. coli* taxon. A higher ANI (99.9%) and a tool other than dRep would be required for strain differentiation. We were not able to assemble a MAG for *M. smithii* which was present at 0.1% abundance and detected in 16S and shotgun metagenomic sequencing. Across methods (16S, shotgun metagenomics, MAGs), samples extracted using Bacteremia and DNA Microbiome most closely matched the expected microbial taxonomic composition of the Standard (**Fig S12**).

Functional Profiling of Urine Microbes

Relatively few studies have performed shotgun metagenomics in urine, and one, to our knowledge, has generated MAGs, which has limited our understanding of the functional

potential of the urobiome. In this study, as proof-of-concept, we mined the urine-derived MAGs for key functions. We first identified core metabolic pathways (e.g., glycolysis, citrate cycle) across all MAGs (**Fig. S13A**). Then we identified core pathways associated with carbohydrate, nitrogen, acid, and alcohol metabolism and searched for pathways that could indicate host adaptation, including urease degradation. (**Fig. S13B**).

Next, we looked for microbial metabolic pathways associated with environmental chemical metabolism. There are a number of environmental chemicals (e.g., arsenic, polycyclic aromatic hydrocarbons) that have been linked to urinary tract diseases like bladder cancer.[173] The kidney filters many of these toxicants out of the blood and into the urine. Therefore, it is important to understand if and how urine microbes metabolize these chemicals and how that could impact disease risk. As such, we mined the urine MAGs for pathways associated with polycyclic aromatic hydrocarbon (PAH) and long-chain alkane degradation. PAHs and long-chain alkanes are common environmental pollutants produced during the combustion process and found in vehicle exhaust and industrial output.[175–177] We did not identify genes (>80% gather cutoff) associated with PAH degradation but we did identify genes for long chain alkane utilization: *ladB* (91% of noise cutoff) in *Bacillus_A cereus* and *ladA* alpha (97% of trusted cutoff) in *Staphylococcus pseudintermedius*. Moreover, in *B. cereus*, we identified a full metabolic pathway starting with an alkanesulfonate monooxygenase (*ssuD*) that desulfonates organosulfonates to yield sulfite and an aldehyde (**Fig. 8A**). The presence of this pathway supports the possibility that *B. cereus* may be capable of utilizing a variety of hydrocarbons as potential carbon sources or electron donors. In *S. pseudintermedius*, we did not identify a complete metabolic pathway for long-chain alkane degradation but the presence of alcohol and aldehyde dehydrogenase protein families suggest that long chain alkanes activated by *ladA* may

be further oxidized by this organism (**Fig. 8B**). Taken together, these results suggest that urine-derived microbes can metabolize environmental chemicals, and that microbial metabolism merits further investigation in relation to urinary tract disease risk.

Discussion

Studies of the urobiome are poised to reveal key insights in urinary tract health and disease; however, validation of approaches to profiling the urine microbial community are urgently needed. Here, we tested urine sampling volume and DNA extraction methods with host depletion using urine from healthy dogs. We identified a minimum urine volume threshold for for 16S and shotgun metagenomic sequencing, and we report on best host depletion methods for obtaining representative and reproducible microbial profiles. Finally, we demonstrate that MAG assembly is feasible in low-microbial, high-host biomass urine samples, and that even in this limited study, we were able to gain novel functional insights into urine-associated microbes.

In relation to urine volume, we observed that greater urine volumes (≥ 3 mL) resulted in improved microbial community capture, increased read depth (although not significant), reduced stochastisticity / variability between samples, and reduced contaminant abundance (**Fig. 1, 2, S5, S6**). The largest urine volume tested in this study was 5 mL. It is possible that urine volumes >5 mL may further increase recovery of rare taxa, though previous work has suggested that urine sample volume does not necessarily influence total biomass or sequencing depth.[110] Notably, one recent review anectodally reccomended 30 mL-50 mL of catheter-collected urine for 16S profiling.[29] Our study focused on mid-stream free catch urine, which can include microbes from the urethra or skin in addition to the bladder, and would therefore contain a higher microbial biomass than catheter-collected samples,[55] which would be more representative of the bladder

microbiota alone. Thus, it is reasonable to suggest that greater urine volumes would be advisable for urobiome studies that utilize catheter-collected urine; although further study is warranted.

We next assessed the impact of DNA extraction methods with and without host depletion on multiple sample types (unspiked and host-spiked urine) and sequencing platforms (16S rRNA, shotgun metagenomics). In unspiked (low host biomass) urine, Bacteremia (no host depletion) and DNA Microbiome (host depletion) consistently yielded the greatest DNA concentrations and highest microbial diversity (16S) (**Fig. 3, 4**). Additionally, DNA Microbiome and Bacteremia-extracted samples were the most similar compositionally, and both of these methods most accurately captured the taxa and abudances of the ZymoBIOMICs Gut Microbiome Standard (**Fig. S12**). Notably, we were only able to reliably capture taxa that were found at ≥0.1% abundance in 16S and shotgun metagenomic data, and generate MAGs from taxa found at ≥1.5% abundance. As observed in other studies, interindividual variation (between dogs) generally outweighed differences due to extraction method.[22,110] However, when we employed phylogeny-aware metrics (Unweighted UniFrac), we saw significant differences in microbial composition by extraction method and by dog, suggesting that some host depletion methods bias microbial community profiles through preferential lysis of specific bacterial clades. Importantly, Bacteremia and DNA Microbiome have been identified as accurate and effective DNA extraction methods in other high-host, low-microbial biomass substrates (i.e., nasal swabs, vaginal swabs, urine, biopsies).[110,125,127,178]

In host-spiked (high host biomass) urine samples, DNA Microbiome, Zymo and Molzym yielded the greatest percent of microbial reads (22, 70, and 71% respectively) (shotgun metagenomics, **Fig. 5**). DNA Microbiome also recovered the greatest microbial diversity (MetaPhlAn4, **Fig. 6**). Notably, Bacteremia, with no host depletion, was not effective in

capturing the microbial community in high host biomass urine. As in our 16S analysis, interindividual variation (between dogs) overwhelmed differences by extraction method, though we did not assess the MetaPhlAn4-profiled communities according to phylogenetic differences in composition. The Zymo HostZERO kit did not perform as well in this study as it has in studies on other substrates (respiratory, intestinal biopsy), suggesting that certain host depletion strategies may be differentially effective by substrate.[144,178] Other technologies, not tested in this study, may also prove effective at microbial enrichment, including adaptive sequencing[179] and selective mechanical lysis.[136]

Important insights have been revealed via read-level analysis of shotgun-sequenced urobiota. For example, in one study, shifts in microbial functional potential were observed in longitudinally collected urine samples of individuals with and without urinary tract symptoms.[4] In another study, microbial virulence factor genes were linked to a distinct subset of individuals with urinary tract infections.[25] Whole-genome sequencing of cultured urine isolates has also revealed key insights: For example, genes enriched in strains of *E. faecalis* isolated from urine, were not found in gut or blood isolates, suggesting unique adaptations for the urinary tract niche.[94] MAG generation offers advantages over read-level analyses and culture as it uniquely provides critical, high-resolution information on specific microbes and their potential functions, without a dependence on culture.[180] Thus, we attempted *de novo* assembly of MAGs from our urine samples as proof-of-concept for genome-resolved metagenomics in urine. We assembled a total of five high quality (>90% complete, <10% contaminated), urine-derived microbial genomes: *B. cereus, S. pseudintermedius, S. canis,* and two unassigned *Arcanobacterium spp*. Notably, this study focused on mid-stream free catch urine samples which includes microbes from the bladder, urethra, and skin. Additionally, this study only included a small number of

healthy individuals and was not designed to capture the breadth of urobiome diversity. To our knowledge, this is among the first reports of MAGs assembled from urine. The MAGs we assembled have all been identified as members of the urobiota (or as potential uropathogens) in other studies.[16,22,181] One other group reports draft genomes of urine-associated microbes from voided human urine samples ranging in completeness from 17-98% (median 36.7%).[135] They assessed genomic functional potential in urolithiasis- or health-associated urine samples by annotating assemblies at the contig level, followed by mapping called genes to both the study's draft genomes and a reference database. This allowed for some of the genes that were differentially abundant in the metagenomes to be attributed to specific taxa but differs from our strategy of directly annotating MAGs for functional potential.

Although the overall number of MAGs we recovered was low, we note that DNA Microbiome yielded a greater number of urine-derived MAGs and generally fewer contaminant MAGs as compared to all other extraction methods. Importantly, the fact that we were able to assemble 11 MAGs from contaminants (i.e. microbial DNA present in reagents and identifiable in negative control samples) highlights the need for rigorous negative controls as well as thorough bioinformatic decontamination to avoid spurious results. Well validated tools such as *decontam*[57,107] as well as an awareness of common "kit-ome" taxa[56] are critical for microbiome studies of low biomass substrates.

After assembling MAGs, we went on to identify key functions in each MAG including core carbon and nitrogen metabolic pathways, urea metabolism, and environmental chemical degradation. As urea is a major component of urine,[182] the identification of full urea-degrading complexes (ureABCEFGD) in 2 MAGs (*B. cereus* and *S. pseudintermedius*) in 3 the 7 dogs via MAG-mapping (**Fig. S13**) indicates that urea is a likely nitrogen source for host-adapted

urobiota. As for environmental chemical degradation, there are well-established links between environmental chemical exposures and urinary tract diseases like bladder cancer.[64,173] In fact, a recent meta-analysis reported that bacteria associated with polycyclic aromatic hydrocarbon (PAH) degradation were found at increased abundances in the urine of individuals with bladder cancer.[16] While we did not find evidence for microbial PAH degradation in this limited study on healthy dogs, we did find evidence for long-chain alkane degradation in 2 urine-derived MAGs (*B. cereus* and *S. pseudintermedius*) found in 3 of the 7 dogs. Long-chain alkanes are common environmental pollutants that result from industrial combustion processes [175,183] and can be found in urine.[184,185] Our findings novelly demonstrate that 2 urine-derived MAGs may utilize long-chain alkanes as a substrate for metabolism. This proof-of-concept study highlights the importance of understanding if and how host-associated microbes may be metabolizing environmental chemicals, and the potential impacts of this metabolism on host health or in diseases like bladder cancer.[88–90]

Conclusions

Key takeaways from this study:

1. Urine sample volumes of ≥ 3 mL produced the most consistent urobiome profiles in dogs, which are a robust model for the human urobiome.
2. Microbial taxa found at ≥ 0.1% abundance were reliably detected via 16S and shotgun metagenomic sequencing, but MAG assembly was only feasible at greater abundances (≥ 1.5%), and strain differentiation in metagenomic data may require a higher ANI threshold than employed in this study (99% ANI was used in this study).

3. Generally, interindividual differences in urobiome profiles overwhelmed differences due to DNA extraction method.
4. In urine samples with low host biomass (unspiked), the QIAamp BiOstic Bacteremia kit (with no host depletion) yielded the greatest microbial DNA concentrations and highest microbial diversity (e.g. captured more / rarer urine taxa).
5. In urine samples with high host biomass (host-spiked), the QIAamp DNA Microbiome kit yielded the greatest microbial DNA concentrations, highest microbial diversity, and greatest number of identified metagenome-assembled genomes (MAGs), while effectively depleting host DNA.
6. MAG assembly is feasible but limited in urine samples. Maximizing urine volume to increase microbial reads would likely improve MAG recovery. Gene-based queries to assess functional potential of the urobiome are feasible with shotgun metagenomic data in the absence of MAG assembly; although, linking function (genes) to microbial species is more challenging with this approach.
7. Urine derived MAGs revealed evidence of urea and environmental chemical (long chain alkane) degradation, both of which are relevant for understanding how microbes live and adapt to the urine environment, as well as how they can potentially modulate environmental exposures in a way that could impact host health.

Urobiome research trails the study of other host-associated microbiomes,[142] and continued optimization of urobiome profiling is critical to enable the mechanistic and functional insights necessary for understanding how these microbes impact host health.

Author Contributions:

Conceptualization: VLH, SSJ, AR, JH, ZJL

Methodology (Sample Processing): ZJL, AS, CM

Investigation (Analysis): ZJL, AS, DV, GS, VLH

Visualization: ZJL, GJS

Funding acquisition: VLH, SSJ, AR, JH

Project administration: VLH

Supervision / Consultation: VLH, SSJ, AAZ

Writing – original draft: ZJL, GJS, VLH

Writing – review & editing: All

Acknowledgements:

We thank Dr. Christine Sun, Dr. Michael Sovic, Shashanka Murthy, and Dr. Veena Devi Ganeshan from the Infectious Diseases Institute Genomics and Microbiology Solutions Laboratory for support in optimizing our shotgun metagenomics sequencing approach. We also thank Dr. Tessa Cannon Wilde, Dr. Emily Coffey, Danielle Curcio, and Dr. Andreas Eleftheriou for feedback on the project and manuscript. We also thank members of the Hale laboratory who were involved in the original sample collection and processing: Ryan Mrofchak, Dr. Andrew McGlynn, and Rushil Madan. We further acknowledge Aaron Wiedemer, Nia Tran, Stephanie Majernik, Megan Hill, Afaf Abdelrahim, Dr. Marion Urvoy, Dr. Ahmed Zayed, and Dr. Matthew Sullivan, for data analysis guidance and support. Finally, we gratefully acknowledge the dogs and dog owners involved in this study.

Data Availability Statement

Raw 16S rRNA gene amplicon sequences are available on the Sequence Read Archive via BioProject accession number PRJNA1109516. Shotgun metagenomic sequences are likewise available via PRJNA1123238. Sequence processing scripts are available on Github.

Table S1. Demographics of dogs enrolled in this study. MN = male neutered, FS = female spayed. Experiment 1: Urine Volume Experiment; Experiment 2: Host Cell Removal - 16S; Experiment 3: Host Cell Removal - Shotgun Metagenomics.

Demographics of Dogs Enrolled in this Study

Dog ID	*Sex*	*Age (years)*	*Breed*	*Experiments*
ArB	MN	5	Australian Shepherd	1,2,3
FC	MN	1	Mixed Breed	1
HF	FS	6	Borzoi	1,2,3
IO	FS	3	Dalmatian	2,3
KH	FS	5	Mixed Breed	1,2,3
LM	FS	1.5	Dalmatian	2,3
MS	FS	4	Mixed Breed	1,2,3
SJ	FS	5	Mixed Breed	2,3

Table S2. **Putative microbial contaminants in Urine Volume Experiment (16S).** Contaminants were identified using the R package decontam and bioinformatically removed from the dataset.

Feature ID	Contaminant taxa
1c400e9dc5ae955f121df5ba8bac63a1	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Xanthomonadales; Family: Xanthomonadaceae; Genus: Stenotrophomonas; Species: uncultured_organism
df417ebb162691a49c1c89d12423aea2	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Enterobacteriaceae
f9a27fcf92575fe1239dbef43e15b33f	Phylum: Firmicutes; Class: Bacilli; Order: Staphylococcales; Family: Staphylococcaceae; Genus: Staphylococcus
6424b515572ba4888d5df4d898783859	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Burkholderiaceae; Genus: Burkholderia-Caballeronia-Paraburkholderia
40dd68437b49789dd596779921fe6e72	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Sphingomonadales; Family: Sphingomonadaceae; Genus: Sphingomonas
8e825741e31f18ea748eb119ece60cca	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Micrococcales; Family: Micrococcaceae; Genus: Micrococcus
6ccc82f2cdebff500a2835667202c23a	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium
9cc98843fbce7f87c738e44318be4a41	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Streptococcaceae; Genus: Streptococcus
40e0477d5940487d1bb061ae4008820f	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Acinetobacter
134e1c859ff430e9a868dde840931a76	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Rhizobiales; Family: Xanthobacteraceae; Genus: Bradyrhizobium
28699416c0922cac60ff5c0a536cc0d1	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Xanthomonadales; Family: Xanthomonadaceae; Genus: Stenotrophomonas
bbcd46ffae7d50084e7e4872d96f4635	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Burkholderiaceae; Genus: Ralstonia
cb8b7b696bdf8544a1fe364c97f20f7e	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
6200b82e17a325c07de4aff7539aa429	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Acinetobacter
0f8d1a9bf26751d9354af3481d68ff3b	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Enterobacteriaceae
7cab79099421ae87e9e9377e8c8e3865	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Carnobacteriaceae; Genus: Atopostipes; Species: uncultured_bacterium
3db347c9182e9467089c6e3ef7f0ee30	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Streptococcaceae; Genus: Streptococcus
fda6f6480bbcd0b84dc2e305071f386f	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium
6e2d8a4dc9fb4ea123accb37580db227	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Micrococcales; Family: Micrococcaceae; Genus: Kocuria
2821459d1c7f745877ba92aeac150396	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Aeromonadales; Family: Aeromonadaceae; Genus: Aeromonas
47ee71a277a3cfc3e9d45f014967e996	Phylum: Actinobacteriota; Class: Thermoleophilia; Order: Solirubrobacterales; Family: Solirubrobacteraceae; Genus: Solirubrobacter
c68f1f38b3805f6b2d15d9c1e98f9eba	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Rickettsiales; Family: Mitochondria; Genus: Mitochondria; Species: Fusarium_circinatum
63031312963d7d298d51d23cd2d73b48	Phylum: Bacteroidota; Class: Bacteroidia; Order: Chitinophagales; Family: Chitinophagaceae; Genus: uncultured
c86f2e387e9b317f9870c291a94eb8ee	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
10b841456150ed0fd1db7159bea33075	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
6ddfb30a4e08d68496956b75ec1d0fc8	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Propionibacteriales; Family: Propionibacteriaceae; Genus: Cutibacterium
52bdaafae05f38d1f2fed2ede1bd9f48	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Sphingomonadales; Family: Sphingomonadaceae; Genus: Sphingopyxis
bfdb053fde342059816bd87ae6bb0b06	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae
ed0f768bfe644821ddd8c793eb0abbf6	Phylum: Firmicutes; Class: Clostridia; Order: Clostridiales; Family: Clostridiaceae; Genus: Clostridium_sensu_stricto_1; Species: uncultured_Firmicutes
ad7a2fa1cf79ee2e2023b0441a2cd776	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae

0a31ab523c43d207ae43623ee8505cc5	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Micrococcales; Family: Promicromonosporaceae; Genus: Cellulosimicrobium
0fdbbedd6cef715fe7af5822ac39b371	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium; Species: Corynebacterium_amycolatum
af754e7ad6102e759210305b677c0a34	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae; Genus: Schlegelella; Species: uncultured_bacterium
5e47104de80af59694f2b6672be66b63	Phylum: Acidobacteriota; Class: Acidobacteriae; Order: Subgroup_2; Family: Subgroup_2; Genus: Subgroup_2; Species: uncultured_bacterium
1228e61c5252626be981be89429855ed	Phylum: Chloroflexi; Class: AD3; Order: AD3; Family: AD3; Genus: AD3; Species: uncultured_bacterium
b5a5ff4ff80d50e34212a89a08f0ba7c	Domain: Bacteria
4875adab4a0fb1d62106f2e927583028	Phylum: Firmicutes; Class: Clostridia; Order: Peptostreptococcales-Tissierellales; Family: Anaerovoracaceae; Genus: Family_XIII_AD3011_group
dfdc712087293952b718d5e554968ed5	Phylum: Firmicutes; Class: Bacilli; Order: Erysipelotrichales; Family: Erysipelotrichaceae; Genus: Dubosiella; Species: uncultured_bacterium

Table S3. Taxa present in the ZymoBIOMICS Gut Microbiome Standard. Bold indicates that we were able to assemble a MAG for that taxon. 5 *E. coli* strains are in the standard, but our methods assembled them into a single MAG.

Taxon	Expected Abundance (%)
Bacteroides_fragilis	14
Faecalibacterium_prausnitzii	14
Roseburia_hominis	14
Veillonella_rogosae	14
Bifidobacterium_adolescentis	6
Fusobacterium_nucleatum	6
Limosilactobacillus_fermentum	6
Prevotella_corporis	6
Akkermansia_muciniphila	1.5
Candida_albicans (microbial eukaryote)	1.5
Clostridioides_difficile	1.5
Saccharomyces_cerevisiae (microbial eukaryote)	1.4
Methanobrevibacter_smithii (Archaea)	0.1
Salmonella_enterica	0.01
Enterococcus_faecalis	0.001
Clostridium_perfringens	0.0001
Escherichia_coli (JM109)	2.8
Escherichia_coli (B-3008)	2.8
Escherichia_coli (B-2207)	2.8
Escherichia_coli (B-766)	2.8
Escherichia_coli (B-1109)	2.8

Table S4. Putative microbial contaminants in Host Cell Removal – 16S experiment. Contaminants were identified using the R package decontam and bioinformatically removed from the dataset.

Feature ID	Contaminant Taxa
17e5d4820ffb665283773524f5d33c47	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Neisseriaceae; Genus: uncultured; Species: uncultured_rumen
29dcc2501cb455324526f383a42716d6	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Acinetobacter
5ce05abbdd174f9ba4373413e4946749	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Lactobacillaceae; Genus: Lactobacillus; Species: Lactobacillus_fermentum
ccc22cbde967b46c238aa946097cf3bb	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
f65b0d716de2efe63a6552a7971aa5d2	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Acinetobacter
622f6bc41a295acae9e913220085723a	Phylum: Firmicutes; Class: Bacilli; Order: Staphylococcales; Family: Staphylococcaceae; Genus: Staphylococcus
207b76aa7c4003ee7d1dab5ad6cd27b0	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Aeromonadales; Family: Aeromonadaceae; Genus: Aeromonas
c02b86c7007265d5042b2c6acc0a414c	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Xanthomonadales; Family: Xanthomonadaceae; Genus: Stenotrophomonas; Species: uncultured_organism
6c5c6a22330a148d44274346605dd3a5	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Yersiniaceae
b3822a12e68e84e0d10cb53ccedcfae3	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Enterobacteriaceae
2d4a85134826aec743e9f9305a4d6910	Phylum: Bacteroidota; Class: Bacteroidia; Order: Flavobacteriales; Family: Weeksellaceae; Genus: Empedobacter
e8ac76b63bdef3a480011a9e653fb8e4	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Micrococcales; Family: Micrococcaceae; Genus: Glutamicibacter
28ff515b3e10dfeb248ea2c7ffc3e304	Phylum: Bacteroidota; Class: Bacteroidia; Order: Flavobacteriales; Family: Weeksellaceae; Genus: Empedobacter; Species: Wautersiella_falsenii
37550dffd56e370f65376f114b0030ce	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Xanthomonadales; Family: Xanthomonadaceae; Genus: Stenotrophomonas
586e552517834804774f5130ef9dfdce	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Rhizobiales; Family: Beijerinckiaceae; Genus: Methylobacterium-Methylorubrum
444365067d3dd46681d9bb2cbc4e23cb	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
a6780dab721cda3bec0d4caa07af1687	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Rhizobiales; Family: Rhizobiaceae; Genus: Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium
eef0a22ea2ba0f3c3f1d5cee3499e583	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Streptococcaceae; Genus: Streptococcus
de9c3c38a0a1aa5fac7d0fe8b73c42d8	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Caulobacterales; Family: Caulobacteraceae; Genus: Brevundimonas
4c5fb2a0a2fa32362a103ed4741de4fa	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Morganellaceae; Genus: Providencia
a3aa427840fdac7d5c02cc7b8b447c6f	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Micrococcales; Family: Micrococcaceae; Genus: Micrococcus
6481d93cfe4f3bf9e3872476b47b8eea	Phylum: Bacteroidota; Class: Bacteroidia; Order: Flavobacteriales; Family: Weeksellaceae; Genus: Chryseobacterium
9859ab1943a57665f998be2fe395cd01	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Enhydrobacter
4b95e337657bf8abde13284efb6dd70a	Phylum: Bacteroidota; Class: Bacteroidia; Order: Sphingobacteriales; Family: Sphingobacteriaceae; Genus: Sphingobacterium
c42580510207ac58d658be982db08efa	Phylum: Firmicutes; Class: Bacilli; Order: Bacillales; Family: Planococcaceae; Genus: Lysinibacillus
3bee722e683391eb0cfa1e54850459b3	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae; Genus: Aquabacterium
46de671ffb57bc8c131a7619ab2f1f83	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
4336d36fe7a1fbfd6401d14b61703af7	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Moraxellaceae; Genus: Acinetobacter
d677e3fb1354dadf011e07609e5d704e	Phylum: Bacteroidota; Class: Bacteroidia; Order: Cytophagales; Family: Spirosomaceae; Genus: Flectobacillus

7430c930ced8e580bc67f426d2d15d1d	Phylum: Cyanobacteria; Class: Cyanobacteriia; Order: Chloroplast; Family: Chloroplast; Genus: Chloroplast
acac322a57b2e28c36ecbec06b12ab56	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Sphingomonadales; Family: Sphingomonadaceae; Genus: Sphingobium
ee2d53f46d88723b6e8dfbe63af874b0	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
d19d6c436c77e541c473f2f8b63af36a	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Leuconostocaceae; Genus: Leuconostoc
619e7874a6e17567151a63395eaf20a4	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
0cdba9ac4d15a03bff3c0b52dbdd60f8	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium
76c611ccff665bc9f26eb018ada4fe53	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
e18be3beff7df2c2d6f3dec3309be8d9	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas; Species: Pseudomonas_caeni
ac4879c1409c4ab336b61171d37c9e2e	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
293ddf5cc8aa885cae617751ea7d93bc	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Enterobacteriaceae
ce62d3d55490378ba7015b76aa3d8343	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Frankiales; Family: Geodermatophilaceae; Genus: Blastococcus
61d6d8e9852b837b524821060be0e68c	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Burkholderiaceae; Genus: Ralstonia
0ab346c22567d9bb8fe93678a793c399	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Sphingomonadales; Family: Sphingomonadaceae; Genus: Sphingomonas
e5493a4b94599cbf91a4661be985eeb7	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Acetobacterales; Family: Acetobacteraceae; Genus: Acetobacter
220a7ddaa9be42548fc3efa0e12ed50c	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Burkholderiaceae; Genus: Burkholderia-Caballeronia-Paraburkholderia
6fa3e8a661f4a107aedb46a42e2656e5	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Acetobacterales; Family: Acetobacteraceae; Genus: Acetobacter
c215a76fac3a604bfe62847f46061693	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Rhizobiales; Family: Beijerinckiaceae; Genus: Methylobacterium-Methylorubrum
86da2134ed434b31365a48561727a1bf	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae
57d15a1e1b2e65d89005243403cf5dec	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Aerococcaceae; Genus: Aerococcus
f6384eb225e15674b7bd55958649ce75	Phylum: Firmicutes; Class: Bacilli; Order: Bacillales; Family: Planococcaceae
dc0397419585d848825320e0ea9b86c4	Phylum: Firmicutes; Class: Clostridia; Order: Oscillospirales; Family: [Eubacterium]_coprostanoligenes_group; Genus: [Eubacterium]_coprostanoligenes_group
afe3a919262ff8ebdb4a7f66df7a8aeb	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Streptococcaceae; Genus: Streptococcus; Species: Streptococcus_salivarius
ba628f9a7dc4737274badc88a5968b49	Phylum: Bacteroidota; Class: Bacteroidia; Order: Flavobacteriales; Family: Weeksellaceae; Genus: Empedobacter
1eeb945228d12251e95d102a2cbe8d9d	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Enterococcaceae; Genus: Enterococcus; Species: Enterococcus_cecorum
94a0f635134e2c318274dd0fdf1bc385	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
9a66b182a468e6a3e12bb3109f2cbdca	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Sphingomonadales; Family: Sphingomonadaceae; Genus: Qipengyuania; Species: uncultured_bacterium
ad1b73be285f3c10a5568d1a2c927dfc	Domain: Archaea; Phylum: Crenarchaeota; Class: Nitrososphaeria; Order: Nitrososphaerales; Family: Nitrososphaeraceae; Genus: Nitrososphaeraceae
6ffb110ff6598392f9c8baed2fbbf2d0	Phylum: Bacteroidota; Class: Bacteroidia; Order: Sphingobacteriales; Family: Sphingobacteriaceae; Genus: Sphingobacterium; Species: Sphingobacterium_psychroaquaticum
da5af0e5a7c0ea0603c6ac70888877d6	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Pseudomonadales; Family: Pseudomonadaceae; Genus: Pseudomonas
83711d11b4caed655cb19233c471604a	Phylum: Proteobacteria; Class: Alphaproteobacteria; Order: Acetobacterales; Family: Acetobacteraceae; Genus: Acetobacter
4dc50e8dc3e44a87d99578358fd7a272	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Oxalobacteraceae; Genus: Herbaspirillum
aa0b6139a4b342def4050014e8f23047	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Lactobacillaceae; Genus: Lactobacillus; Species: Lactobacillus_salivarius
f9e9654f45fb5ae1d7259333ea3eb902	Phylum: Firmicutes; Class: Bacilli; Order: Bacillales; Family: Bacillaceae; Genus: Bacillus; Species: Bacillus_thermoamylovorans

eb6c0edfe9b013aaa4f0a6779a66f37c	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium
d978d8f907b58e5499c3f52d23eedc8e	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Carnobacteriaceae; Genus: Atopostipes; Species: uncultured_bacterium
cabc84505d63e214d45d53246018081d	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacterales; Family: Enterobacteriaceae
8e9155bdf3e7d9a42b8ce76849122bd2	Phylum: Firmicutes; Class: Clostridia; Order: Peptostreptococcales-Tissierellales; Family: Peptostreptococcales-Tissierellales; Genus: Fenollaria; Species: uncultured_bacterium
febba0f8d4a2e1c9b8c31373cc35f418	Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Burkholderiales; Family: Comamonadaceae
e3beb290843cb1c4220cb340329517b9	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Corynebacteriales; Family: Corynebacteriaceae; Genus: Corynebacterium; Species: Corynebacterium_kroppenstedtii
561944bb25b498c8a8e1ac03d213629b	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Bifidobacteriales; Family: Bifidobacteriaceae; Genus: Bifidobacterium
6326f1d853ad013f5d35af902310c798	Phylum: Actinobacteriota; Class: Actinobacteria; Order: Propionibacteriales; Family: Propionibacteriaceae; Genus: Cutibacterium
a3ddb44d01a2bbd4182098679a2a42a3	Phylum: Firmicutes; Class: Bacilli; Order: Lactobacillales; Family: Streptococcaceae; Genus: Streptococcus; Species: Streptococcus_agalactiae

Table S5. Putative microbial contaminants in shotgun metagenomic data. Contaminants were identified manually or by the R package *decontam* and bioinformatically removed from MetaPhlAn, MAG, and SingleM abundance tables.

Putative Contaminant Taxa in Metagenomic Data	
Removed from MetaPhlAn abundance table	
s__Corynebacterium_accolens	s__Sphingobium_yanoikuyae
s__Corynebacterium_aurimucosum	s__Sphingomonas_ursincola
s__Corynebacterium_otitidis	s__Cupriavidus_gilardii
s__Gordonia_paraffinivorans	s__Paraburkholderia_fungorum
s__Micrococcus_luteus	s__Aquabacterium_parvum
s__Cutibacterium_acnes	s__Tepidimonas_fonticaldi
s__GGB2722_SGB3663	s__Acidovorax_temperans
s__GGB43920_SGB61282	s__Herminiimonas_contaminans
s__Hydrotalea_flava	s__Acinetobacter_johnsonii
s__Cloacibacterium_caeni	s__Acinetobacter_junii
s__Staphylococcus_epidermidis	s__Pseudomonas_aeruginosa
s__Streptococcus_thermophilus	s__Pseudomonas_putida
s__Brevundimonas_nasdae	s__Pseudomonas_stutzeri
s__Caulobacter_sp_3R27C2_B	s__Malassezia_restricta
s__Afipia_birgiae	s__Bradyrhizobium_diazoefficiens
s__Bradyrhizobium_sp_CCH5_F6	s__Bradyrhizobium_viridifuturi
s__Methylobacterium_sp_B4	s__Sphingomonas_echinoides
s__Methylorubrum_extorquens	s__Acinetobacter_ursingii
s__Reyranella_soli	s__Herbaspirillum_huttiense
s__Agrobacterium_tomkonis	s__Xanthomonas_massiliensis
Removed from the MAG abundance table	
s__Bradyrhizobium_sp003020075	g__U87765
s__Cutibacterium_acnes	g__DSPG01
s__Sphingomonas_echinoides	s__Gemmatimonas_sp016741895
s__Bradyrhizobium_diazoefficiens_A	g__Qipengyuania
g__Lacibacter	s__Cupriavidus_gilardii
Removed from the SingleM abundance table	
f__Lactobacillaceae	g__Hydrotalea

s__Bifidobacterium_adolescentis	s__Methylobacterium_extorquens
d__Archaea	s__Reyranella_sp009377525
c__Gammaproteobacteria	c__Bacilli
c__Alphaproteobacteria	o__Lactobacillales
o__Burkholderiales	f__Mycobacteriaceae
o__Rhizobiales	s__Cutibacterium_acnes
f__Burkholderiaceae	s__Lawsonella_clevelandensis_A
f__Sphingomonadaceae	g__Corynebacterium
f__Rhodobacteraceae	o__Pseudomonadales
f__Xanthobacteraceae	f__Pseudomonadaceae
f__Chitinophagaceae	g__Pseudomonas_E
g__Ga0077523	p__Chloroflexota
g__Erythrobacter	c__Thermoleophilia
g__Qipengyuania	f__Solirubrobacteraceae
g__Novosphingobium	g__Cutibacterium
g__Tsuneonella	s__Pseudomonas_A_stutzeri
g__Hyphomicrobium_A	g__Streptococcus
g__Methylobacterium	s__Sphingomonas_echinoides
g__Reyranella	g__SCUB01
g__Lacibacter	f__Rhizobiaceae
s__Cupriavidus_gilardii	f__Acetobacteraceae
s__Ga0077523_sp001464695	f__Micrococcaceae
s__Blastomonas_ursincola	s__Micrococcus_luteus
s__Afipia_birgiae	g__Nitrobacter
s__Hydrotalea_flava	g__Rhodopseudomonas
s__Gemmatimonas_sp016741895	g__Afipia
c__Bacteroidia	g__Agrobacterium
c__Actinomycetia	f__Acetobacteraceae
g__Blastomonas	f__Micrococcaceae
g__Pelagerythrobacter	s__Micrococcus_luteus

Table S6. Pairwise comparisons of contaminant abundance by urine sample volume (16S). FDR-corrected (q=0.05) multiple comparisons (Friedman) of microbial contaminant abundance across different urine sample volumes. The overall test was significant (p=0.026) with no significant pairwise comparisons.

Pairwise Comparisons of Contaminant Abundance by Urine Sample Volume	q-value
0.1mL vs. 0.2mL	0.5448
0.1mL vs. 0.5mL	0.8926
0.1mL vs. 1.0mL	0.2002
0.1mL vs. 3.0mL	0.0642
0.1mL vs. 5.0mL	0.1543
0.2mL vs. 0.5mL	0.4936
0.2mL vs. 1.0mL	0.4936
0.2mL vs. 3.0mL	0.1543
0.2mL vs. 5.0mL	0.4936
0.5mL vs. 1.0mL	0.1543
0.5mL vs. 3.0mL	0.0642
0.5mL vs. 5.0mL	0.1543
1.0mL vs. 3.0mL	0.4936
1.0mL vs. 5.0mL	0.8926
3.0mL vs. 5.0mL	0.5448

Table S7. Pairwise comparisons of microbial richness by urine sample volume (16S). FDR-corrected (q=0.05) multiple comparisons (Friedman) of microbial richness (unique ASVs) across different urine sample volumes. The overall test was significant (p=0.011). Significant pairwise comparisons are denoted in bold.

Pairwise Comparisons of Microbial Richness by Urine Sample Volume	q-value
0.1mL vs. 0.2mL	0.6349
0.1mL vs. 0.5mL	0.8400
0.1mL vs. 1.0mL	0.1828
0.1mL vs. 3.0mL	0.1481
0.1mL vs. 5.0mL	**0.0157**
0.2mL vs. 0.5mL	0.6349
0.2mL vs. 1.0mL	0.2942
0.2mL vs. 3.0mL	0.1828
0.2mL vs. 5.0mL	**0.0342**
0.5mL vs. 1.0mL	0.1828
0.5mL vs. 3.0mL	0.1481
0.5mL vs. 5.0mL	**0.0157**
1.0mL vs. 3.0mL	0.6349
1.0mL vs. 5.0mL	0.1828
3.0mL vs. 5.0mL	0.2942

Table S8. Pairwise comparisons of microbial diversity by extraction method (16S). FDR-corrected (q=0.05) pairwise comparisons (Friedman) of microbial richness (Unique ASVs) (overall p=0.0025) and Shannon entropy (overall p=0.0075) by extraction method. Significant pairwise comparisons are denoted in bold.

Extraction Methods	Unique ASVs q-value	Shannon entropy q-value
Bacteremia vs. DNA Microbiome	0.8400	0.6177
Bacteremia vs. Molzym MolYsis	0.0714	**0.0287**
Bacteremia vs. Propidium Monoazide	0.1851	0.5713
Bacteremia vs. Zymo HostZERO	**0.0030**	0.1528
DNA Microbiome vs. Molzym MolYsis	0.0714	0.0503
DNA Microbiome vs. Propidium Monoazide	0.1851	0.4179
DNA Microbiome vs. Zymo HostZERO	**0.0030**	0.2468
Molzym MolYsis vs. Propidium Monoazide	0.4657	**0.0111**
Molzym MolYsis vs. Zymo HostZERO	0.1851	0.3726
Propidium Monoazide vs. Zymo HostZERO	0.0714	0.0588

Table S9. Pairwise comparisons of diversity metric distances to Bacteremia-extracted samples (16S). FDR-corrected (Q=0.05) pairwise comparisons (Friedman) of distances to Bacteremia-extracted samples by extraction method. The overall tests were significant (p=0.037, 0.0342, 0.007, respectively). Significant pairwise comparisons are denoted in bold.

Pairwise Comparison of Microbial Composition by Extraction Method	Bray Curtis q value	Jaccard q value	Unweighted Unifrac q value
DNA Microbiome vs. Molzym MolYsis	0.1799	0.7128	0.2249
DNA Microbiome vs. Propoidium Monoazide	0.7000	0.7016	0.5852
DNA Microbiome vs. Zymo HostZERO	**0.0273**	**0.0374**	**0.0079**
Molzym MolYsis vs. Propoidium Monoazide	0.1799	0.7315	0.2525
Molzym MolYsis vs. Zymo HostZERO	0.1799	0.0598	0.0874
Propoidium Monoazide vs. Zymo HostZERO	**0.0273**	0.0673	**0.0079**

Figure i. Loss of sparse microbes at threshold sample volumes of urine. Image generated using Biorender.

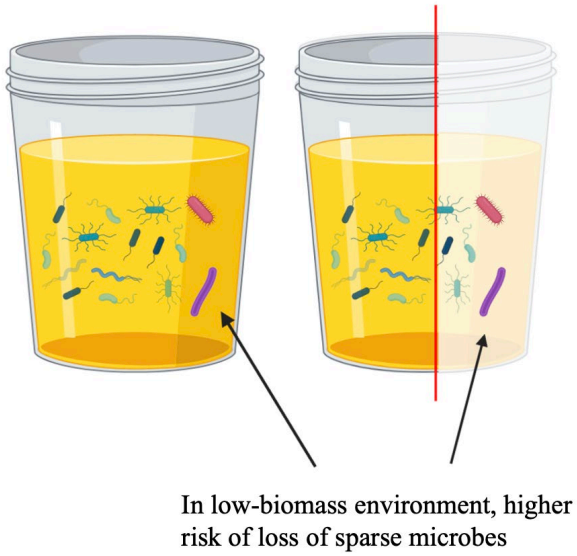


Figure 1. Urine sample volume influences contaminant abundance and microbial diversity (16S). A) The abundance of contaminants (contaminating microbial sequences) decreased significantly as sample volume increased (overall p=0.026, Friedman, no pairwise comparisons were significant, **Table S6**). B) Microbial richness, or the number of unique amplicon sequence variants (ASVs), increased significantly with increased sample volume (p=0.015, Friedman) and 5.0mL samples had significantly greater numbers of unique ASVs compared to 0.5mL (p=0.031), 0.2mL (p=0.031), and 0.1mL samples (p=0.048), (multiple comparisons were FDR-corrected at 0.05, **Table S7**). C) Sequencing depth (reads) was increased at greater urine sample volumes although this difference was not significant (p=0.075, Friedman). Box and whisker plots show the median, IQR, and min/max.

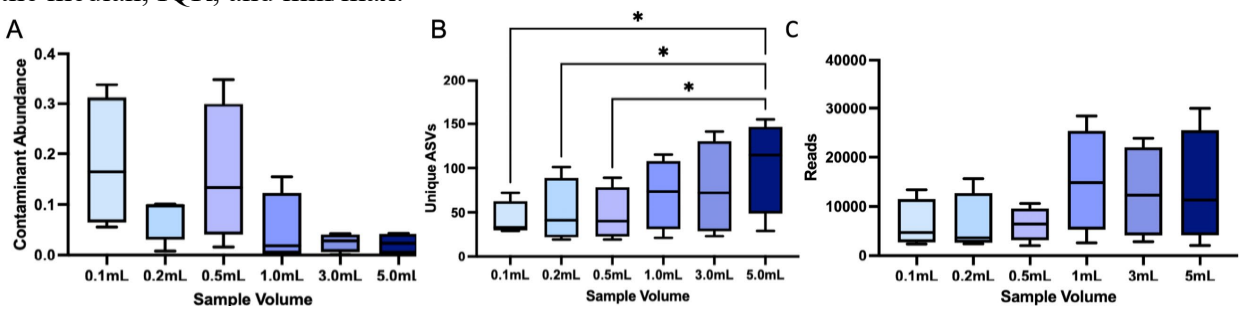


Figure 2. Urine sample volume and microbial composition (16S). A) Microbial composition (Bray-Curtis) of urine samples differed significantly by dog but not by sample volume (PERMANOVA: by dog p = 0.001; by sample volume p = 0.98). B) Representative Bray-Curtis plot of a single dog's (Dog = MS) urine samples. C) High volume (≥ 3 mL) samples were significantly less variable (shorter distance to centroid) than low volume (≤ 1 mL) samples (Bray Curtis, p = 0.0017, PERMDISP).

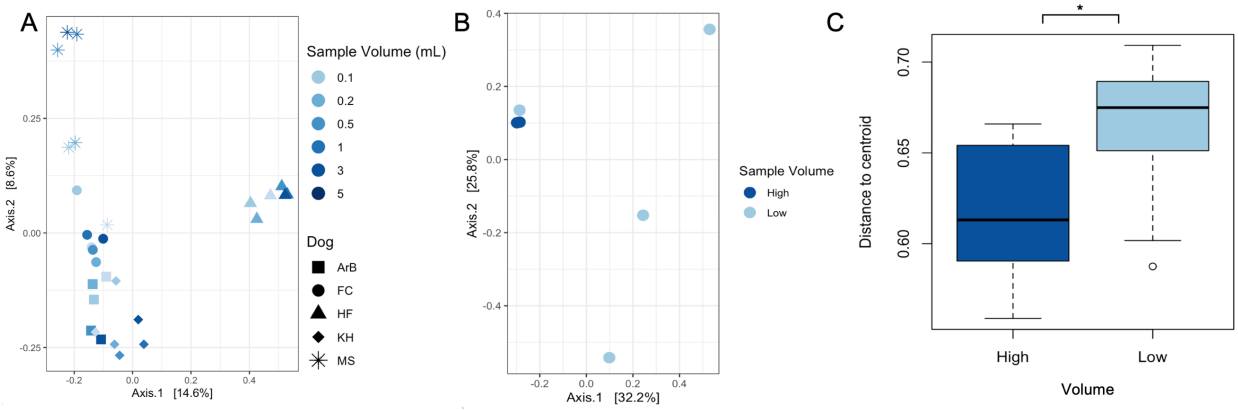


Figure 3. Total and bacterial DNA recovery differed by extraction method. A) Total DNA concentrations (ng/ul, Qubit fluorometry) did not differ by extraction method (p=0.62, Friedman). B) Bacterial DNA concentrations (qPCR) differed significantly by extraction method (overall p=0.014, Friedman); although no pairwise comparisons were significant. C) Total DNA concentrations recovered from urine samples spiked with canine (CTAC) cells varied significantly by extraction method (p<0.0001, Friedman). Pairwise comparisons are indicated by letter above each bar. Bars with differing letters were significantly different (p<0.05, FDR 0.05). D) Bacterial DNA concentration from spiked urine samples marginally differed by extraction method (p=0.051, Friedman). Bars represent the mean with standard error. Pipettor icon in C) and D) indicates that all samples shown in these graphs were spiked with canine thyroid adenocarcinoma cells (CTAC).

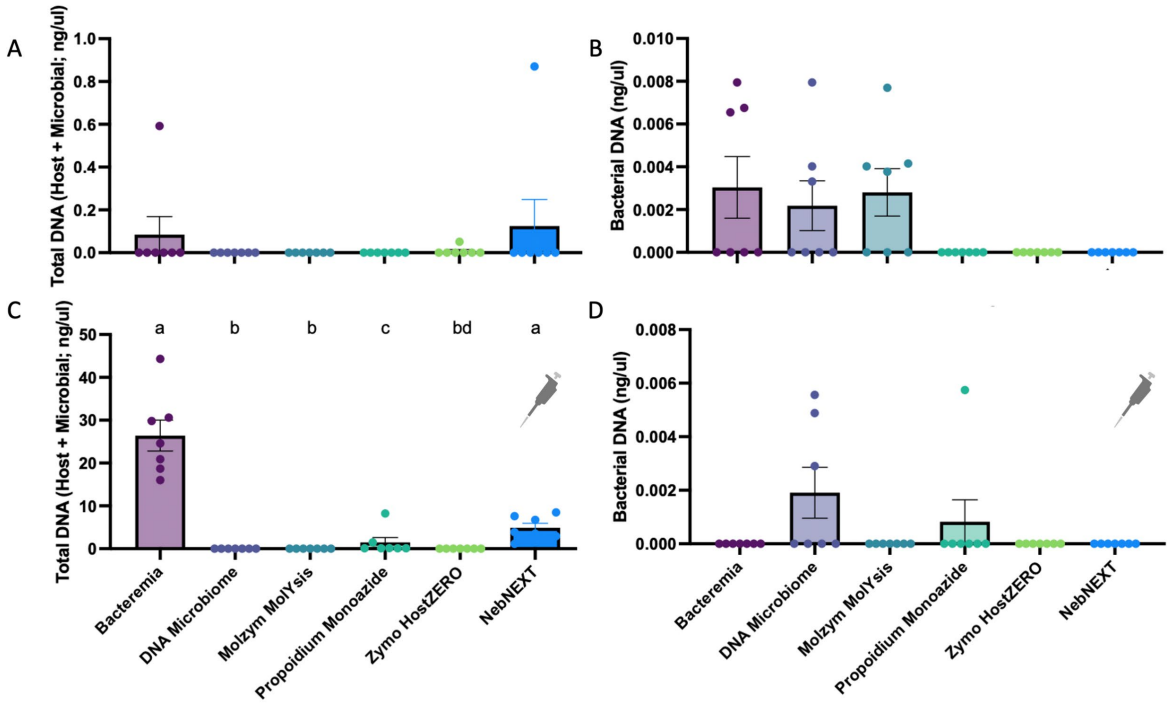


Figure 4. Microbial diversity and composition by extraction method (16S). A) Microbial richness, or number of unique ASVs, and B) Microbial diversity (Shannon entropy) differed significantly by extraction method (Richness p=0.0018, Shannon p=0.0091, Friedman, multiple comparisons with FDR at 0.05, **Table S8**). Whiskers represent minimum, maximum, and median. *p<0.05. C) Microbial composition (Bray-Curtis) differed significantly by dog (PERMANOVA p=0.001), but not extraction method (PERMANOVA p=0.92) D) When microbial composition was weighted by phylogeny (Unweighted UniFrac), composition differed significantly by both extraction method (PERMANOVA p=0.002), and by dog (PERMANOVA p=0.001). E) Bray-Curtis, F) Jaccard, and G) Unweighted UniFrac distances from Bacteremia-extracted samples to samples of the same dog extracted via other extraction methods. E) Bray-Curtis distances differed significantly (p=0.034, Friedman) with DNA Microbiome samples being the most similar shortest distance) to Bacteremia-extracted samples. F) Jaccard distances to Bacteremia-extracted samples differed significantly (p=0.0342) following the same pattern. G) Unweighted UniFrac distances to Bacteremia-extracted samples differed significantly (p=0.0071), again following the same pattern. Pairwise comparison p-values are outlined in **Table S9.** *p<0.05. Bar represents median.

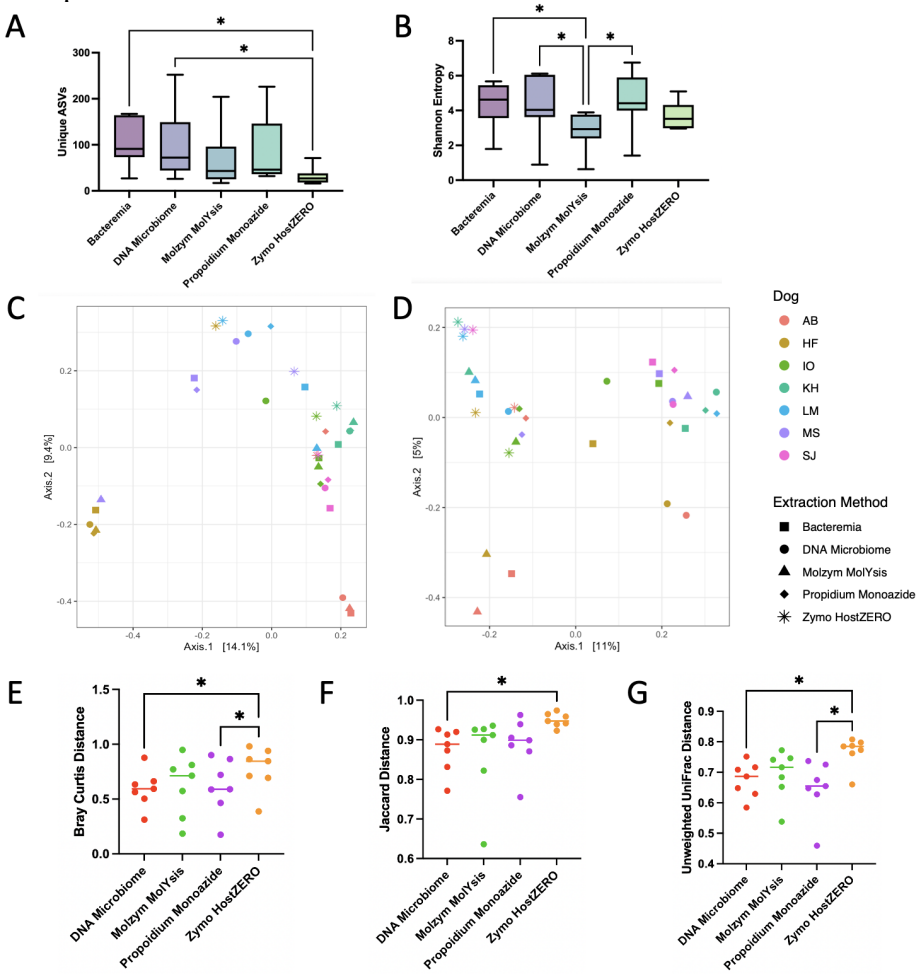


Figure 5. Extraction method impacted host and microbial read abundances (Shotgun metagenomics). A) Total sequencing reads did not vary by extraction method (p=0.12, Friedman). B) However, *microbial* reads did vary significantly by extraction method (p=0.0039, Friedman), with DNA Microbiome, Molzym MolYsis, and Zymo HostZERO exhibiting a greater number of microbial reads compared to Bacteremia (all pairwise p=0.01). C) The *proportion* of microbial reads also varied significantly by extraction method with Molzym MolYsis and Zymo HostZERO yielding the greatest proportion of microbial reads (overall p<0.0001, all pairwise p<0.02, Friedman). D) The abundance of contaminant reads also differed significantly by extraction method (overall Friedman p=0.014) and was lowest in DNA Microbiome (DNA microbiome vs. Zymo HostZERO pairwise p=0.01). (See **Table S5** for a list of contaminants).

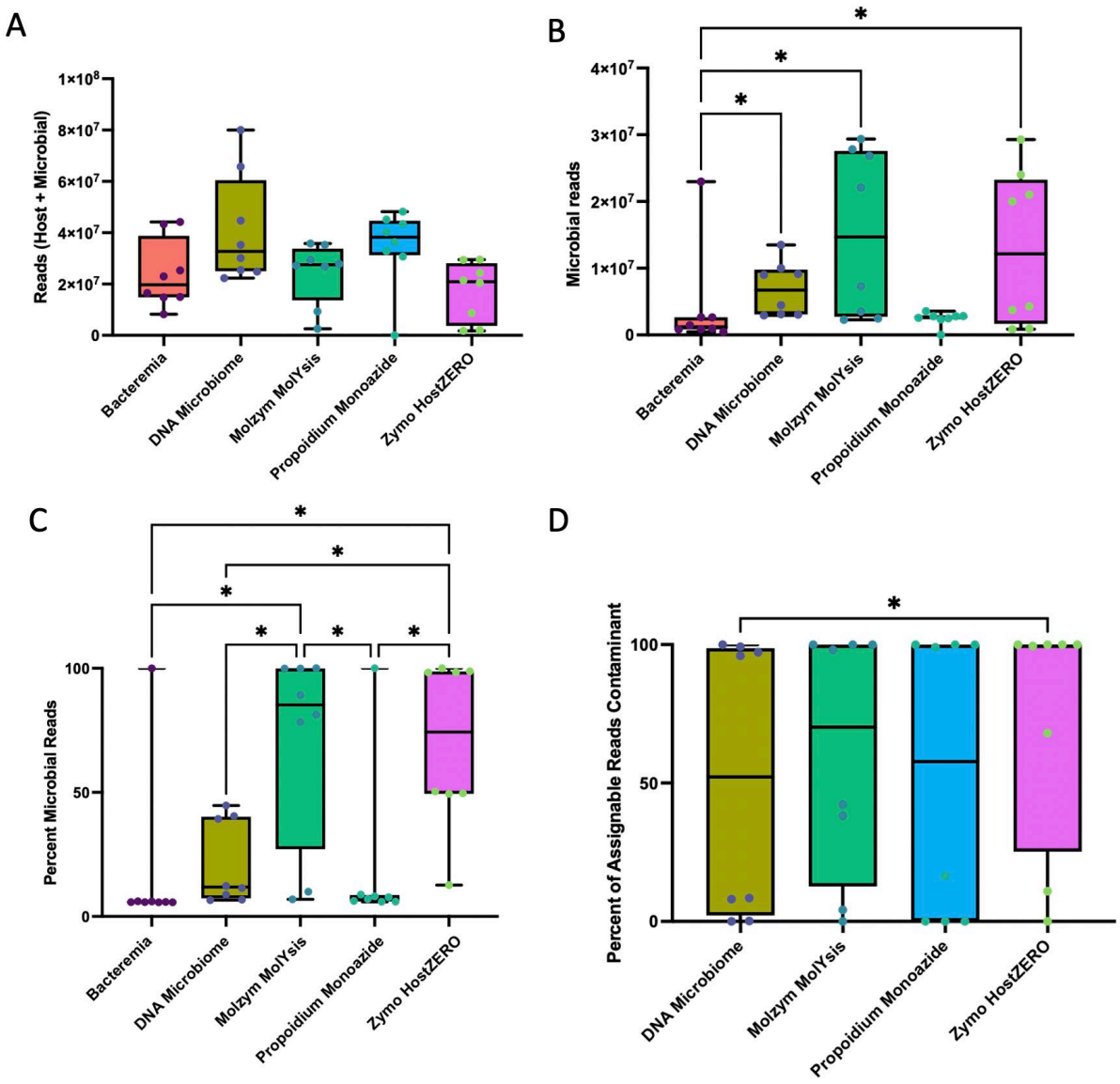


Figure 6. Extraction method and microbial diversity and composition (Shotgun metagenomics). Microbial diversity as measured by A) Observed Species (richness) and B) Shannon Entropy varied significantly by extraction method (Observed Species p=0.011, Shannon entropy, p=0.002, Friedman) with DNA Microbiome yielding significantly greater microbial diversity than other extraction methods (Observed Species DNA Microbiome vs. Bacteremia pairwise p=0.014, Shannon DNA Microbiome vs. all other methods (except Molzym MolYsis) pairwise p=0.014). Microbial species were identified via MetaPhlAn4. C) Microbial composition as measured by Jaccard or D) Bray-Curtis differed significantly by dog (Jaccard p=0.001, Bray-Curtis p=0.001, PERMANOVA), but not extraction method (Jaccard p=0.67, Bray-Curtis p=0.96, PERMANOVA). Urine samples extracted with Bacteremia contained little microbial DNA and did not produce reads that were assignable to a taxa by MetaPhlAn4. As such, Bacteremia samples were excluded from C) and D) and beta diversity testing.

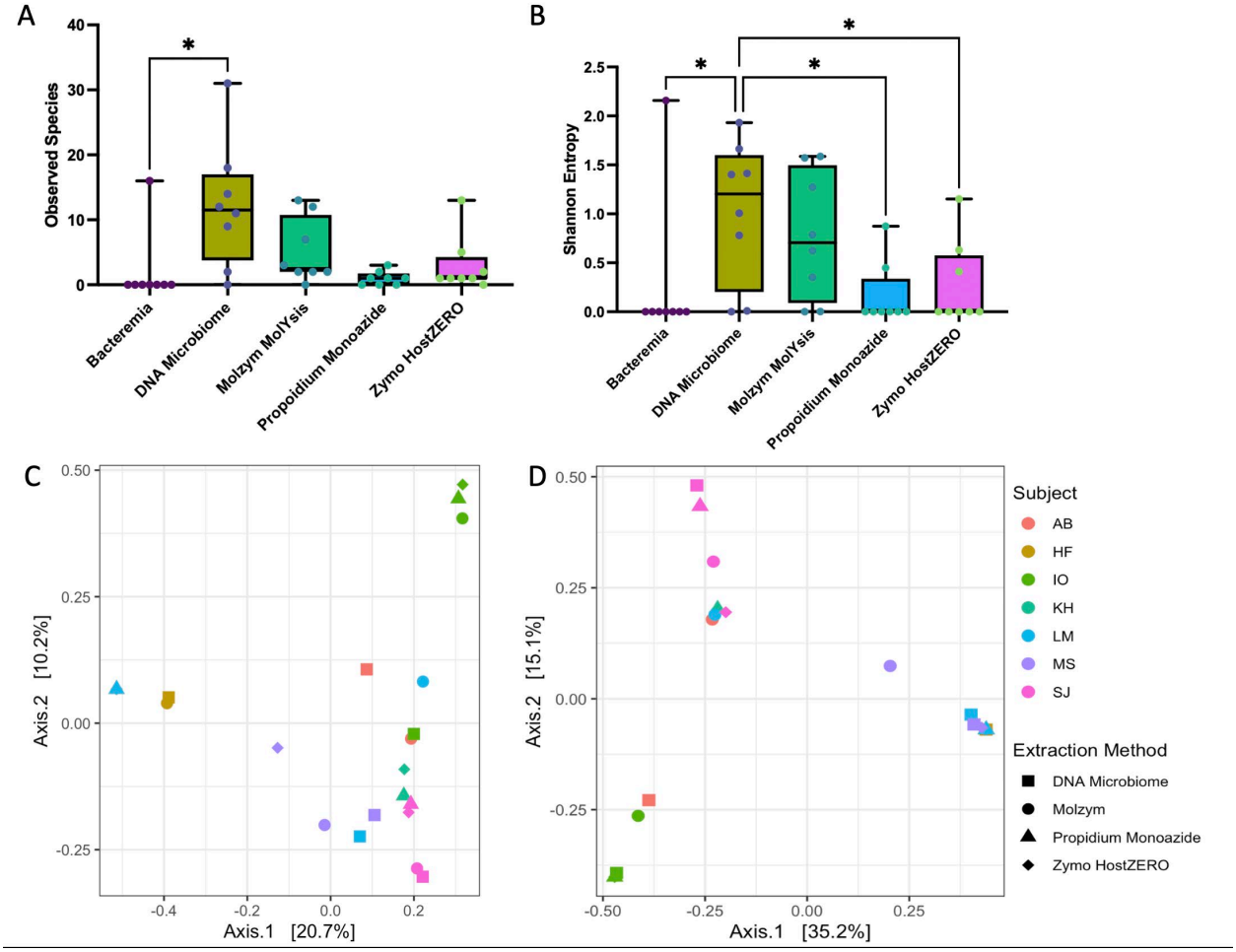


Figure 7. Top 20 microbial genera represented in urine samples. Relative abundances of the top 20 microbial genera identified in A) 16S rRNA sequencing of *unspiked* urine samples B) Shotgun metagenomic sequencing (MetaPhlAn4) of *spiked* urine samples, C) Metagenome-assembled-genomes (MAG) generated from *spiked* urine samples. The same urine samples were used for 16S and shotgun metagenomic sequencing. Across methods, *Staphylococcus* (*pseudintermedius*), *Bacillus* (*cereus*), *Streptococcus* (*canis*), and *Arcanobacterium* repeatedly emerge as abundant taxa. For 16S samples, ASVs were filtered to a minimum 0.5% abundance in at least 10% of samples.

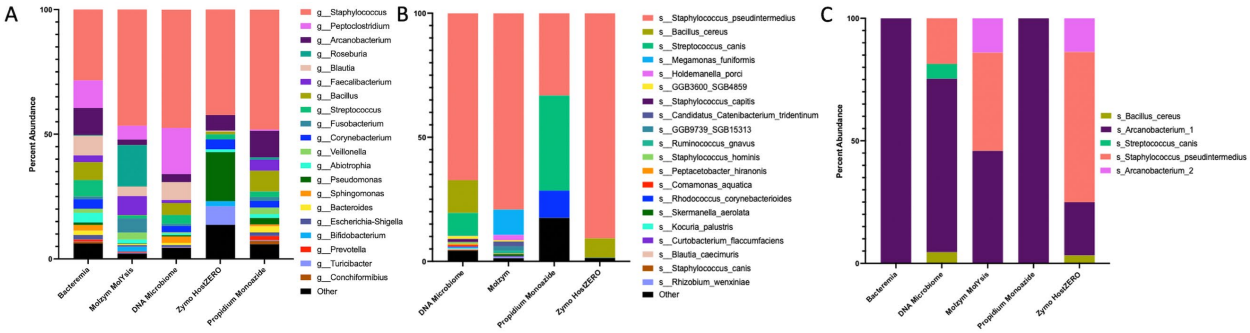
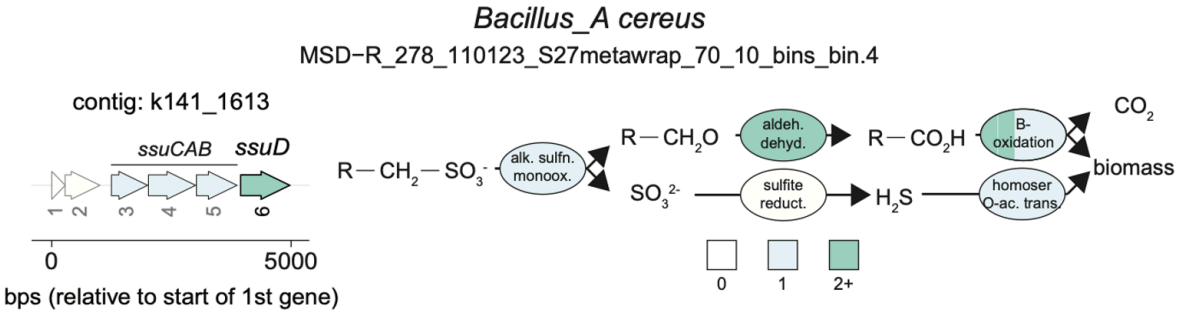


Figure 8. Metabolic potential of urine-associated MAGs. A) and B) feature reconstructed alkane metabolism in urine-associated MAGs. Shown are the regulon in which the predicted alkane metabolism gene occurs, as well as the reconstructed relevant pathway. For the depiction of the regulon, only up to ten neighboring genes on each side were included, and the coloring denotes arbitrary groupings with the gene responsible for alkane activation the darkest (*i.e.*, *ssuD* and *ladB*), and genes that weren't directly related to the predicted alkane metabolism colored grey. The numbers below the arrows indicate the gene number on the contig. For the depiction of the reconstructed alkane metabolism pathways, colors denote the number of genes that may be involved at each reaction, noting that for simplicity beta-oxidation has been summarized in one ellipse broken into five pieces. Annotation of the *Bacillus_A ladB* relative indicated that it is an alkanesulfonate monooxygenase (*ssuD*), which desulfonates organosulfonates to yield sulfite and an aldehyde. Depending on the downstream pathways available, organosulfonates degraded in this manner may be used as carbon, sulfur, or electron sources (Reichenbecher & Murrell, 1999). This *ssuD* was on a fractured contig (~5k bp) containing only genes relevant for the uptake of sulfonates (*ssuCAB*,). However, annotations from other contigs in the *Bacillus_A* bin suggest that the remaining long-chain aldehyde may be used as a carbon source via two members of the aldehyde dehydrogenase protein family (PF00171) that would produce fatty-acid resembling molecules capable of proceeding to beta-oxidation. The bin contained multiple genes for fatty acid degradation, acetyl-CoA sythetases, and acyl-CoA dehydrogenases, suggesting an ability to utilize multiple alkane groups (Jimenez-Diaz et al., 2017). Further supporting the possibility of utilizing a variety of hydrocarbons as potential carbon sources or electron donors was the recovery of 3 cytochrome p450 family genes as well as 4 divergent zinc- or iron-dependent alcohol dehydrogenases (PF00465, PF13685, PF00107, PF13602). Annotation of the *ladA* in *Staphylococcus* was less certain, as the neighboring genes in well-assembled contigs (>20 Kbp) did not reveal genes relevant for downstream pathways. However, several genes in the regulon were unable to be annotated but may be relevant. Additionally, reconstruction of degradation pathways among the rest of the contigs suggested that long-chain alkanes activated by *ladA* could be further oxidized by divergent members of the alcohol dehydrogenase protein family (PF08240, PF00107, PF13602) to an aldehyde, and then oxidized to a carboxylic acid using up to two divergent members of the aldehyde dehydrogenase protein family (PF00171) (Jimenez-Diaz et al., 2017). The fate of this carboxylic acid is uncertain -- genes for the first or third steps of beta-oxidation were not found in this bin, but there were candidate genes for the other three steps that would yield acetyl-CoA for complete oxidation or assimilation – up to two members of the acyl-CoA dehydrogenase protein family (PF08028, PF02770, PF00441, PF02771), 3-hydroxyacyl-CoA dehydrogenase, and up to two acetyl-CoA C-acetyltransferases. In the setting of odd-chain alkane activation by *Staphylococcus*, propionate secretion is likely, given the inability to introduce the 3-carbon carboxylic acid into central carbon metabolism and the presence of a propionate CoA-transferase to retain the CoA molecule.

A


Bacillus_A cereus
MSD−R_278_110123_S27metawrap_70_10_bins_bin.4
contig: k141_1613
ssuCAB
ssuD
1 2 3 4 5 6
0
5000
bps (relative to start of 1st gene)
R—CH2—SO3-
alk. sulfn. monoox.
R—CH2O
aldeh. dehyd.
R—CO2H
B-oxidation
CO2
biomass
SO3 2-
sulfite reduct.
H2S
homoser O-ac. trans.
0
1
2+

B

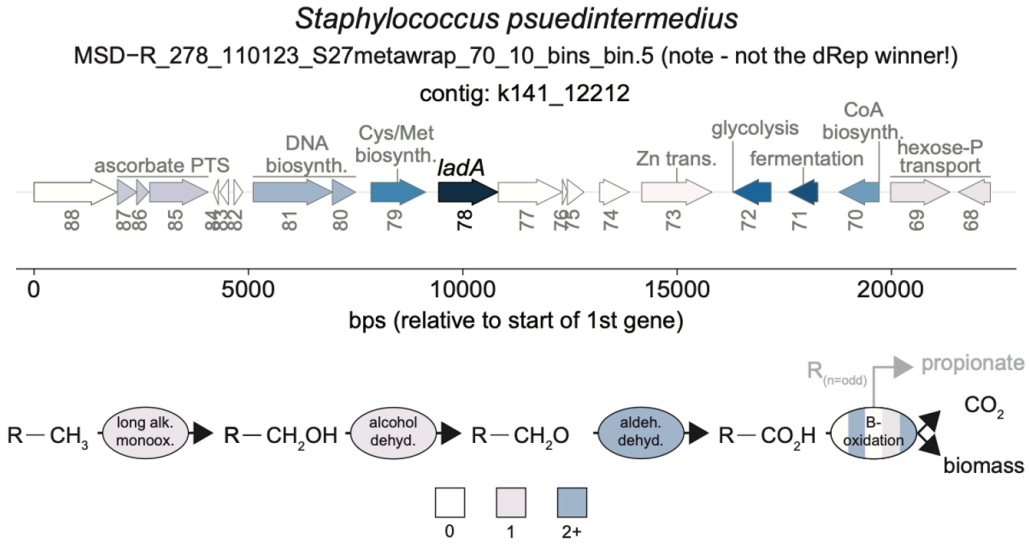

Staphylococcus psuedintermedius
MSD−R_278_110123_S27metawrap_70_10_bins_bin.5 (note - not the dRep winner!)
contig: k141_12212
ascorbate PTS
DNA biosynth.
Cys/Met biosynth.
ladA
Zn trans.
glycolysis
fermentation
CoA biosynth.
hexose-P transport
88 87 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68
0
5000
10000
15000
20000
bps (relative to start of 1st gene)
R—CH3
long alk. monoox.
R—CH2OH
alcohol dehyd.
R—CH2O
aldeh. dehyd.
R—CO2H
B-oxidation
R(n=odd)
propionate
CO2
biomass
0
1
2+

Figure S1. Experimental Design for Urine Volume Experiment. Figure generated using BioRender.

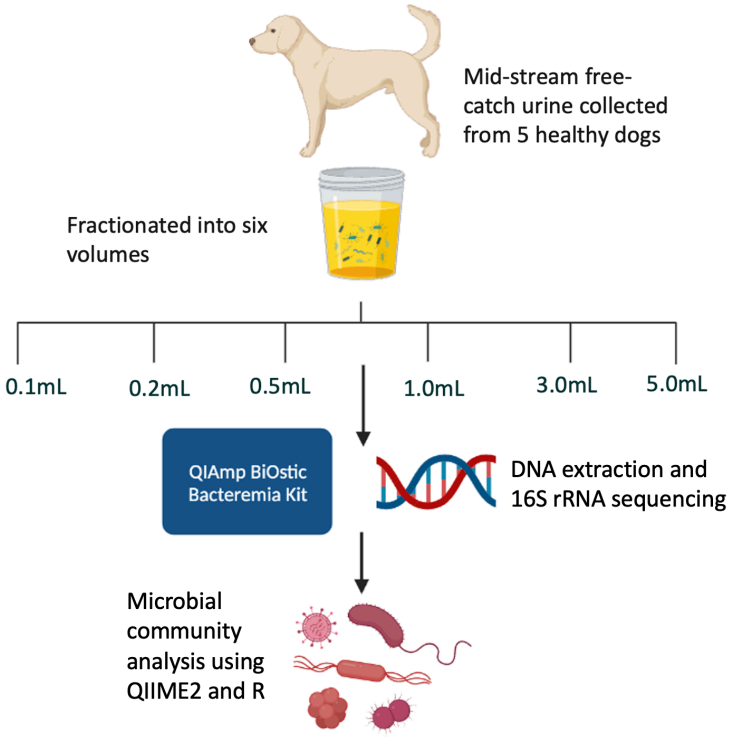


Figure S2. Experimental design for Host Cell Removal Experiments. Figure generated using BioRender.

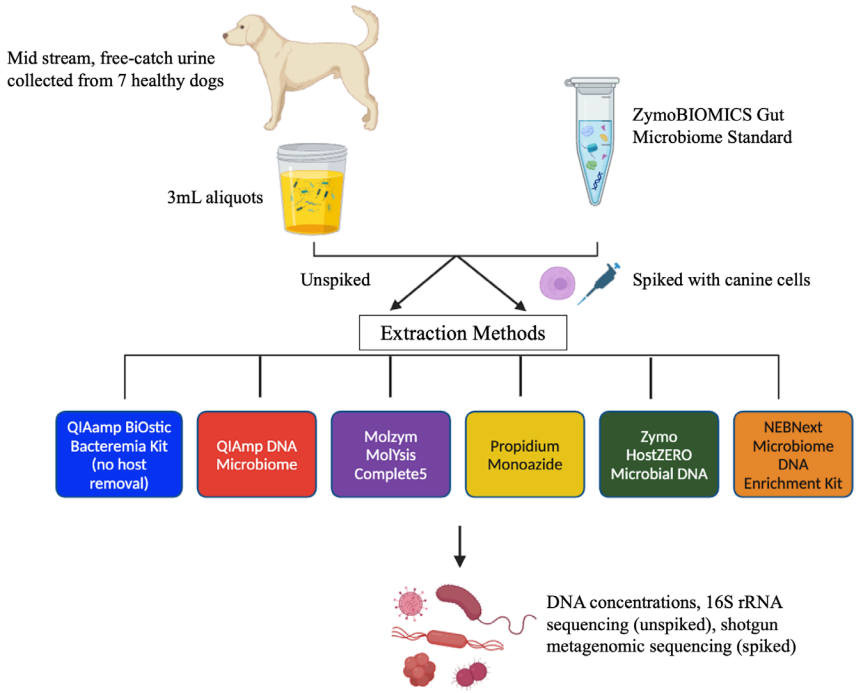


Figure S3. Cross contamination from positive control (16S). One urine sample (Dog SJ, Extraction Method Molzym MolYsis, see red arrow) exhibited a microbial composition similar to that of the ZymoBIOMICS Gut Microbiome Standard (Zymo Pos Control, pink) and was excluded from analysis for possible cross contamination**.** A) Bray-Curtis and B) weighted UniFrac PCoAs showing microbial composition of urine samples in relation to the Zymo Gut Microbiome Standard. Urine sample denoted by red arrow was removed from further analysis.

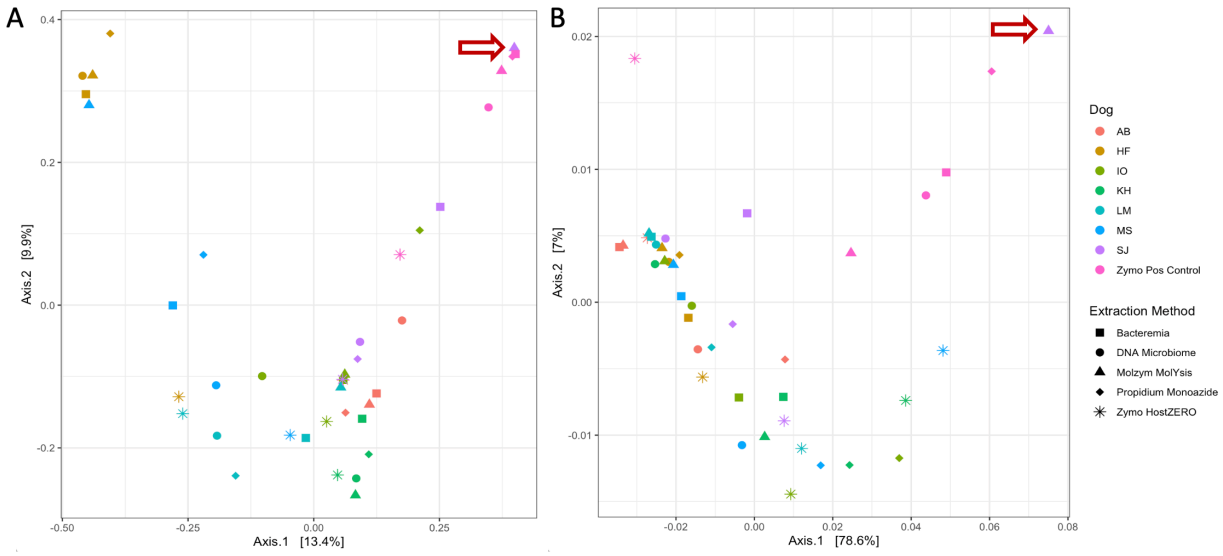


Figure S4. Comparing 9 and 12 PCR amplification cycles for shotgun metagenomics of low biomass samples. DNA from the ZymoBIOMICS Gut Microbiome Standard (n=21 microbial taxa at known relative abundances; Zymo Research, Tustin, CA) was extracted and sequenced at 3 ten-fold dilutions (D1 = 0.07 ng/uL, D2 = 0.7 ng/uL, D3 = 7 ng/uL). These dilutions were selected due to their relevance to healthy urine DNA concentrations. DNA was then amplified with 9 or 12 PCR cycles and sequenced to a depth of 50 million reads. A) Taxonomic profile assigned using microbial marker genes via SingleM. B) Metagenome assembled genome (MAG) abundances across samples assigned using the Genome Taxonomy Database. In A) and B) "9" or "12" indicates amplification cycles. Similar microbial profiles were observed at all dilutions and using 9 or 12 amplification cycles. NC, Negative Control. MAGs, metagenome-assembled genomes.

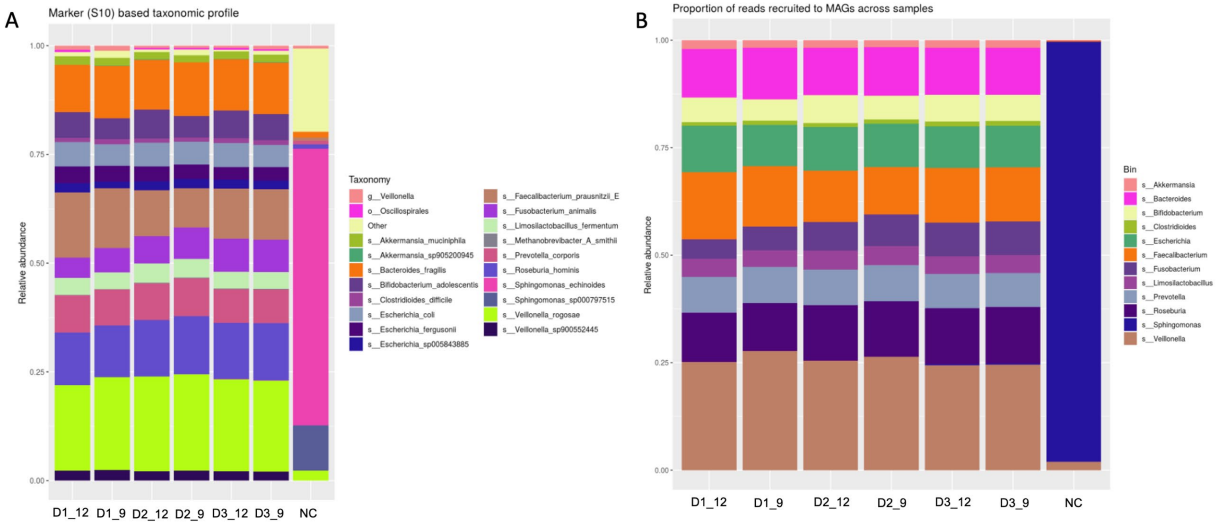


Figure S5. Urine sample volume and microbial diversity and composition. A) Average species accumulation, saturation, and achievable sequencing depth were greater in higher-volume samples. B) Microbial diversity (Shannon Entropy) modestly increased with urine sample volume, although this difference was not significant (p=0.366, Friedman). C) Urine microbial composition (Jaccard) differed significantly by dog (p=0.001, PERMANOVA) but not by sample volume (p=0.99, PERMANOVA).

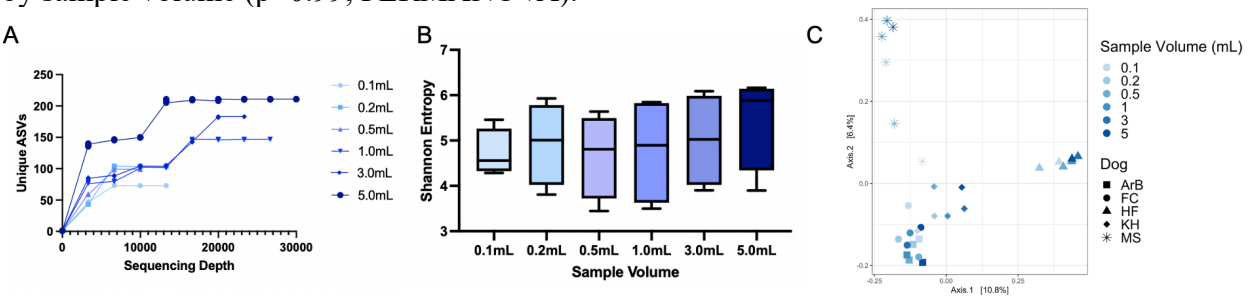


Figure S6. **Microbial composition was more variable in urine samples of lower volume.** A) - I) Microbial composition of urine samples of varying volumes within a single dog. A) - D) are based on Bray-Curtis distances. E) – I) are based on Jaccard distances. A), F) are dog HF. B), D) are dog FC. C), H) are dog ArB. D), I) are dog KH. E) is dog MS. Within dogs, microbial composition varied by sample volume with High volume samples (dark blue, 3mL, 5mL) clustering more closely (showing more similar microbial communities) than Low volume samples (light blue, 0.1, 0.2, 0.5, 1.0 mL). J) The microbial composition (Jaccard) of low volume samples was significantly more variable (greater distance to the centroid) than that of high volume samples (p=0.0006, PERMDISP).

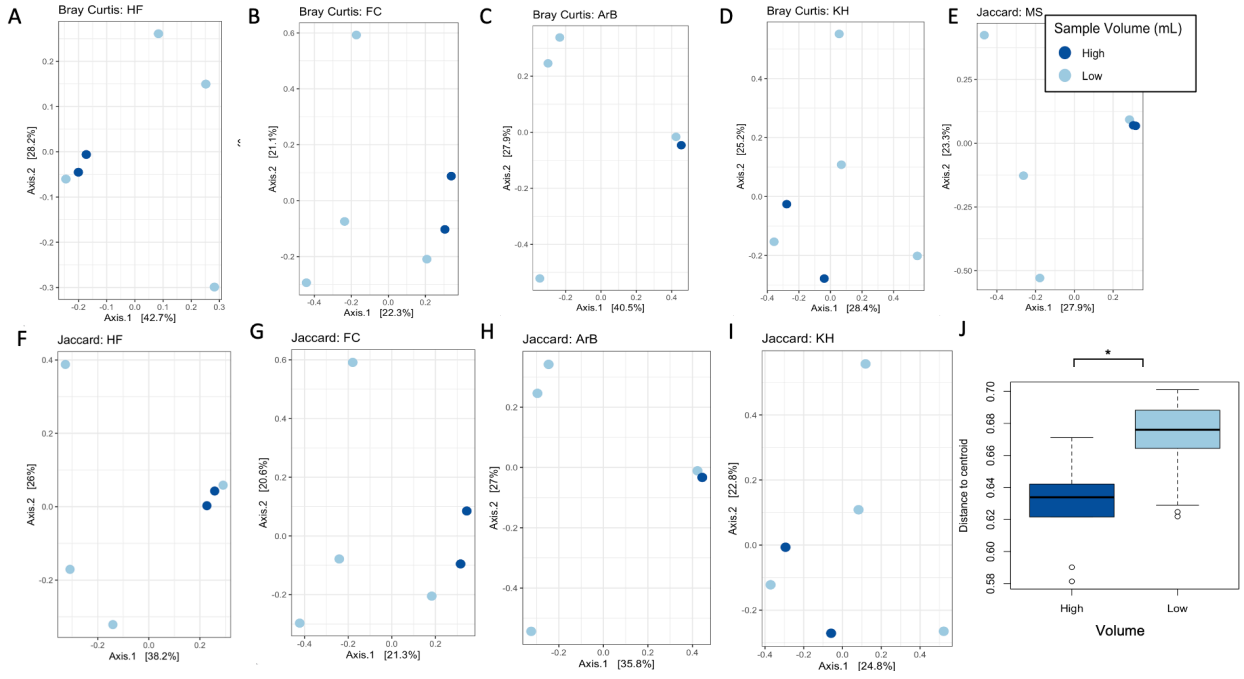


Figure S7. Total and Bacterial DNA recovery did not differ by dog. Neither A) Total DNA concentrations (ng/ul, Qubit fluorometry; p=0.15, Kruskal-Walis) nor B) bacterial DNA concentrations (qPCR; p=0.09, Kruskal-Walis) differed by dog. Neither C) Total DNA concentrations nor D) bacterial DNA concentrations from urine samples spiked with canine cells differed by dog (Total DNA: p=0.99, Bacterial DNA: p=0.28, Kruskal-Wallis). Bars represent the mean with standard error. Pipettor icon in C) and D) indicates that all samples shown in these graphs were spiked with canine thyroid adenocarcinoma (CTAC) cells.

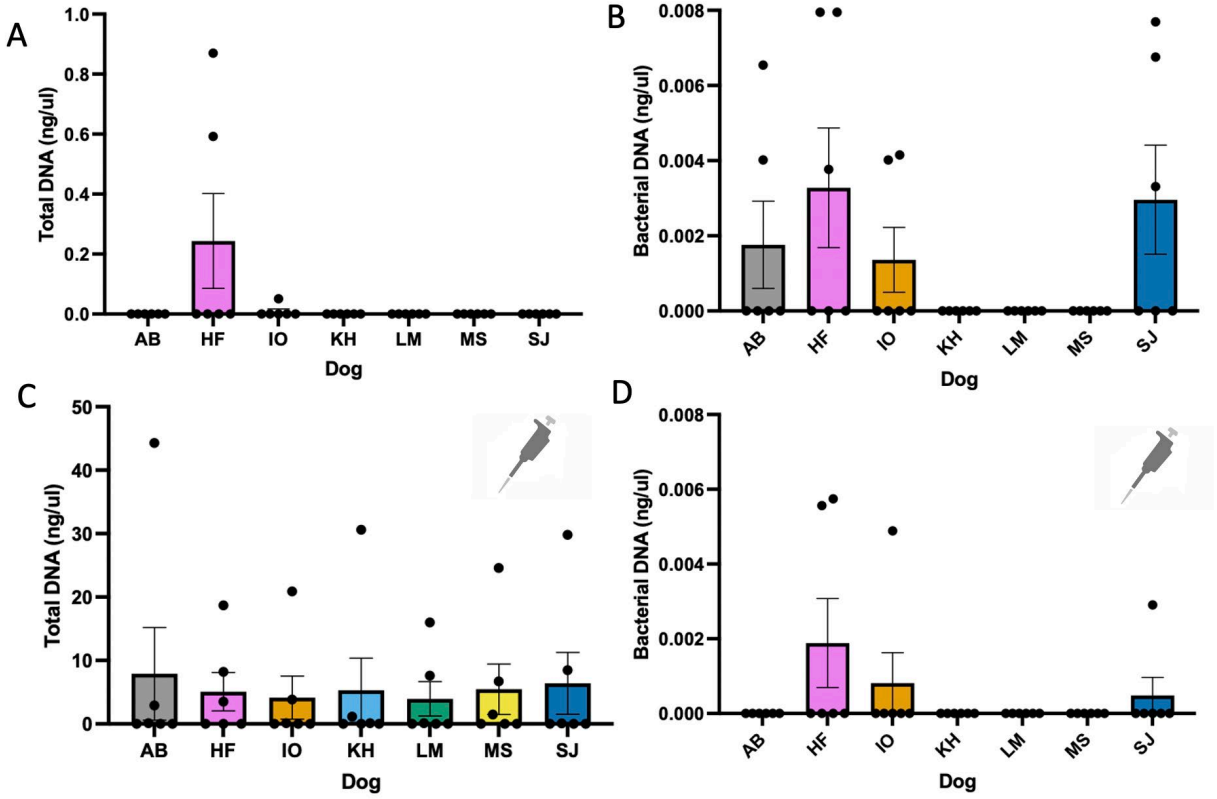


Figure S8. **Microbial diversity and composition by extraction method (16S).** A) Microbial diversity (Faith's phylogenetic diversity) differed significantly by extraction method (p=0.0004, Friedman, multiple comparisons with FDR q=0.05). *p<0.05, **p<0.01 B) Microbial composition (Jaccard) differed significantly by dog (p=0.001) but not by kit (p=0.72, PERMANOVA). C) When weighted by microbial phylogeny (unweighted UniFrac), microbial composition differed significantly both by dog (p=0.005) and kit (p=0.005, PERMANOVA).

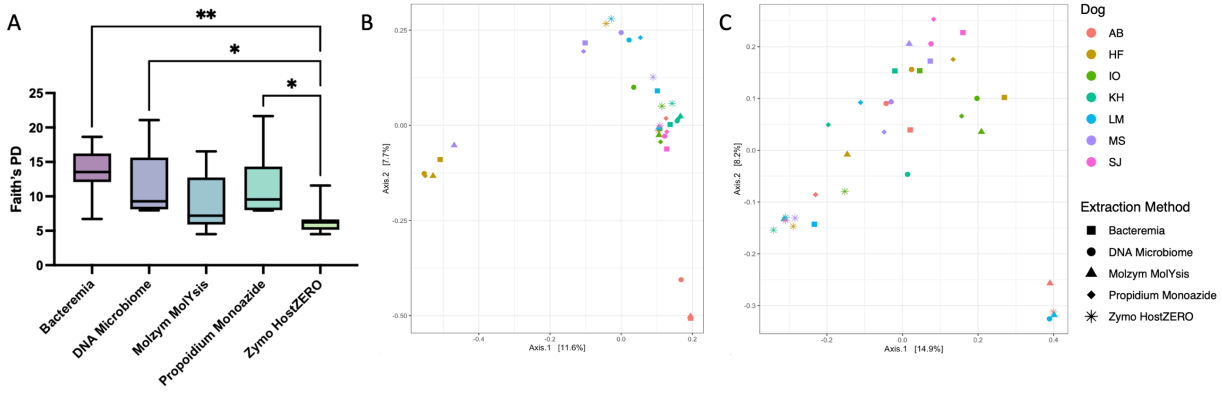
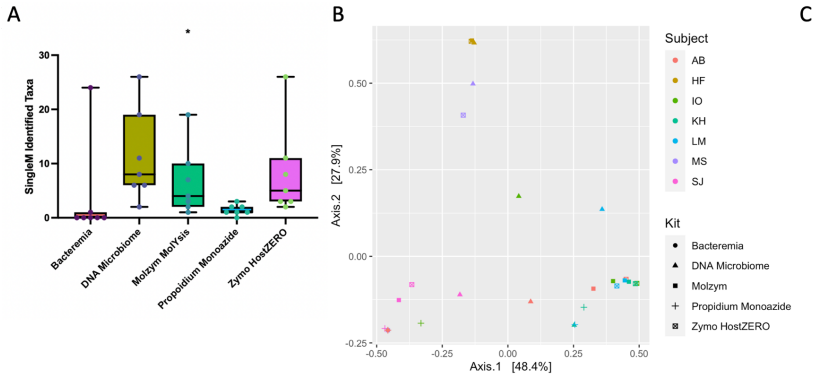


Figure S9. Metagenomic marker gene profiling using SingleM and SingleM "condense" was largely congruent with MetaPhlAn4-based profiling of spiked urine samples. A) Microbial richness (unique taxa) differed significantly by extraction method (p=0.017, Friedman), with DNA Microbiome generally recovering the greatest number of taxa. B) Microbial composition also differed significantly by dog (Bray Curtis, p=0.001, PERMANOVA) but not by extraction method (Bray Curtis, p=0.1, PERMANOVA). C) Top 20 taxa represented in the SingleM dataset. Bolded taxa indicate taxa that were also identified as "Top 20" taxa by



C

Taxon
g__Sphingomonas
s__Staphylococcus_pseudintermedius
s__Limosilactobacillus_fermentum
g__Arcanobacterium
p__Actinobacteriota
o__Actinomycetales
s__Streptococcus_canis
p__Proteobacteria
f__Actinomycetaceae
g__Nocardioides
s__Bacillus_A_cereus
f__Microbacteriaceae
f__Gaiellaceae
s__Veillonella_rogosae
o__Acidimicrobiales
o__Mycobacteriales
f__Nocardioidaceae
f__Propionibacteriaceae
f__70-9
s__Megamonas_funiformis

Figure S10. Assembly metrics of all sequenced samples, including spiked urine, spiked ZymoBIOMICs Gut Microbiome Standards. A) All bases assembled into contigs by sample (assembled bases), B) bases assembled into contigs of 1kb or greater (1kb+), C) assembled bases 5kb+, D) assembled bases 25kb+, E) assembled bases 50kb+, F) All assembled contigs by sample (contigs), G) contigs 1kb+, H) contigs 5kb+, I) contigs 25kb+, J) contigs 50kb+, K) percent of reads mapping to MAGs within each sample.

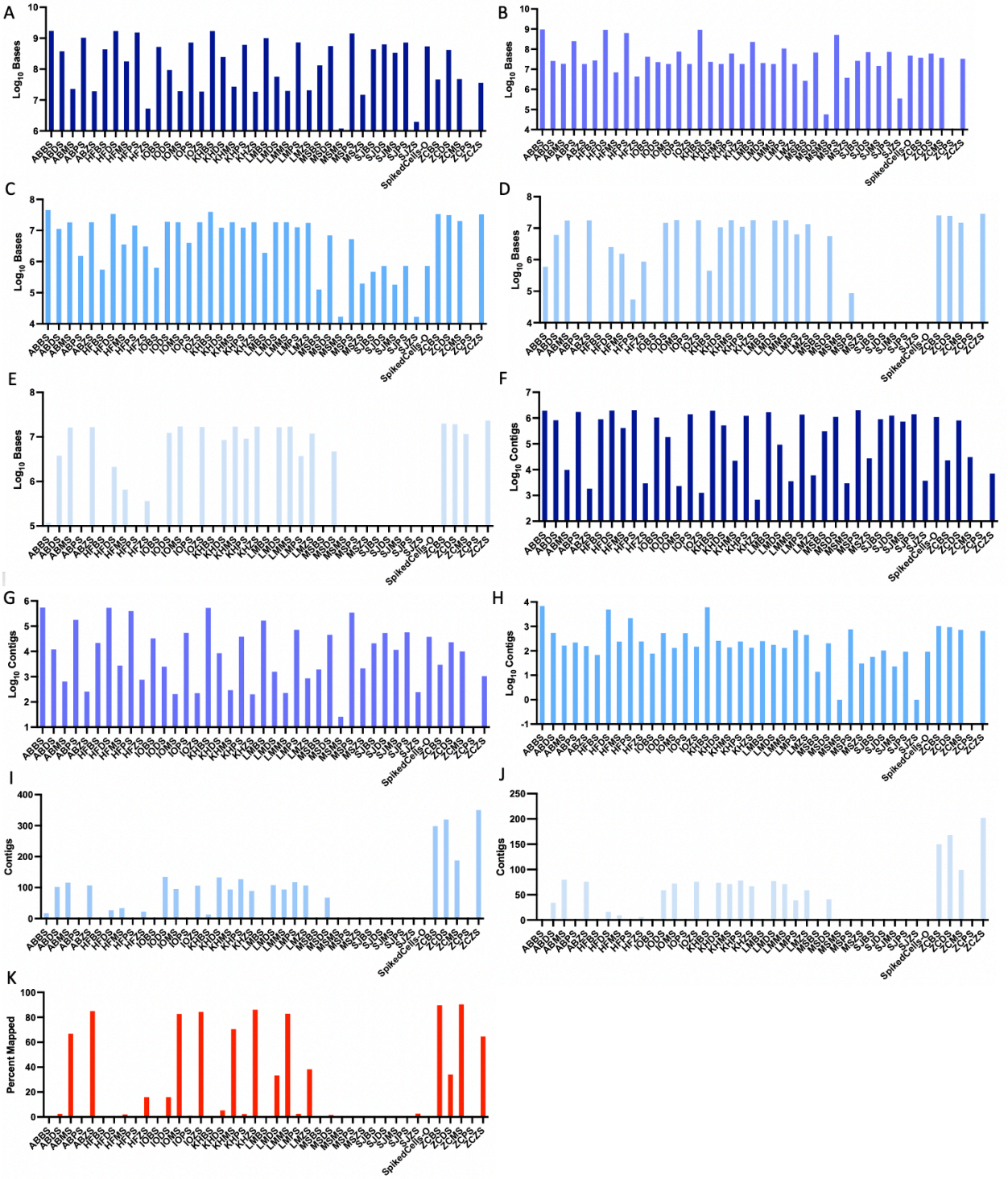


Figure S11. Metagenome-assembled genome (MAG) counts by extraction method. A total of 26 MAGs were generated across the entire dataset: 11 were assembled from the the Zymo Gut Microbiome Standard, 5 were urine-derived, and 10 were identified as contaminants and were assembled from negative controls or urine samples. A) There was no significant difference in the number of non-contaminant MAGs by extraction method (p=0.3, Friedman). B) Contaminant MAG counts varied significantly by extraction method (overall p=0.018, Friedman, no pairwise significant) with DNA Microbiome producing fewer contaminant MAGs as compared other extraction methods. Contaminant MAGs were identified using *decontam.*

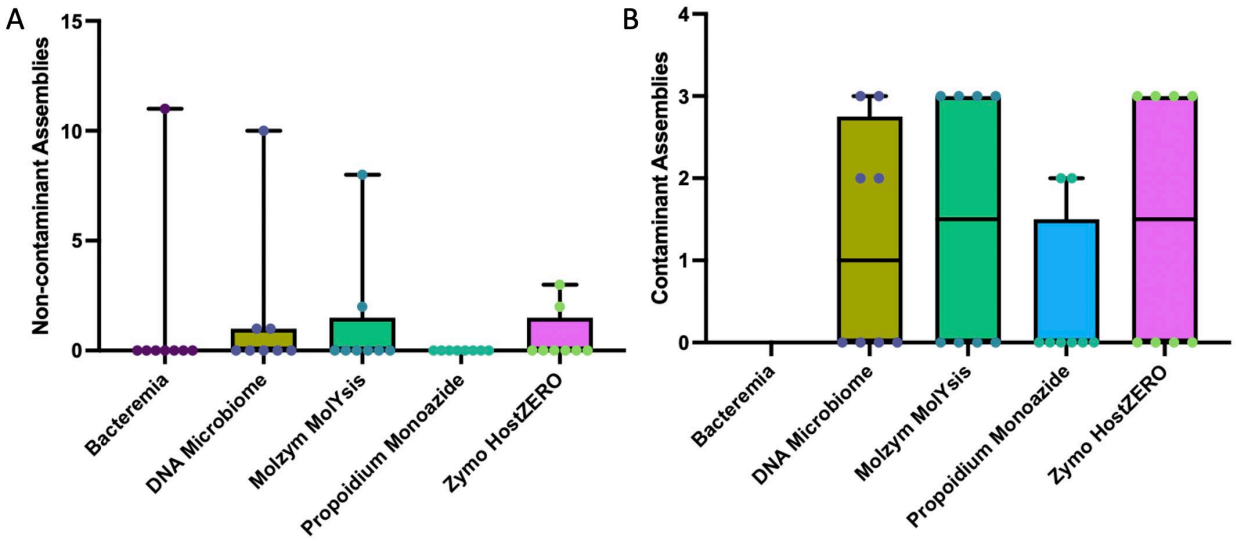


Figure S12. The expected and actual composition of the ZymoBIOMICs Gut Microbiome Standard (positive control) across extraction methods. Taxa bar plots of expected and actual profiles of the ZymoBIOMICs Gut Microbiome Standard based on A) 16S rRNA sequencing (unspiked), B) shotgun metagenomic sequencing (MetaPhlAn4, samples spiked with canine cells), and C) MAG assembly from shotgun metagenomic sequencing (spiked). Bacteremia and DNA Microbiome yielded taxonomic profiles and abundances most similar to expected.

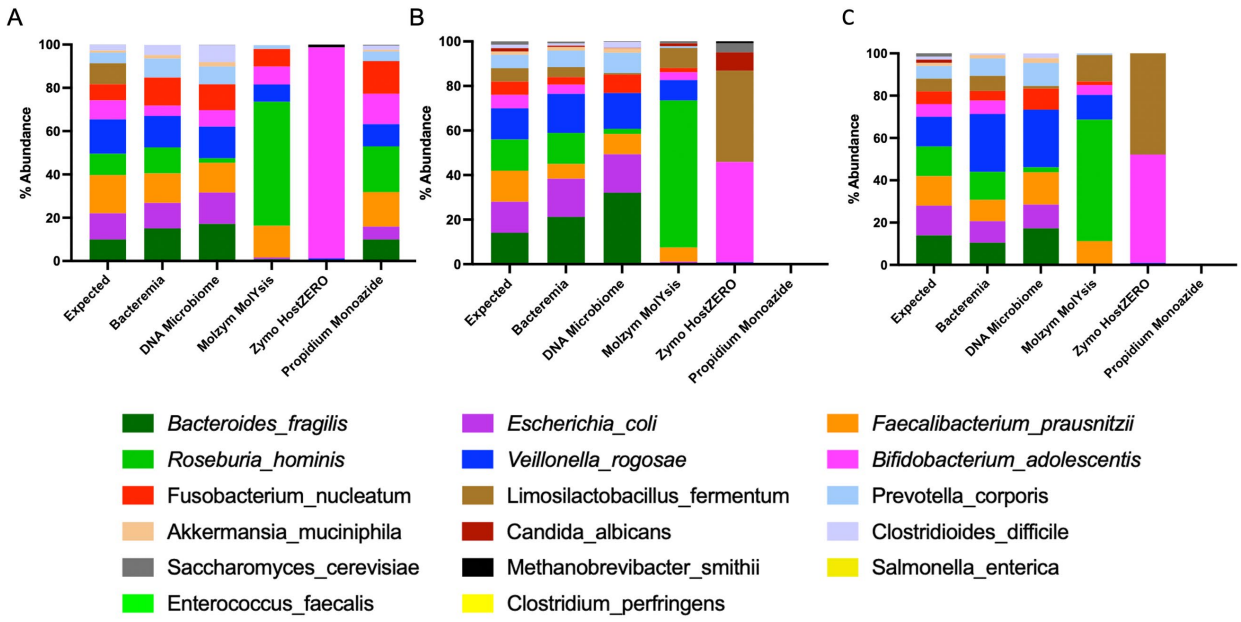
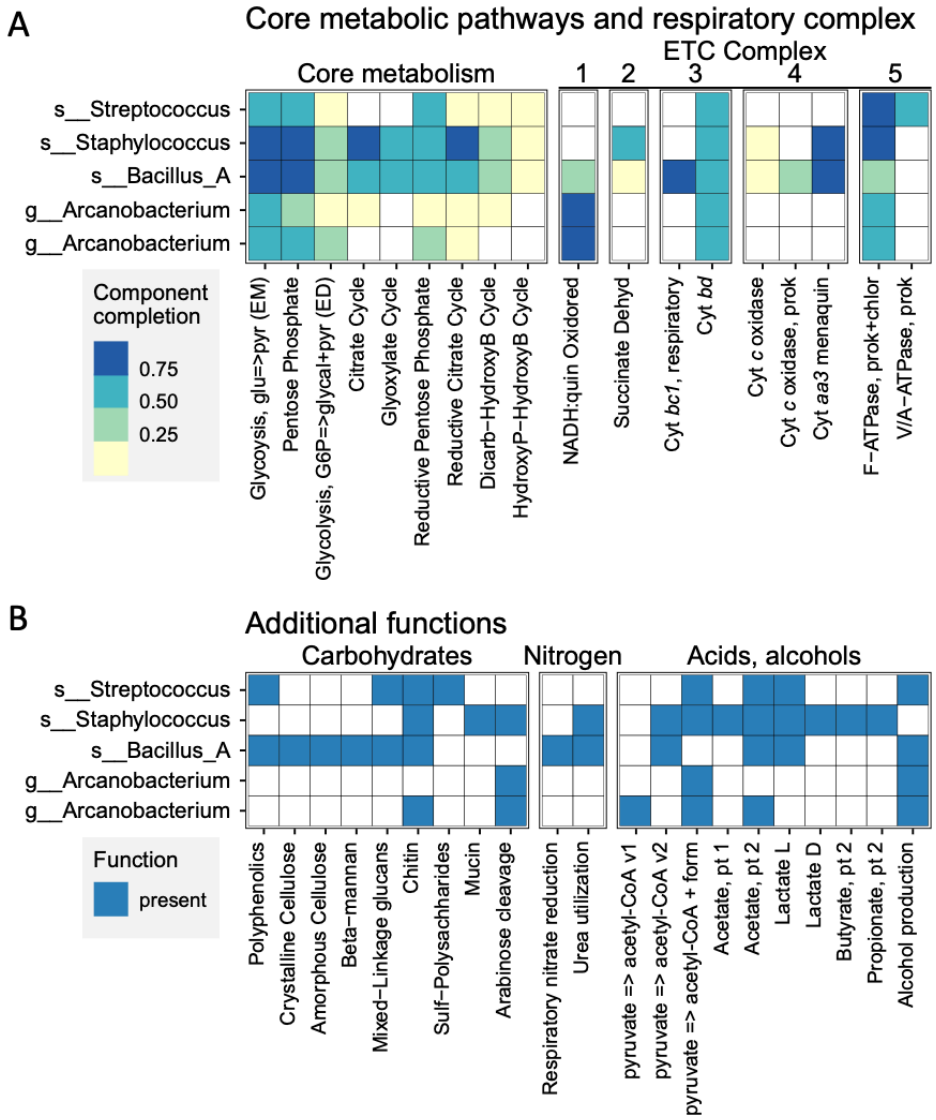


Figure S13. Metabolic potential of urine-associated MAGs. A) Core metabolic pathways and B) additional metabolic pathways identified in urine-associated metagenome assembled genomes (MAGs). Genomic potential from dereplicated genomes are shown. Functions absent across dataset (*e.g.*, starch degradation) have been removed for clarity.



References

1. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214. https://doi.org/10.1038/nature11234.

2. Hou, K., Wu, Z.-X., Chen, X.-Y., Wang, J.-Q., Zhang, D., Xiao, C., Zhu, D., Koya, J.B., Wei, L., Li, J., et al. (2022). Microbiota in health and diseases. Sig Transduct Target Ther *7*, 1–28. https://doi.org/10.1038/s41392-022-00974-4.

3. Reichenbecher, W., and Murrell, J.C. (1999). Linear alkanesulfonates as carbon and energy sources for gram-positive and gram-negative bacteria. Archives of Microbiology *171*, 430–438. https://doi.org/10.1007/s002030050730.

4. Hilt, E.E., McKinley, K., Pearce, M.M., Rosenfeld, A.B., Zilliox, M.J., Mueller, E.R., Brubaker, L., Gai, X., Wolfe, A.J., and Schreckenberger, P.C. (2014). Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. J Clin Microbiol *52*, 871–876. https://doi.org/10.1128/JCM.02876-13.

5. Jimenez-Diaz, L., Caballero, A., and Segura, A. (2017). Pathways for the Degradation of Fatty Acids in Bacteria. In Aerobic Utilization of Hydrocarbons, Oils and Lipids, F. Rojo, ed. (Springer International Publishing), pp. 1–23. https://doi.org/10.1007/978-3-319-39782-5_42-1.

6. Claeys, K.C., Blanco, N., Morgan, D.J., Leekha, S., and Sullivan, K.V. (2019). Advances and Challenges in the Diagnosis and Treatment of Urinary Tract Infections: the Need for Diagnostic Stewardship. Curr Infect Dis Rep *21*, 11. https://doi.org/10.1007/s11908-019-0668-7.

7. Vap, L.M., and Shropshire, S.B. (2017). Urine Cytology: Collection, Film Preparation, and Evaluation. Veterinary Clinics of North America: Small Animal Practice *47*, 135–149. https://doi.org/10.1016/j.cvsm.2016.07.009.

8. Cai, T., Mazzoli, S., Mondaini, N., Meacci, F., Nesi, G., D'Elia, C., Malossini, G., Boddi, V., and Bartoletti, R. (2012). The Role of Asymptomatic Bacteriuria in Young Women With Recurrent Urinary Tract Infections: To Treat or Not to Treat? Clinical Infectious Diseases *55*, 771–777. https://doi.org/10.1093/cid/cis534.

9. Foxman, B. (2014). Urinary Tract Infection Syndromes. Infectious Disease Clinics of North America *28*, 1–13. https://doi.org/10.1016/j.idc.2013.09.003.

10. Al, K.F., Joris, B.R., Daisley, B.A., Chmiel, J.A., Bjazevic, J., Reid, G., Gloor, G.B., Denstedt, J.D., Razvi, H., and Burton, J.P. (2023). Multi-site microbiota alteration is a hallmark of kidney stone formation. Microbiome *11*, 263. https://doi.org/10.1186/s40168-023-01703-x.

11. Coffey, E.L., Gomez, A.M., Burton, E.N., Granick, J.L., Lulich, J.P., and Furrow, E. (2022). Characterization of the urogenital microbiome in Miniature Schnauzers with and without calcium oxalate urolithiasis. Journal of Veterinary Internal Medicine *36*, 1341–1352. https://doi.org/10.1111/jvim.16482.

12. Nelson, D.E., Dong, Q., Van der Pol, B., Toh, E., Fan, B., Katz, B.P., Mi, D., Rong, R., Weinstock, G.M., Sodergren, E., et al. (2012). Bacterial communities of the coronal sulcus and distal urethra of adolescent males. PLoS One *7*, e36298. https://doi.org/10.1371/journal.pone.0036298.

13. Mrofchak, R., Madden, C., Evans, M.V., Kisseberth, W.C., Dhawan, D., Knapp, D.W., and Hale, V.L. (2022). Urine and fecal microbiota in a canine model of bladder cancer and comparison of canine and human urine microbiota. All Life *15*, 1245–1263. https://doi.org/10.1080/26895293.2022.2154858.

14. Wolfe, A.J., Toh, E., Shibata, N., Rong, R., Kenton, K., FitzGerald, M., 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. https://doi.org/10.1128/JCM.05852-11.

15. Wolfe, A.J., and Brubaker, L. (2015). "Sterile Urine" and the Presence of Bacteria. Eur Urol *68*, 173–174. https://doi.org/10.1016/j.eururo.2015.02.041.

16. Bukavina, L., Isali, I., Ginwala, R., Sindhani, M., Calaway, A., Magee, D., Miron, B., Correa, A., Kutikov, A., Zibelman, M., et al. (2023). Global Meta-analysis of Urine Microbiome: Colonization of Polycyclic Aromatic Hydrocarbon–degrading Bacteria Among Bladder Cancer Patients. European Urology Oncology *6*, 190–203. https://doi.org/10.1016/j.euo.2023.02.004.

17. Price, T.K., Wolff, B., Halverson, T., Limeira, R., Brubaker, L., Dong, Q., Mueller, E.R., and Wolfe, A.J. (2020). Temporal Dynamics of the Adult Female Lower Urinary Tract Microbiota. mBio *11*, e00475-20. https://doi.org/10.1128/mBio.00475-20.

18. Adu-Oppong, B., Thänert, R., Wallace, M.A., Burnham, C.-A.D., and Dantas, G. (2022). Substantial overlap between symptomatic and asymptomatic genitourinary microbiota states. Microbiome *10*, 6. https://doi.org/10.1186/s40168-021-01204-9.

19. Thomas-White, K., Forster, S.C., Kumar, N., Van Kuiken, M., Putonti, C., Stares, M.D., Hilt, E.E., Price, T.K., Wolfe, A.J., and Lawley, T.D. (2018). Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. Nat Commun *9*, 1557. https://doi.org/10.1038/s41467-018-03968-5.

20. 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. mBio *5*, e01283-14. https://doi.org/10.1128/mBio.01283-14.

21. Du, J., Khemmani, M., Halverson, T., Ene, A., Limeira, R., Tinawi, L., Hochstedler-Kramer, B.R., Noronha, M.F., Putonti, C., and Wolfe, A.J. (2024). Cataloging the phylogenetic diversity of human bladder bacterial isolates. Genome Biology *25*, 75. https://doi.org/10.1186/s13059-024-03216-8.

22. Mrofchak, R., Madden, C., Evans, M.V., and Hale, V.L. (2021). Evaluating extraction methods to study canine urine microbiota. PLOS ONE *16*, e0253989. https://doi.org/10.1371/journal.pone.0253989.

23. Coffey, E.L., Gomez, A.M., Ericsson, A.C., Burton, E.N., Granick, J.L., Lulich, J.P., and Furrow, E. (2023). The impact of urine collection method on canine urinary microbiota detection: a cross-sectional study. BMC Microbiol *23*, 101. https://doi.org/10.1186/s12866-023-02815-y.

24. Hoffman, C., Siddiqui, N.Y., Fields, I., Gregory, W.T., Simon, H.M., Mooney, M.A., Wolfe, A.J., and Karstens, L. (2021). Species-Level Resolution of Female Bladder Microbiota from 16S rRNA Amplicon Sequencing. mSystems *6*, 10.1128/msystems.00518-21. https://doi.org/10.1128/msystems.00518-21.

25. HICKLING, D.R., SUN, T.-T., and WU, X.-R. (2015). Anatomy and Physiology of the Urinary Tract: Relation to Host Defense and Microbial Infection. Microbiol Spectr *3*, 10.1128/microbiolspec.UTI-0016–2012. https://doi.org/10.1128/microbiolspec.UTI-0016-2012.

26. 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. https://doi.org/10.1038/nrurol.2014.361.

27. 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*, 4333. https://doi.org/10.1038/s41598-018-22660-8.

28. Martin, A., Woolbright, B.L., Umar, S., Ingersoll, M.A., and Taylor, J.A. (2022). Bladder cancer, inflammageing and microbiomes. Nat Rev Urol *19*, 495–509. https://doi.org/10.1038/s41585-022-00611-3.

29. Isali, I., Helstrom, E.K., Uzzo, N., Lakshmanan, A., Nandwana, D., Valentine, H., Sindhani, M., Abbosh, P., and Bukavina, L. (2024). Current Trends and Challenges of Microbiome Research in Bladder Cancer. Curr Oncol Rep *26*, 292–298. https://doi.org/10.1007/s11912-024-01508-7.

30. Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., et al. (2018). Best practices for analysing microbiomes. Nat Rev Microbiol *16*, 410–422. https://doi.org/10.1038/s41579-018-0029-9.

31. Pohl, H.G., Groah, S.L., Pérez-Losada, M., Ljungberg, I., Sprague, B.M., Chandal, N., Caldovic, L., and Hsieh, M. (2020). The Urine Microbiome of Healthy Men and Women

Differs by Urine Collection Method. Int Neurourol J *24*, 41–51. https://doi.org/10.5213/inj.1938244.122.

32. Adebayo, A.S., Ackermann, G., Bowyer, R.C.E., Wells, P.M., Humphreys, G., Knight, R., Spector, T.D., and Steves, C.J. (2020). The Urinary Tract Microbiome in Older Women Exhibits Host Genetic and Environmental Influences. Cell Host & Microbe *28*, 298-305.e3. https://doi.org/10.1016/j.chom.2020.06.022.

33. Hrbacek, J., Morais, D., Cermak, P., Hanacek, V., and Zachoval, R. (2021). Alpha-diversity and microbial community structure of the male urinary microbiota depend on urine sampling method. Sci Rep *11*, 23758. https://doi.org/10.1038/s41598-021-03292-x.

34. Hrbáček, J., Tláskal, V., Čermák, P., Hanáček, V., and Zachoval, R. (2023). Bladder cancer is associated with decreased urinary microbiota diversity and alterations in microbial community composition. Urologic Oncology: Seminars and Original Investigations *41*, 107.e15-107.e22. https://doi.org/10.1016/j.urolonc.2022.09.018.

35. Lev-Sagie, A., Goldman-Wohl, D., Cohen, Y., Dori-Bachash, M., Leshem, A., Mor, U., Strahilevitz, J., Moses, A.E., Shapiro, H., Yagel, S., et al. (2019). Vaginal microbiome transplantation in women with intractable bacterial vaginosis. Nat Med *25*, 1500–1504. https://doi.org/10.1038/s41591-019-0600-6.

36. Wang, Z., He, Y., and Zheng, Y. (2019). Probiotics for the Treatment of Bacterial Vaginosis: A Meta-Analysis. Int J Environ Res Public Health *16*, 3859. https://doi.org/10.3390/ijerph16203859.

37. Chee, W.J.Y., Chew, S.Y., and Than, L.T.L. (2020). Vaginal microbiota and the potential of Lactobacillus derivatives in maintaining vaginal health. Microb Cell Fact *19*, 203. https://doi.org/10.1186/s12934-020-01464-4.

38. Komesu, Y.M., Dinwiddie, D.L., Richter, H.E., Lukacz, E.S., Sung, V.W., Siddiqui, N.Y., Zyczynski, H.M., Ridgeway, B., Rogers, R.G., Arya, L.A., et al. (2020). Defining the relationship between vaginal and urinary microbiomes. American Journal of Obstetrics and Gynecology *222*, 154.e1-154.e10. https://doi.org/10.1016/j.ajog.2019.08.011.

39. Neugent, M.L., Hulyalkar, N.V., Nguyen, V.H., Zimmern, P.E., and De Nisco, N.J. (2020). Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. mBio *11*, e00218-20. https://doi.org/10.1128/mBio.00218-20.

40. Stapleton, A.E., Au-Yeung, M., Hooton, T.M., Fredricks, D.N., Roberts, P.L., Czaja, C.A., Yarova-Yarovaya, Y., Fiedler, T., Cox, M., and Stamm, W.E. (2011). Randomized, Placebo-Controlled Phase 2 Trial of a Lactobacillus crispatus Probiotic Given Intravaginally for Prevention of Recurrent Urinary Tract Infection. Clin Infect Dis *52*, 1212–1217. https://doi.org/10.1093/cid/cir183.

41. De Nisco, N.J., Neugent, M., Mull, J., Chen, L., Kuprasertkul, A., de Souza Santos, M., Palmer, K.L., Zimmern, P., and Orth, K. (2019). Direct detection of tissue-resident bacteria and chronic inflammation in the bladder wall of postmenopausal women with recurrent

urinary tract infection. J Mol Biol *431*, 4368–4379. https://doi.org/10.1016/j.jmb.2019.04.008.

42. Flores, C., Ling, J., Loh, A., Maset, R.G., Aw, A., White, I.J., Fernando, R., and Rohn, J.L. (2023). A human urothelial microtissue model reveals shared colonization and survival strategies between uropathogens and commensals. Sci. Adv. *9*, eadi9834. https://doi.org/10.1126/sciadv.adi9834.

43. Dubourg, G., Morand, A., Mekhalif, F., Godefroy, R., Corthier, A., Yacouba, A., Diakite, A., Cornu, F., Cresci, M., Brahimi, S., et al. (2020). Deciphering the Urinary Microbiota Repertoire by Culturomics Reveals Mostly Anaerobic Bacteria From the Gut. Front. Microbiol. *11*, 513305. https://doi.org/10.3389/fmicb.2020.513305.

44. Worby, C.J., Schreiber, H.L., Straub, T.J., van Dijk, L.R., Bronson, R.A., Olson, B.S., Pinkner, J.S., Obernuefemann, C.L.P., Muñoz, V.L., Paharik, A.E., et al. (2022). Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women. Nat Microbiol *7*, 630–639. https://doi.org/10.1038/s41564-022-01107-x.

45. Graziani, C., Laterza, L., Talocco, C., Pizzoferrato, M., Di Simone, N., D'Ippolito, S., Ricci, C., Gervasoni, J., Persichilli, S., Del Chierico, F., et al. (2022). Intestinal Permeability and Dysbiosis in Female Patients with Recurrent Cystitis: A Pilot Study. J Pers Med *12*, 1005. https://doi.org/10.3390/jpm12061005.

46. Burton, E.N., Cohn, L.A., Reinero, C.N., Rindt, H., Moore, S.G., and Ericsson, A.C. (2017). Characterization of the urinary microbiome in healthy dogs. PLOS ONE *12*, e0177783. https://doi.org/10.1371/journal.pone.0177783.

47. Brubaker, L., Nager, C.W., Richter, H.E., Visco, A., Nygaard, I., Barber, M.D., Schaffer, J., Meikle, S., Wallace, D., Shibata, N., et al. (2014). Urinary bacteria in adult women with urgency urinary incontinence. Int Urogynecol J *25*, 1179–1184. https://doi.org/10.1007/s00192-013-2325-2.

48. Pearce, M.M., Zilliox, M.J., Thomaswhite, K.J., Richter, H.E., Nager, C.W., Visco, A.G., Nygaard, I., Barber, M.D., Schaffer, J., Moalli, P., et al. (2015). The Female Urinary Microbiota in Urgency Urinary Incontinence. Am J Obstet Gynecol *213*, 347.e1-347.e11. https://doi.org/10.1016/j.ajog.2015.07.009.

49. Costea, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos, W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., et al. (2018). Enterotypes in the landscape of gut microbial community composition. Nat Microbiol *3*, 8–16. https://doi.org/10.1038/s41564-017-0072-8.

50. Altenhoefer, A., Oswald, S., Sonnenborn, U., Enders, C., Schulze, J., Hacker, J., and Oelschlaeger, T.A. (2004). The probiotic Escherichia coli strain Nissle 1917 interferes with invasion of human intestinal epithelial cells by different enteroinvasive bacterial pathogens. FEMS Immunology & Medical Microbiology *40*, 223–229. https://doi.org/10.1016/S0928-8244(03)00368-7.

51. Terlizzi, M.E., Gribaudo, G., and Maffei, M.E. (2017). UroPathogenic Escherichia coli (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. Front Microbiol *8*, 1566. https://doi.org/10.3389/fmicb.2017.01566.

52. Geerlings, S.E. (2016). Clinical Presentations and Epidemiology of Urinary Tract Infections. Microbiology Spectrum *4*, 10.1128/microbiolspec.uti-0002–2012. https://doi.org/10.1128/microbiolspec.uti-0002-2012.

53. Irwin, D.E., Milsom, I., Hunskaar, S., Reilly, K., Kopp, Z., Herschorn, S., Coyne, K., Kelleher, C., Hampel, C., Artibani, W., et al. (2006). Population-Based Survey of Urinary Incontinence, Overactive Bladder, and Other Lower Urinary Tract Symptoms in Five Countries: Results of the EPIC Study. European Urology *50*, 1306–1315. https://doi.org/10.1016/j.eururo.2006.09.019.

54. Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L.S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends in Microbiology *27*, 105–117. https://doi.org/10.1016/j.tim.2018.11.003.

55. Karstens, L., Asquith, M., Caruso, V., Rosenbaum, J.T., Fair, D., Braun, J., Gregory, T., Nardos, R., and McWeeney, S. (2018). Community profiling of the urinary microbiota: Methodological considerations for low microbial biomass biological samples. Nat Rev Urol *15*, 735–749. https://doi.org/10.1038/s41585-018-0104-z.

56. Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology *12*, 87. https://doi.org/10.1186/s12915-014-0087-z.

57. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226. https://doi.org/10.1186/s40168-018-0605-2.

58. Coelho, L.P., Kultima, J.R., Costea, P.I., Fournier, C., Pan, Y., Czarnecki-Maulden, G., Hayward, M.R., Forslund, S.K., Schmidt, T.S.B., Descombes, P., et al. (2018). Similarity of the dog and human gut microbiomes in gene content and response to diet. Microbiome *6*, 72. https://doi.org/10.1186/s40168-018-0450-3.

59. Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D., Caporaso, J.G., Knights, D., Clemente, J.C., Nakielny, S., et al. (2013). Cohabiting family members share microbiota with one another and with their dogs. eLife *2*, e00458. https://doi.org/10.7554/eLife.00458.

60. Amend, B., Harland, N., Knoll, J., Stenzl, A., and Aicher, W.K. (2021). Large Animal Models for Investigating Cell Therapies of Stress Urinary Incontinence. Int J Mol Sci *22*, 6092. https://doi.org/10.3390/ijms22116092.

61. Knapp, D.W., Dhawan, D., Ramos-Vara, J.A., Ratliff, T.L., Cresswell, G.M., Utturkar, S., Sommer, B.C., Fulkerson, C.M., and Hahn, N.M. (2020). Naturally-Occurring Invasive Urothelial Carcinoma in Dogs, a Unique Model to Drive Advances in Managing Muscle Invasive Bladder Cancer in Humans. Front. Oncol. *9*, 1493. https://doi.org/10.3389/fonc.2019.01493.

62. Johnson, J.R., Clabots, C., and Kuskowski, M.A. (2008). Multiple-Host Sharing, Long-Term Persistence, and Virulence of Escherichia coli Clones from Human and Animal Household Members. J Clin Microbiol *46*, 4078–4082. https://doi.org/10.1128/JCM.00980-08.

63. de Brot, S., Robinson, B.D., Scase, T., Grau-Roma, L., Wilkinson, E., Boorjian, S.A., Gardner, D., and Mongan, N.P. (2018). The dog as an animal model for bladder and urethral urothelial carcinoma: Comparative epidemiology and histology. Oncology Letters *16*, 1641–1649. https://doi.org/10.3892/ol.2018.8837.

64. Knapp, D.W., Dhawan, D., Ruple, A., Cooper, B.R., Zhang, M., Liu, D., Ramos-Vara, J.A., Bonney, P.L., Fourez, L.M., Enstrom, A.W., et al. (2024). Association between cigarette smoke exposure and urinary bladder cancer in Scottish terriers in a cohort study. The Veterinary Journal *303*, 106044. https://doi.org/10.1016/j.tvjl.2023.106044.

65. Jubber, I., Ong, S., Bukavina, L., Black, P.C., Compérat, E., Kamat, A.M., Kiemeney, L., Lawrentschuk, N., Lerner, S.P., Meeks, J.J., et al. (2023). Epidemiology of Bladder Cancer in 2023: A Systematic Review of Risk Factors. European Urology *84*, 176–190. https://doi.org/10.1016/j.eururo.2023.03.029.

66. Kobayashi, T., Owczarek, T.B., McKiernan, J.M., and Abate-Shen, C. (2015). Modeling bladder cancer in mice: opportunities and challenges. Nat Rev Cancer *15*, 42–54. https://doi.org/10.1038/nrc3858.

67. Sepich-Poore, G.D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J.A., and Knight, R. (2021). The microbiome and human cancer. Science *371*, eabc4552. https://doi.org/10.1126/science.abc4552.

68. Rubinstein, M.R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y.W. (2013). Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe *14*, 195–206. https://doi.org/10.1016/j.chom.2013.07.012.

69. Baruch, E.N., Youngster, I., Ben-Betzalel, G., Ortenberg, R., Lahat, A., Katz, L., Adler, K., Dick-Necula, D., Raskin, S., Bloch, N., et al. (2021). Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. Science *371*, 602–609. https://doi.org/10.1126/science.abb5920.

70. Dizman, N., Meza, L., Bergerot, P., Alcantara, M., Dorff, T., Lyou, Y., Frankel, P., Cui, Y., Mira, V., Llamas, M., et al. (2022). Nivolumab plus ipilimumab with or without live bacterial supplementation in metastatic renal cell carcinoma: a randomized phase 1 trial. Nat Med *28*, 704–712. https://doi.org/10.1038/s41591-022-01694-6.

71. Chen, Y.E., Bousbaine, D., Veinbachs, A., Atabakhsh, K., Dimas, A., Yu, V.K., Zhao, A., Enright, N.J., Nagashima, K., Belkaid, Y., et al. (2023). Engineered skin bacteria induce antitumor T cell responses against melanoma. Science *380*, 203–210. https://doi.org/10.1126/science.abp9563.

72. Bukavina, L., Ginwala, R., Eltoukhi, M., Sindhani, M., Prunty, M., Geynisman, D.M., Ghatalia, P., Valentine, H., Calaway, A., Correa, A.F., et al. (2024). Role of Gut Microbiome in Neoadjuvant Chemotherapy Response in Urothelial Carcinoma: A Multi-Institutional Prospective Cohort Evaluation. Cancer Research Communications. https://doi.org/10.1158/2767-9764.CRC-23-0479.

73. Chipollini, J., Wright, J.R., Nwanosike, H., Kepler, C.Y., Batai, K., Lee, B.R., Spiess, P.E., Stewart, D.B., and Lamendella, R. (2020). Characterization of urinary microbiome in patients with bladder cancer: Results from a single-institution, feasibility study. Urologic Oncology: Seminars and Original Investigations *38*, 615–621. https://doi.org/10.1016/j.urolonc.2020.04.014.

74. Bučević Popović, V., Šitum, M., Chow, C.-E.T., Chan, L.S., Roje, B., and Terzić, J. (2018). The urinary microbiome associated with bladder cancer. Sci Rep *8*, 12157. https://doi.org/10.1038/s41598-018-29054-w.

75. Bi, H., Tian, Y., Song, C., Li, J., Liu, T., Chen, Z., Chen, C., Huang, Y., and Zhang, Y. (2019). Urinary microbiota – a potential biomarker and therapeutic target for bladder cancer. Journal of Medical Microbiology *68*, 1471–1478. https://doi.org/10.1099/jmm.0.001058.

76. Hussein, A.A., Elsayed, A.S., Durrani, M., Jing, Z., Iqbal, U., Gomez, E.C., Singh, P.K., Liu, S., Smith, G., Tang, L., et al. (2021). Investigating the association between the urinary microbiome and bladder cancer: An exploratory study. Urologic Oncology: Seminars and Original Investigations *39*, 370.e9-370.e19. https://doi.org/10.1016/j.urolonc.2020.12.011.

77. Liu, F., Liu, A., Lu, X., Zhang, Z., Xue, Y., Xu, J., Zeng, S., Xiong, Q., Tan, H., He, X., et al. (2019). Dysbiosis signatures of the microbial profile in tissue from bladder cancer. Cancer Med *8*, 6904–6914. https://doi.org/10.1002/cam4.2419.

78. Mansour, B., Monyók, Á., Makra, N., Gajdács, M., Vadnay, I., Ligeti, B., Juhász, J., Szabó, D., and Ostorházi, E. (2020). Bladder cancer-related microbiota: examining differences in urine and tissue samples. Sci Rep *10*, 11042. https://doi.org/10.1038/s41598-020-67443-2.

79. Pederzoli, F., Ferrarese, R., Amato, V., Locatelli, I., Alchera, E., Lucianò, R., Nebuloni, M., Briganti, A., Gallina, A., Colombo, R., et al. (2020). Sex-specific Alterations in the Urinary and Tissue Microbiome in Therapy-naïve Urothelial Bladder Cancer Patients. European Urology Oncology *3*, 784–788. https://doi.org/10.1016/j.euo.2020.04.002.

80. Zeng, J., Zhang, G., Chen, C., Li, K., Wen, Y., Zhao, J., and Wu, P. (2020). Alterations in Urobiome in Patients With Bladder Cancer and Implications for Clinical Outcome: A Single-Institution Study. Front Cell Infect Microbiol *10*, 555508. https://doi.org/10.3389/fcimb.2020.555508.

81. Chen, C., Huang, Z., Huang, P., Li, K., Zeng, J., Wen, Y., Li, B., Zhao, J., and Wu, P. (2022). Urogenital Microbiota:Potentially Important Determinant of PD-L1 Expression in Male Patients with Non-muscle Invasive Bladder Cancer. BMC Microbiol *22*, 7. https://doi.org/10.1186/s12866-021-02407-8.

82. Oresta, B., Braga, D., Lazzeri, M., Frego, N., Saita, A., Faccani, C., Fasulo, V., Colombo, P., Guazzoni, G., Hurle, R., et al. (2021). The Microbiome of Catheter Collected Urine in Males with Bladder Cancer According to Disease Stage. The Journal of Urology. https://doi.org/10.1097/JU.0000000000001336.

83. Mai, G., Chen, L., Li, R., Liu, Q., Zhang, H., and Ma, Y. (2019). Common Core Bacterial Biomarkers of Bladder Cancer Based on Multiple Datasets. Biomed Res Int *2019*, 4824909. https://doi.org/10.1155/2019/4824909.

84. Chorbińska, J., Krajewski, W., Nowak, Ł., Bardowska, K., Żebrowska-Różańska, P., Łaczmański, Ł., Pacyga-Prus, K., Górska, S., Małkiewicz, B., and Szydełko, T. (2023). Is the Urinary and Gut Microbiome Associated With Bladder Cancer? Clin Med Insights Oncol *17*, 11795549231206796. https://doi.org/10.1177/11795549231206796.

85. Wu, P., Zhang, G., Zhao, J., Chen, J., Chen, Y., Huang, W., Zhong, J., and Zeng, J. (2018). Profiling the Urinary Microbiota in Male Patients With Bladder Cancer in China. Front Cell Infect Microbiol *8*, 167. https://doi.org/10.3389/fcimb.2018.00167.

86. Boada, L.D., Henríquez-Hernández, L.A., Navarro, P., Zumbado, M., Almeida-González, M., Camacho, M., Álvarez-León, E.E., Valencia-Santana, J.A., and Luzardo, O.P. (2015). Exposure to polycyclic aromatic hydrocarbons (PAHs) and bladder cancer: evaluation from a gene-environment perspective in a hospital-based case-control study in the Canary Islands (Spain). Int J Occup Environ Health *21*, 23–30. https://doi.org/10.1179/2049396714Y.0000000085.

87. Letašiová, S., Medveďová, A., Šovčíková, A., Dušinská, M., Volkovová, K., Mosoiu, C., and Bartonová, A. (2012). Bladder cancer, a review of the environmental risk factors. Environ Health *11*, S11. https://doi.org/10.1186/1476-069X-11-S1-S11.

88. Sowada, J., Schmalenberger, A., Ebner, I., Luch, A., and Tralau, T. (2014). Degradation of benzo[a]pyrene by bacterial isolates from human skin. FEMS Microbiology Ecology *88*, 129–139. https://doi.org/10.1111/1574-6941.12276.

89. Sowada, J., Lemoine, L., Schön, K., Hutzler, C., Luch, A., and Tralau, T. (2017). Toxification of polycyclic aromatic hydrocarbons by commensal bacteria from human skin. Arch Toxicol *91*, 2331–2341. https://doi.org/10.1007/s00204-017-1964-3.

90. Van De Wiele, T., Vanhaecke, L., Boeckaert, C., Peru, K., Headley, J., Verstraete, W., and Siciliano, S. (2005). Human Colon Microbiota Transform Polycyclic Aromatic Hydrocarbons to Estrogenic Metabolites. Environ Health Perspect *113*, 6–10. https://doi.org/10.1289/ehp.7259.

91. Goto, T., and Miyamoto, H. (2021). The Role of Estrogen Receptors in Urothelial Cancer. Front Endocrinol (Lausanne) *12*, 643870. https://doi.org/10.3389/fendo.2021.643870.

92. Berkel, C., and Cacan, E. (2021). Estrogen- and estrogen receptor (ER)-mediated cisplatin chemoresistance in cancer. Life Sciences *286*, 120029. https://doi.org/10.1016/j.lfs.2021.120029.

93. Bukowska, B., Mokra, K., and Michałowicz, J. (2022). Benzo[a]pyrene—Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. International Journal of Molecular Sciences *23*, 6348. https://doi.org/10.3390/ijms23116348.

94. Sharon, B.M., Arute, A.P., Nguyen, A., Tiwari, S., Reddy Bonthu, S.S., Hulyalkar, N.V., Neugent, M.L., Palacios Araya, D., Dillon, N.A., Zimmern, P.E., et al. (2023). Genetic and functional enrichments associated with *Enterococcus faecalis* isolated from the urinary tract. mBio *14*, e02515-23. https://doi.org/10.1128/mbio.02515-23.

95. Yoo, J.-J., Song, J.S., Kim, W.B., Yun, J., Shin, H.B., Jang, M.-A., Ryu, C.B., Kim, S.S., Chung, J.C., Kuk, J.C., et al. (2022). Gardnerella vaginalis in Recurrent Urinary Tract Infection Is Associated with Dysbiosis of the Bladder Microbiome. J Clin Med *11*, 2295. https://doi.org/10.3390/jcm11092295.

96. Vermeulen, S.H., Hanum, N., Grotenhuis, A.J., Castaño-Vinyals, G., van der Heijden, A.G., Aben, K.K., Mysorekar, I.U., and Kiemeney, L.A. (2015). Recurrent urinary tract infection and risk of bladder cancer in the Nijmegen bladder cancer study. Br J Cancer *112*, 594–600. https://doi.org/10.1038/bjc.2014.601.

97. Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L.T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., et al. (2020). The human tumor microbiome is composed of tumor type–specific intracellular bacteria. Science *368*, 973–980. https://doi.org/10.1126/science.aay9189.

98. Walker, A.W., and Hoyles, L. (2023). Human microbiome myths and misconceptions. Nat Microbiol *8*, 1392–1396. https://doi.org/10.1038/s41564-023-01426-7.

99. Chrisman, B., He, C., Jung, J.-Y., Stockham, N., Paskov, K., Washington, P., and Wall, D.P. (2022). The human "contaminome": bacterial, viral, and computational contamination in whole genome sequences from 1000 families. Sci Rep *12*, 9863. https://doi.org/10.1038/s41598-022-13269-z.

100. Lusk, R.W. (2014). Diverse and Widespread Contamination Evident in the Unmapped Depths of High Throughput Sequencing Data. PLOS ONE *9*, e110808. https://doi.org/10.1371/journal.pone.0110808.

101. Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. Sci Transl Med *6*, 237ra65. https://doi.org/10.1126/scitranslmed.3008599.

102. Lauder, A.P., Roche, A.M., Sherrill-Mix, S., Bailey, A., Laughlin, A.L., Bittinger, K., Leite, R., Elovitz, M.A., Parry, S., and Bushman, F.D. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome *4*, 29. https://doi.org/10.1186/s40168-016-0172-3.

103. Leiby, J.S., McCormick, K., Sherrill-Mix, S., Clarke, E.L., Kessler, L.R., Taylor, L.J., Hofstaedter, C.E., Roche, A.M., Mattei, L.M., Bittinger, K., et al. (2018). Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. Microbiome *6*, 196. https://doi.org/10.1186/s40168-018-0575-4.

104. Poore, G.D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., Kosciolek, T., Janssen, S., Metcalf, J., Song, S.J., et al. (2020). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature *579*, 567–574. https://doi.org/10.1038/s41586-020-2095-1.

105. Gihawi, A., Ge, Y., Lu, J., Puiu, D., Xu, A., Cooper, C.S., Brewer, D.S., Pertea, M., and Salzberg, S.L. (2023). Major data analysis errors invalidate cancer microbiome findings. mBio *14*, e01607-23. https://doi.org/10.1128/mbio.01607-23.

106. Melgarejo, T., Oakley, B.B., Krumbeck, J.A., Tang, S., Krantz, A., and Linde, A. (2021). Assessment of bacterial and fungal populations in urine from clinically healthy dogs using next-generation sequencing. Journal of Veterinary Internal Medicine *35*, 1416–1426. https://doi.org/10.1111/jvim.16104.

107. Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W.T., Wolfe, A.J., Braun, J., and McWeeney, S. (2019). Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. mSystems *4*, e00290-19. https://doi.org/10.1128/mSystems.00290-19.

108. Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., and Kelley, S.T. (2011). Bayesian community-wide culture-independent microbial source tracking. Nat Methods *8*, 761–763. https://doi.org/10.1038/nmeth.1650.

109. Tito, E., Ahmad, A., Gongolli, J., Issack, W., and Johnson, A. Sphingomonas paucimobilis Bacteremia in a Patient With Retropharyngeal Abscess. Cureus *14*, e25407. https://doi.org/10.7759/cureus.25407.

110. Karstens, L., Siddiqui, N.Y., Zaza, T., Barstad, A., Amundsen, C.L., and Sysoeva, T.A. (2021). Benchmarking DNA isolation kits used in analyses of the urinary microbiome. Sci Rep *11*, 6186. https://doi.org/10.1038/s41598-021-85482-1.

111. Greathouse, K.L., Sinha, R., and Vogtmann, E. (2019). DNA extraction for human microbiome studies: the issue of standardization. Genome Biol *20*, 212. https://doi.org/10.1186/s13059-019-1843-8.

112. Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A.A., Ren, B., Amir, A., Schwager, E., Crabtree, J., Ma, S., Abnet, C.C., et al. (2017). Assessment of variation in microbial

community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol *35*, 1077–1086. https://doi.org/10.1038/nbt.3981.

113. Knudsen, B.E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F.M., and Pamp, S.J. (2016). Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. mSystems *1*, 10.1128/msystems.00095-16. https://doi.org/10.1128/msystems.00095-16.

114. Stinson, L.F., Keelan, J.A., and Payne, M.S. (2018). Comparison of Meconium DNA Extraction Methods for Use in Microbiome Studies. Front. Microbiol. *9*. https://doi.org/10.3389/fmicb.2018.00270.

115. Vesty, A., Biswas, K., Taylor, M.W., Gear, K., and Douglas, R.G. (2017). Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. PLOS ONE *12*, e0169877. https://doi.org/10.1371/journal.pone.0169877.

116. Walden, C., Carbonero, F., and Zhang, W. (2017). Assessing impacts of DNA extraction methods on next generation sequencing of water and wastewater samples. Journal of Microbiological Methods *141*, 10–16. https://doi.org/10.1016/j.mimet.2017.07.007.

117. QIAamp BiOstic Bacteremia DNA Kit https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/qiaamp-biostic-bacteremia-dna-kit.

118. Nadkarni, M.A., Martin, F.E., Jacques, N.A., and Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology *148*, 257–266. https://doi.org/10.1099/00221287-148-1-257.

119. Yoo, J.-J., Shin, H.B., Song, J.S., Kim, M., Yun, J., Kim, Z., Lee, Y.M., Lee, S.W., Lee, K.W., Kim, W. bin, et al. (2021). Urinary Microbiome Characteristics in Female Patients with Acute Uncomplicated Cystitis and Recurrent Cystitis. J Clin Med *10*, 1097. https://doi.org/10.3390/jcm10051097.

120. Kim, M., Parrish, R.C., Shah, V.S., Ross, M., Cormier, J., Baig, A., Huang, C.-Y., Brenner, L., Neuringer, I., Whiteson, K., et al. (2024). Host DNA depletion on frozen human respiratory samples enables successful metagenomic sequencing for microbiome studies. Res Sq, rs.3.rs-3638876. https://doi.org/10.21203/rs.3.rs-3638876/v1.

121. Montassier, E., Kitsios, G.D., Radder, J.E., Le Bastard, Q., Kelly, B.J., Panzer, A., Lynch, S.V., Calfee, C.S., Dickson, R.P., and Roquilly, A. (2023). Robust airway microbiome signatures in acute respiratory failure and hospital-acquired pneumonia. Nat Med *29*, 2793–2804. https://doi.org/10.1038/s41591-023-02617-9.

122. Duarte, V. da S., and Porcellato, D. (2023). Host DNA depletion methods and genome-centric metagenomics of bovine hindmilk microbiome. mSphere *0*, e00470-23. https://doi.org/10.1128/msphere.00470-23.

123. Yap, M., Feehily, C., Walsh, C.J., Fenelon, M., Murphy, E.F., McAuliffe, F.M., van Sinderen, D., O'Toole, P.W., O'Sullivan, O., and Cotter, P.D. (2020). Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. Sci Rep *10*, 21665. https://doi.org/10.1038/s41598-020-78773-6.

124. Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome *6*, 42. https://doi.org/10.1186/s40168-018-0426-3.

125. Ong, C.T., Boe-Hansen, G., Ross, E.M., Blackall, P.J., Turni, C., Hayes, B.J., and Tabor, A.E. (2022). Evaluation of Host Depletion and Extraction Methods for Shotgun Metagenomic Analysis of Bovine Vaginal Samples. Microbiology Spectrum *10*, e00412-21. https://doi.org/10.1128/spectrum.00412-21.

126. Rajar, P., Dhariwal, A., Salvadori, G., Junges, R., Åmdal, H.A., Berild, D., Fugelseth, D., Saugstad, O.D., Lausten-Thomsen, U., Greisen, G., et al. (2022). Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies. Frontiers in Microbiology *13*.

127. Ahannach, S., Delanghe, L., Spacova, I., Wittouck, S., Van Beeck, W., De Boeck, I., and Lebeer, S. (2021). Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. iScience *24*, 103306. https://doi.org/10.1016/j.isci.2021.103306.

128. Cheng, Y., Chen, Z., Gawthorne, J.A., Mukerjee, C., Varettas, K., Mansfield, K.J., Schembri, M.A., and Moore, K.H. (2016). Detection of intracellular bacteria in exfoliated urothelial cells from women with urge incontinence. Pathogens and Disease *74*, ftw067. https://doi.org/10.1093/femspd/ftw067.

129. Flores-Mireles, A.L., Walker, J.N., Caparon, M., and Hultgren, S.J. (2015). Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol *13*, 269–284. https://doi.org/10.1038/nrmicro3432.

130. Urinalysis and Body Fluids - F.A. Davis Company https://www.fadavis.com/product/urinalysis-and-body-fluids-strasinger-6.

131. Chivian, D., Jungbluth, S.P., Dehal, P.S., Wood-Charlson, E.M., Canon, R.S., Allen, B.H., Clark, M.M., Gu, T., Land, M.L., Price, G.A., et al. (2023). Metagenome-assembled genome extraction and analysis from microbiomes using KBase. Nat Protoc *18*, 208–238. https://doi.org/10.1038/s41596-022-00747-x.

132. Pereira-Marques, J., Hout, A., Ferreira, R.M., Weber, M., Pinto-Ribeiro, I., van Doorn, L.-J., Knetsch, C.W., and Figueiredo, C. (2019). Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Front. Microbiol. *10*. https://doi.org/10.3389/fmicb.2019.01277.

133. Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., McCracken, C., Giglio, M.G., et al. (2017). Strains, functions and dynamics

in the expanded Human Microbiome Project. Nature *550*, 61–66. https://doi.org/10.1038/nature23889.

134. De Filippis, F., Valentino, V., Sequino, G., Borriello, G., Riccardi, M.G., Pierri, B., Cerino, P., Pizzolante, A., Pasolli, E., Esposito, M., et al. (2024). Exposure to environmental pollutants selects for xenobiotic-degrading functions in the human gut microbiome. Nat Commun *15*, 4482. https://doi.org/10.1038/s41467-024-48739-7.

135. Kachroo, N., Monga, M., and Miller, A.W. (2022). Comparative functional analysis of the urinary tract microbiome for individuals with or without calcium oxalate calculi. Urolithiasis *50*, 303–317. https://doi.org/10.1007/s00240-022-01314-5.

136. Wu-Woods, N.J., Barlow, J.T., Trigodet, F., Shaw, D.G., Romano, A.E., Jabri, B., Eren, A.M., and Ismagilov, R.F. (2023). Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples. Nat Methods *20*, 1672–1682. https://doi.org/10.1038/s41592-023-02025-4.

137. Ganda, E., Beck, K.L., Haiminen, N., Silverman, J.D., Kawas, B., Cronk, B.D., Anderson, R.R., Goodman, L.B., and Wiedmann, M. (2021). DNA Extraction and Host Depletion Methods Significantly Impact and Potentially Bias Bacterial Detection in a Biological Fluid. mSystems *6*. https://doi.org/10.1128/mSystems.00619-21.

138. Amar, Y., Lagkouvardos, I., Silva, R.L., Ishola, O.A., Foesel, B.U., Kublik, S., Schöler, A., Niedermeier, S., Bleuel, R., Zink, A., et al. (2021). Pre-digest of unprotected DNA by Benzonase improves the representation of living skin bacteria and efficiently depletes host DNA. Microbiome *9*, 123. https://doi.org/10.1186/s40168-021-01067-0.

139. NEBNext® Microbiome DNA Enrichment Kit | NEB https://www.neb.com/en-us/products/e2612-nebnext-microbiome-dna-enrichment-kit.

140. Nelson, M.T., Pope, C.E., Marsh, R.L., Wolter, D.J., Weiss, E.J., Hager, K.R., Vo, A.T., Brittnacher, M.J., Radey, M.C., Hayden, H.S., et al. (2019). Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. Cell Reports *26*, 2227-2240.e5. https://doi.org/10.1016/j.celrep.2019.01.091.

141. Stern, J.M., Moazami, S., Qiu, Y., Kurland, I., Chen, Z., Agalliu, I., Burk, R., and Davies, K.P. (2016). Evidence for a distinct gut microbiome in kidney stone formers compared to non-stone formers. Urolithiasis *44*, 399–407. https://doi.org/10.1007/s00240-016-0882-9.

142. Hilt, E.E., McKinley, K., Pearce, M.M., Rosenfeld, A.B., Zilliox, M.J., Mueller, E.R., Brubaker, L., Gai, X., Wolfe, A.J., and Schreckenberger, P.C. (2014). Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. J Clin Microbiol *52*, 871–876. https://doi.org/10.1128/JCM.02876-13.

143. Walker, S.P., Barrett, M., Hogan, G., Flores Bueso, Y., Claesson, M.J., and Tangney, M. (2020). Non-specific amplification of human DNA is a major challenge for 16S rRNA gene sequence analysis. Sci Rep *10*, 16356. https://doi.org/10.1038/s41598-020-73403-7.

144. Kim, M., Ii, R.C.P., Shah, V.S., Ross, M., Cormier, J., Baig, A., Huang, C.-Y., Brenner, L., Neuringer, I., Whiteson, K., et al. (2024). Host DNA depletion on frozen human respiratory samples enables successful metagenomic sequencing for microbiome studies. Preprint, https://doi.org/10.21203/rs.3.rs-3638876/v1 https://doi.org/10.21203/rs.3.rs-3638876/v1.

145. Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research *48*, 8883–8900. https://doi.org/10.1093/nar/gkaa621.

146. Marques, C., Belas, A., Aboim, C., Trigueiro, G., Cavaco-Silva, P., Gama, L.T., and Pomba, C. (2019). Clonal relatedness of Proteus mirabilis strains causing urinary tract infections in companion animals and humans. Veterinary Microbiology *228*, 77–82. https://doi.org/10.1016/j.vetmic.2018.10.015.

147. Johnson, J.R., O'Bryan, T.T., Low, D.A., Ling, G., Delavari, P., Fasching, C., Russo, T.A., Carlino, U., and Stell, A.L. (2000). Evidence of Commonality between Canine and Human Extraintestinal Pathogenic Escherichia coli Strains That Express papG Allele III. Infect Immun *68*, 3327–3336.

148. McGlynn, A., Mrofchak, R., Madan, R., Madden, C., Jahid, M.J., Mollenkopf, D., Wittum, T., Justice, S.S., Rudinsky, A., Hokamp, J., et al. (2023). Longitudinal examination of urine pH, specific gravity, protein, culture, and antimicrobial resistance profiles in healthy dogs. Journal of Veterinary Internal Medicine *37*, 2219–2229. https://doi.org/10.1111/jvim.16860.

149. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.

150. Benjamini, Y., Krieger, A.M., and Yekutieli, D. (2006). Adaptive Linear Step-up Procedures That Control the False Discovery Rate. Biometrika *93*, 491–507.

151. Hassan, B.B., Altstadt, L.A., Dirksen, W.P., Elshafae, S.M., and Rosol, T.J. (2020). Canine Thyroid Cancer: Molecular Characterization and Cell Line Growth in Nude Mice. Vet Pathol *57*, 227–240. https://doi.org/10.1177/0300985819901120.

152. Szóstak, N., Szymanek, A., Havránek, J., Tomela, K., Rakoczy, M., Samelak-Czajka, A., Schmidt, M., Figlerowicz, M., Majta, J., Milanowska-Zabel, K., et al. (2022). The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. Sci Rep *12*, 8470. https://doi.org/10.1038/s41598-022-12037-3.

153. Ohio Supercomputer Center (1987). Columbus, OH: Ohio Supercomputer Center. http://osc.edu/ark:/19495/f5s1ph73 http://osc.edu/ark:/19495/f5s1ph73.

154. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

155. Woodcroft, B.J. (2023). CoverM.

156. Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A.M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol *41*, 1633–1644. https://doi.org/10.1038/s41587-023-01688-w.

157. Woodcroft, B.J. (2023). wwood/singlem.

158. Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

159. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

160. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome *6*, 158. https://doi.org/10.1186/s40168-018-0541-1.

161. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*. https://doi.org/10.7717/peerj.7359.

162. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605–607. https://doi.org/10.1093/bioinformatics/btv638.

163. Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nat Methods *11*, 1144–1146. https://doi.org/10.1038/nmeth.3103.

164. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

165. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res *25*, 1043–1055. https://doi.org/10.1101/gr.186072.114.

166. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

167. McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

168. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2001). vegan: Community Ecology Package. https://doi.org/10.32614/CRAN.package.vegan https://doi.org/10.32614/CRAN.package.vegan.

169. Tidyverse https://www.tidyverse.org/.

170. Huang, Y., Li, L., Yin, X., and Zhang, T. (2023). Polycyclic aromatic hydrocarbon (PAH) biodegradation capacity revealed by a genome-function relationship approach. Environmental Microbiome *18*, 39. https://doi.org/10.1186/s40793-023-00497-7.

171. Khot, V., Zorz, J., Gittins, D.A., Chakraborty, A., Bell, E., Bautista, M.A., Paquette, A.J., Hawley, A.K., Novotnik, B., Hubert, C.R.J., et al. (2022). CANT-HYD: A Curated Database of Phylogeny-Derived Hidden Markov Models for Annotation of Marker Genes Involved in Hydrocarbon Degradation. Front. Microbiol. *12*. https://doi.org/10.3389/fmicb.2021.764058.

172. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res *39*, W29–W37. https://doi.org/10.1093/nar/gkr367.

173. Pelucchi, C., Bosetti, C., Negri, E., Malvezzi, M., and La Vecchia, C. (2006). Mechanisms of Disease: the epidemiology of bladder cancer. Nat Rev Urol *3*, 327–340. https://doi.org/10.1038/ncpuro0510.

174. Bukavina, L., Isali, I., Ginwala, R., Sindhani, M., Calaway, A., Magee, D., Miron, B., Correa, A., Kutikov, A., Zibelman, M., et al. (2022). Global Meta-Analysis of Urine Microbiome: Colonization of PAH-degrading bacteria among bladder cancer patients (In Review) https://doi.org/10.21203/rs.3.rs-2003199/v1.

175. Li, J., Li, K., Li, H., Wang, X., Wang, W., Wang, K., and Ge, M. (2022). Long-chain alkanes in the atmosphere: A review. Journal of Environmental Sciences *114*, 37–52. https://doi.org/10.1016/j.jes.2021.07.021.

176. Chen, J., Rodopoulou, S., Strak, M., de Hoogh, K., Taj, T., Poulsen, A.H., Andersen, Z.J., Bellander, T., Brandt, J., Zitt, E., et al. (2022). Long-term exposure to ambient air pollution and bladder cancer incidence in a pooled European cohort: the ELAPSE project. Br J Cancer *126*, 1499–1507. https://doi.org/10.1038/s41416-022-01735-4.

177. Lima, A.L.C., Farrington, J.W., and Reddy, C.M. (2005). Combustion-Derived Polycyclic Aromatic Hydrocarbons in the Environment—A Review. Environmental Forensics *6*, 109–131. https://doi.org/10.1080/15275920590952739.

178. Marchukov, D., Li, J., Juillerat, P., Misselwitz, B., and Yilmaz, B. (2023). Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. Front. Genet. *14*. https://doi.org/10.3389/fgene.2023.1184473.

179. Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., Pletz, M.W., Ehricht, R., and Brandt, C. (2022). Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. Sci Rep *12*, 4000. https://doi.org/10.1038/s41598-022-08003-8.

180. Arikawa, K., Ide, K., Kogawa, M., Saeki, T., Yoda, T., Endoh, T., Matsuhashi, A., Takeyama, H., and Hosokawa, M. (2021). Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics. Microbiome *9*, 202. https://doi.org/10.1186/s40168-021-01152-4.

181. Bottone, E.J. (2010). Bacillus cereus, a Volatile Human Pathogen. Clin Microbiol Rev *23*, 382–398. https://doi.org/10.1128/CMR.00073-09.

182. Mobley, H.L., and Hausinger, R.P. (1989). Microbial ureases: significance, regulation, and molecular characterization. Microbiol Rev *53*, 85–108.

183. Glasius, M., Thomsen, D., Wang, K., Iversen, L.S., Duan, J., and Huang, R.-J. (2022). Chemical characteristics and sources of organosulfates, organosulfonates, and carboxylic acids in aerosols in urban Xi'an, Northwest China. Science of The Total Environment *810*, 151187. https://doi.org/10.1016/j.scitotenv.2021.151187.

184. Llambrich, M., Brezmes, J., and Cumeras, R. (2022). The untargeted urine volatilome for biomedical applications: methodology and volatilome database. Biological Procedures Online *24*, 20. https://doi.org/10.1186/s12575-022-00184-w.

185. Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P., et al. (2013). The Human Urine Metabolome. PLOS ONE *8*, e73076. https://doi.org/10.1371/journal.pone.0073076.