



Forming Consensus To Advance Urobiome Research

Linda Brubaker,^a Jean-Philippe F. Gourdine,^b Nazema Y. Siddiqui,^c Amanda Holland,^d Thomas Halverson,^e Roberto Limeria,^f David Pride,^{g,h} Lenore Ackerman,ⁱ Catherine S. Forster,^j Kristin M. Jacobs,^k Krystal J. Thomas-White,^l Catherine Putonti,^{e,m} Qunfeng Dong,^{n,o} Michael Weinstein,^{p,q} Emily S. Lukacz,^a Lisa Karstens,^r Alan J. Wolfe^{e,f}

^aDepartment of Obstetrics, Gynecology and Reproductive Sciences, Division of Female Pelvic Medicine and Reconstructive Surgery, University of California, San Diego, San Diego, California, USA

^bOregon Clinical & Translational Research Institute (OCTRI), Oregon Health & Science University, Portland, Oregon, USA

^cDepartment of Obstetrics and Gynecology, Division of Female Pelvic Medicine and Reconstructive Surgery, Duke University, Durham, North Carolina, USA

^dDepartment of Obstetrics and Gynecology, Oregon Health & Science University, Portland, Oregon, USA

^eDepartment of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois, USA

^fLoyola Genomics Facility, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois, USA

^gDepartment of Pathology, University of California, San Diego, San Diego, California, USA

^hDepartment of Medicine, University of California, San Diego, San Diego, California, USA

ⁱDepartment of Urology, Division of Female Pelvic Medicine and Reconstructive Surgery, University of California, Los Angeles, Los Angeles, California, USA

^jDepartment of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

^kDepartment of Obstetrics and Gynecology, Division of Female Pelvic Medicine and Reconstructive Surgery, Rush University, Chicago, Illinois, USA

^lDepartment of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California, USA

^mDepartment of Biology, Bioinformatics Program, Loyola University Chicago, Chicago, Illinois, USA

ⁿDepartment of Medicine, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois, USA

^oCenter for Biomedical Informatics, Loyola University Chicago, Chicago, Illinois, USA

^pMolecular, Cell, and Developmental Biology, University of California, Los Angeles, Los Angeles, California, USA

^qZymo Research, Irvine, California, USA

^rDepartment of Clinical Epidemiology and Medical Informatics, Oregon Health & Science University, Portland, Oregon, USA

ABSTRACT Urobiome research has the potential to advance the understanding of a wide range of diseases, including lower urinary tract symptoms and kidney disease. Many scientific areas have benefited from early research method consensus to facilitate the greater, common good. This consensus document, developed by a group of expert investigators currently engaged in urobiome research (UROBIOME 2020 conference participants), aims to promote standardization and advances in this field by the adoption of common core research practices. We propose a standardized nomenclature as well as considerations for specimen collection, preservation, storage, and processing. Best practices for urobiome study design include our proposal for standard metadata elements as part of core metadata collection. Although it is impractical to follow fixed analytical procedures when analyzing urobiome data, we propose guidelines to document and report data originating from urobiome studies. We offer this first consensus document with every expectation of subsequent revision as our field progresses.

KEYWORDS consensus, guideline, human microbiome, research, statement, urinary microbiome, urobiome


Since the discovery of the human urinary microbiome (urobiome), urobiome research has been impacted by inconsistent sampling conditions, technical conditions, and participant-related factors (1). The number of investigators currently working in urobiome research is still relatively small; however, the rapid growth of the field and the variety of approaches used to date have highlighted an urgent need for consensus on optimal strategies for the scientific investigation of the urobiome. A group of expert

Citation Brubaker L, Gourdine J-PF, Siddiqui NY, Holland A, Halverson T, Limeria R, Pride D, Ackerman L, Forster CS, Jacobs KM, Thomas-White KJ, Putonti C, Dong Q, Weinstein M, Lukacz ES, Karstens L, Wolfe AJ. 2021. Forming consensus to advance urobiome research. *mSystems* 6:e01371-20. <https://doi.org/10.1128/mSystems.01371-20>.

Editor Jian Xu, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences

Copyright © 2021 Brubaker et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Linda Brubaker, lbrubaker@health.ucsd.edu.

 Faster advances with shared approaches - see the consensus statement of the UROBIOME 2020 conference. Much to learn if we all pull together. #urobiome #urinary microbiome @brubaker1030

Published 20 July 2021

- ❖ Use appropriate nomenclature to describe the urine specimen: “Urogenital” sample for voided urine samples versus “Urinary Bladder” for catheterized (transurethral or suprapubic aspirates).
- ❖ Align urine sample technique with research purpose. For studies of bladder urinary microbiota in adult women, we recommend collection of catheterized urine samples when possible. When using voided specimens, limit or control for vulvo-vaginal contamination.
- ❖ Use nucleic acid preservative and maintain cold temperatures while expediting transfer to laboratory.
- ❖ Align sample processing plan with research question (culture ± culture-independent techniques).
- ❖ Minimize batch effects as much as possible (differences in kits and sequencing runs; record information to account for differences in downstream analyses if necessary).
- ❖ Adopt the standard metadata checklist for urobiome studies
- ❖ Collect and describe “required” metadata in all urobiome research
- ❖ Collect and report “desired” metadata when feasible
- ❖ Collaboration with bioinformaticians, biostatisticians and computational scientists is critical during the planning phase
- ❖ No single bioinformatic platform or fixed analytic procedures are recommended
- ❖ Statistical methods tailored to microbiome data should be used
- ❖ Documentation of computational steps, software, and data used is essential.

FIG 1 Key recommendations for urobiome research.

investigators currently engaged in urobiome research gathered to share research progress and exchange ideas at the National Institutes of Health (NIH)-sponsored international UROBIOME conferences in 2019 and 2020. This consensus document, developed by UROBIOME 2020 conference participants and their collaborators, aims to promote standardization and advances in this field by the adoption of common core research practices (Fig. 1).

TERMINOLOGY

Standard terminology for urine specimens is necessary (Fig. 2). Many descriptors, including “bladder,” “urinary,” “urogenital,” and “genitourinary,” have been used, and these terms are often conflated. We propose a standardized nomenclature to explicitly describe the specimen as it relates to the collection method. The preferred, recommended terminology for a voided urine sample is “urogenital sample.” The preferred, recommended terminology for a catheterized urine sample (either transurethral or suprapubic) is “urinary bladder.” Samples obtained by urethral swabs, by urothelial/tissue biopsy, or from the kidney pelvis should be so named.

URINE SPECIMEN COLLECTION

The urine specimen collection method must guide analysis and data interpretation, appropriately recognizing anatomical differences between sexes. Although the microbial biomass increases as the urine moves from the kidney to the bladder, urethra, and external genitalia, the urobiome has a low microbial biomass compared to other human microbial niches. Several studies have provided convincing evidence that the female urobiome includes vulvovaginal microbes (2, 3) when conventional “clean catch” midstream voided urine is used; thus, this type of sample should be referred to as a urogenital sample. A catheterized urine sample minimizes the inclusion of vulvovaginal microbes. When catheterization is not feasible or not desired (due to the potential disturbance of the urobiome itself) or when researchers wish to answer questions concerning the lower urinary tract microbiota, voided urine samples can be collected with a urinal device (i.e., Peezy midstream [Forte Medical]) that decreases microbial abundance and diversity, apparently by decreasing posturethral contamination (4). When multiple samples are collected from the same research participant, the order of collection should be specified. An alternative is to include a periurethral swab to allow

Sample Type	Collection Method	Recommended Terminology	Notes
Urine			
	Catheter or suprapubic aspirate	Urinary Bladder	Indicates that the sample was collected by directly accessing urine from the bladder
	Voided (clean catch, midstream urine, or collected with a device such as a Peezy).	Urogenital	Indicates that the sample includes microbiota in the bladder, urinary tract and potentially genitalia
	Renal Urine Aspirate	Urinary Kidney	Indicates that the sample was directly obtained from the pelvic of the kidney
Tissue			
	Bladder biopsy or scraping	Urothelial/Bladder Tissue	Indicates that the sample is from bladder tissue/cells
	Urethral swab	Urethral	Indicates that the sample is from tissue/cells from the urethral opening.

FIG 2 Recommended terminology for urobiome samples.

the separate detection of genital microbes (5). Separate urethral swabs should be used for studies of the urethral microbiota. There is much less research informing urinary sample collection techniques in men; however, the currently available evidence supports the following conclusions: (i) the microbiome of voided urine most closely resembles that of urethral swabs, and (ii) catheterized urine does not tend to resemble voided urine (6, 7). Therefore, for males, we recommend using the term “urogenital” for voided urine and “urinary bladder” for catheterized urine or suprapubic aspirates.

SPECIMEN PRESERVATION AND STORAGE

Immediately upon procurement, specimens should be cooled on ice or in a 4°C refrigerator and should be received by research personnel within 4 h. To avoid inappropriate microbial growth or degradation of nucleic acids, specimens should be appropriately preserved.

For all culture-based techniques, we recommend the use of BD Vacutainer Plus C&S boric acid sodium borate/formate (“gray top”) tubes (Becton, Dickinson and Company, Franklin Lakes, NJ). These are commonly used for clinical culture and antimicrobial sensitivity testing because they maintain microbial viability for at least 24 h under ambient conditions while inhibiting growth. This 24-h period gives research personnel some flexibility and permits overnight shipping.

For culture-independent analyses, we recommend the addition of AssayAssure (Sierra Molecular Corporation, Princeton, NJ) directly to the sample in a 1:10 ratio. This reagent is designed to inhibit 31 enzyme families known to degrade nucleic acids and thus stabilizes nucleic acids (DNA and RNA) over extended time periods without freezing or refrigeration. Importantly, it does not inhibit the amplification fundamental to PCR-based analyses such as 16S rRNA gene sequencing. It is recommended that the specimen be frozen at –80°C upon receipt. However, a benchmarking study showed that AssayAssure in combination with immediate cooling to 4°C or freezing at –20°C allowed storage for up to 4 days with a minimal impact on alpha diversity (8). Although the AssayAssure product guide states that samples can be maintained for up to 4 days at room temperature, we recommend caution when interpreting data from specimens held in this fashion compared to those immediately cooled in the presence of AssayAssure as different taxa may be recovered under different temperatures (8). We recommend rapid shipment (overnight if possible) on dry ice; however, the 4-day window allows flexibility as long as the samples remain cool. Other nucleic acid preservatives exist (e.g., DNA/RNA Shield [Zymo Research Corporation, Irvine, CA]) and can

be used if AssayAssure is unavailable. As there is no current evidence that either pelleting/freezing bacteria or boric acid will adequately preserve nucleic acid, we recommend that this preservative/storage method should be included as a study limitation when nucleic acid preservatives are not utilized (due to affordability or other reasons). Studies reporting on urobiome findings should explicitly describe the use of preservative and storage conditions.

SAMPLE PROCESSING

Traditional and enhanced culture techniques, as well as culture-independent methods, can be used for microbial detection. Culture techniques facilitate microbial detection and demonstrate that the microbe is alive, allowing subsequent experiments with the microbe itself. Lists of known urinary microbes and their growth conditions have been published (9, 10). Enhanced culture techniques, also known as metaculturomics, move beyond the traditional method described by Kass (11), allowing detection of microbes similar to that achieved with sequencing techniques (9, 10). Several enhanced culture methods have been reported (12, 13), including the expanded quantitative urine culture (EQUC) protocol, which has been used extensively for urobiome studies (9, 10, 13). To account for the very low biomass of catheterized urine specimens, we recommend plating 100 μ l, which allows the detection of 10 CFU/ml. However, smaller volumes (1 or 10 μ l) are recommended to achieve accurate counts of CFU per milliliter for voided urine samples or swabs (e.g., urethral or vaginal). Compared to the standard method, EQUC uses additional growth media (9, 10). The selection of media will depend on the research question, the cohort under study, the sample type, and resource constraints. The use of Columbia CNA (colistin naladixic acid) blood agar plates is critical to detect underlying Gram-positive bacteria that are often overwhelmed by more numerous and faster-growing Gram-negative bacteria such as *Escherichia coli* (9, 10). EQUC uses more atmospheric conditions than the standard method; 5% CO₂ allows the growth of most urinary species, which prefer less oxygen. Anaerobic conditions are used for obligate anaerobes; when possible, we recommend an anaerobic chamber. If a chamber is not available, anaerobic jars can suffice for many but not all anaerobes. Finally, an extended incubation period (48 instead of 24 h) allows for the growth of slow-growing microbes and for the morphological differences between species to develop (8).

For sequencing, investigators should have a complete and detailed workflow (including nucleic acid isolation, library preparation, and sequencing) that aligns with the study hypotheses and bioinformatic analysis. Currently, marker gene (amplicon) sequencing is most commonly used for urobiome investigations. Studies of the bacterial communities rely on a hypervariable region of the 16S rRNA gene, while fungal community surveys target the internal transcribed spacer (ITS) region (14). Whereas amplicon sequencing can be used for taxonomic assignment and to determine relative quantities, shotgun metagenomic sequencing can provide insight into urobiome functionality and can also detect the viral fraction, which lacks a conserved marker gene (15).

Nucleic acid isolation techniques affect sequencing results, with some nucleic acid isolation kits showing biases that could specifically affect urobiome information (15). Enzymatic lysis is generally more reproducible among a range of laboratory environments (16). When establishing an enzymatic lysis protocol within a new laboratory, testing must be performed to determine whether the lysing enzymes contain nucleic acids from their manufacturing process (contaminants known as “the kitome”). Lysozyme and mutanolysin have been shown to contain a minimum amount of kitome contamination while having the best lysis efficiency (16).

Purification methodologies can be done with either silica column or magnetic bead protocols. Silica columns are easy to use; however, as they tend to shear DNA during extraction, they should be used only for short-read sequencing. Magnetic beads are easier to automate and can provide similar yields and purities (17, 18). We recommend

that, whenever possible, all samples from entire projects be sequenced at once on the same machine to minimize technical variations. When that is not possible, we recommend that machines with the most similar chemistries/flow cells be used and that available reagent lot numbers be recorded so that these metadata can be considered during analysis. We recommend running positive-control samples with each batch to identify any differences due to the batch.

For 16S amplicon sequencing and Illumina's paired-end 250-bp chemistry, one must choose between longer sequences that span multiple variable regions of the 16S rRNA gene (e.g., V1-V3) or shorter regions (e.g., V4). Longer regions possess more sequence information for downstream taxonomic assignment. However, sequence read quality diminishes at the ends. For shorter regions, this problem is reduced because the reads in both directions overlap, and sequencing errors can be eliminated by comparing complementary reads. For longer regions, poor-quality sequence overlap in the middle region can yield artifacts, which artificially increase sample diversity.

The choice of sequencing chemistries for whole-genome sequencing of purified isolates is important. Short-read chemistries (e.g., Illumina and Ion Torrent) are recommended if draft assemblies are sufficient. If complete genome assemblies are required, then long-read sequencing chemistries (e.g., PacBio or Nanopore) can be used to provide scaffolding to assemble data from the short-read chemistries (19).

CORE METADATA AND UROBIOME STUDY DESIGN

In clinical research, standardized guidelines for reporting randomized trials and observational studies have led to increased reporting quality and transparency for readers (20–22). In microbiome research, metadata guidelines function in a similar capacity to improve transparency, enhance interpretation, and facilitate integration and comparison of results among studies (23–25). Readers should be able to understand the design, conduct, and analysis of a microbiome study in order to comprehend and interpret results. Detailed and thorough reporting of metadata, the information that describes a sampling event and subsequent data generation efforts, facilitates a shared understanding of the relevance of research findings. In addition, collection and reporting of a common, minimal set of metadata across different projects will foster data comparisons and analysis; they will facilitate comparisons across studies and combining of studies to allow more powerful meta-analyses.

Following a review of other consensus-based guidelines and based on iterative discussions within the urobiome research community, we propose standard metadata elements for urobiome studies. These include the minimum required metadata elements as well as those that are optional but highly desired for publication (Table 1). Since urobiome studies commonly involve human subject research, protected health information must not be included in the sequencing data or metadata.

Within the proposed metadata elements, “required” elements refer to the absolute minimum information needed to make data interpretable. The “desired” elements include characteristics that enhance the reader's ability to interpret findings within specific cohorts. These elements have been associated with differences in microbiota in previous studies and thus are considered potentially confounding elements. We suggest that study teams aiming for a high level of rigor should collect information pertaining to the desired elements and either include this information when disseminating their research or explain the lack of inclusion. Researchers are highly encouraged to consider additional items relevant to their study design or specific research question. The recommended metadata elements in Table 1 are organized based on important biological, environmental, and technical factors that could introduce variability or confound results.

For studies that include marker gene sequencing (e.g., 16S rRNA gene sequencing), we have complied with the Genome Standards Consortium (GSC) recommendations for minimum information standards (MixS) for describing and publicly sharing these data (26). In collaborating with the GSC, we have created an environmental package (MixS-Urobiome) consisting of a checklist for describing minimum and desired

TABLE 1 Proposed elements to be included in the minimum metadata standards for reporting of urobiome research

Element(s)	Required/desired	Description ^d
Biological elements		
Age	Required	Age in years or months/days if appropriate for infant/young child population ^b
Sex	Required	Biological sex; gender if relevant for the study
Antibiotic usage	Desired	There is a lack of knowledge about postantibiotic microbiome recovery; when possible, we recommend recording of use in the prior 3 months or length of time between last antibiotic exposure and sample collection
Hormone status	Desired	Pubertal stage ^a Pregnant/postpartum Menopausal status: perimenopausal, postmenopausal Also specify if taking supplemental hormones (estrogen) and route (oral, transdermal, or vaginal, etc.) Last menstrual period (if menstruating)
Contraception	Desired	Use of oral contraceptives, other hormonal or nonhormonal/barrier, or none
Body mass index	Desired	Body mass index at the time of the study visit, calculated from height and weight
Race, ethnicity	Desired	If possible, use standard terminology from sources such as the U.S. census and SNOMED CT
Surgery	Desired	Performed in the prior 3 months Prior GU surgeries Prior implanted GU materials
Birth details ^b	Desired	Gestational age Mode of delivery NICU stay Method of feeding
Medical history	Desired	Diabetes/prediabetes Other relevant medical comorbidities Use of steroids or immunosuppressant medications GU anatomical abnormalities Recurrent GU infections Recent GU instrumentation
Urine characteristics	Desired	pH, specific gravity, leukocyte esterase, blood
Environmental variables		
Method of collection	Required	Void, collection device (Peezy) Catheter (use of Mitrofanoff ^a) Suprapubic aspirate
Geographic location ^c	Required	Can be discrete, including geographic coordinates, or broad, such as region or country
Seasonal	Desired	Month of collection
Dietary	Desired	Consumption of a special diet, use of fiber supplementation, yogurt consumption
Sexual activity	Desired	Time interval between last sexual activity and sample collection, if sexually active
Technical variables		
Date and time of collection ^c with conditions	Required	Used to ensure that samples stored at room temp for long periods are highlighted as such, potentially impacting the validity of results Ensure that the date is generic enough to be included or use a date range
Date and time of freezing	Required	Time interval between sample collection and freezing Omit if samples undergo immediate DNA extraction
Preservative	Required	If used, name
DNA extraction	Required	Method/kit used
Sequencing method ^c	Required	e.g., Illumina, Ion Torrent, Nanopore, PacBio, Sanger, pyrosequencing; include amplicon/variable region(s) used
Processing details	Desired	Including, but not limited to, details of sample transfer method and extraction protocol (sterile hood or technique), etc.

^aAdditional recommendation for pediatric populations.

^bAdditional recommendation for infant populations.

^cRequired when uploading sequence data to the Sequence Read Archive (SRA) (27) or the European Nucleotide Archive (ENA) (36) public data repository.

^dOCP, oral contraceptive pill; GU, genitourinary; NICU, neonatal intensive care unit.

information about marker gene analyses (26). Table S1 in the supplemental material displays a checklist structured to facilitate the uploading of information to public databases such as the Sequence Read Archive (SRA), where raw sequencing data are often shared (27). A Research Electronic Data Capture (REDCap) database template encompasses required and desired metadata elements should study teams wish to use a standard template for prospective studies (28).

TABLE 2 Guidelines for processing sequencing data for urobiome research^a

Data processing step	Description (reference[s])
Marker gene sequencing	
Grouping reads	Sequencing reads can be grouped into OTUs or ASVs; ASVs offer several advantages over OTUs, such as better accuracy and resolution, and hence are preferred (37); current ASV algorithms include DADA2 (38) and Deblur (39); significantly outdated OTU clustering algorithms (such as <i>uclust</i> [40]) should be avoided
Assigning taxonomy	Algorithm: taxonomy can be assigned with taxonomic classifiers such as naive Bayes or BLCA classifiers (41, 42); species-level assignment needs to be performed with algorithms designed for species-level assignments, such as BLCA or the exact matching approach implemented in DADA2 (43) Database: the Silva (42) and NCBI 16S (44) databases are preferred, as they are more representative of microbiota in the urobiome than the currently available version of the Greengenes database (v13_8) (43)
Data cleaning	Chimeras: chimeras arise from PCR and should be removed using an algorithm such as <i>ChimeraSlayer</i> (45) or <i>UCHIME</i> (37, 46) Contaminants: since catheter-collected specimens are typically low-biomass specimens, computational strategies for bacterial contaminants, identification, and removal should be used; <i>Decontam</i> is currently the preferred approach in conjunction with an exptl design that includes negative controls and/or a mock microbial dilution series to evaluate performance (47)
Whole-genome sequencing	
Data cleaning	Host DNA needs to be removed using tools such as <i>Bowtie2</i> with the current human reference genome (48)
Read processing	Sequencing reads can be processed using metagenomic <i>de novo</i> sequence assembly using tools such as <i>metaSPAdes</i> (49) or binned, where reads are clustered by sequence similarity, using tools such as <i>MaxBin</i> (50)
Annotation	Taxonomic annotation: marker genes such as 16S rRNA and well-characterized functional genes can be used for genus- and species-level annotations using tools such as <i>Metaphlan</i> (51) Gene annotation: identifying relevant features of bacterial genomes can be performed using tools such as <i>Prokka</i> (52) Metabolic pathway analysis: the metabolic functional potential of a microbial community can be modeled and explored using tools such as <i>CarveMe</i> (53); as with marker gene sequencing, annotation is highly dependent on the reference databases used and how well the urobiome microbiota are represented
Software pipelines for data analysis	
Marker genes	QIIME2 (54), <i>mothur</i> (55), and DADA2 (38)
WGS	MG-RAST (56), EBI MetaGenomics (57), and IMG/M (58)
Viral	Classification of eukaryotic viruses and bacteriophage: <i>Virmine</i> (59) Classification of bacteriophage: <i>VirSorter</i> (60)

^aOTUs, operational taxonomic units; ASVs, amplicon sequence variants; WGS, whole genome sequencing.

BIOINFORMATIC APPROACHES AND DATA ANALYSIS

Analyzing urobiome data is often tailored to the specific research questions addressed in a particular project, making it impractical to follow fixed analytical procedures. Table 2 displays guidance for documenting and reporting urobiome study data (29, 30). To ensure that urobiome data are appropriately handled and interpreted, it is essential to collaborate with bioinformaticians or computational biologists; consultation in the early stages of study design is recommended.

Several manipulations are needed to distill sequencing reads into biologically meaningful data for statistical analysis. Standard steps include quality filtering and denoising, grouping sequences by similarity for marker gene studies or binning approaches for whole-genome sequencing (WGS) studies, assembly for WGS studies, removing technical artifacts and noise, and assigning taxonomy (31). While the approach for a specific study depends on the data generated, the steps can be completed using freely available sequence processing platforms. Table 2 displays current guidelines and recommendations.

Urobiome studies are typically limited by a small sample size yet a large number of measured variables (taxa or genes). Thus, ecological community analyses such as alpha diversity (e.g., the Chao1, Simpson, Shannon, and Pielou indices) and beta diversity (e.g., Bray-Curtis and UniFrac) using nonmetric multidimensional scaling (NMDS) and principal-coordinate analysis (PCoA) are applied for multivariate analyses of microbiomes (32). These measures can identify overall differences between study groups. Drilling down to the level of taxa or genes is often desired, but the process is complex. Although standard statistical methods are often applied, it is important to realize that

TABLE 3 Minimum information for reporting bioinformatics methods in urobiome studies

Information to be included	Description (reference) ^a
Software	Include software package and version; if using a package such as QIIME (61), reference key algorithms for OTU/ASV generation, taxonomy assignment, chimera removal, and contaminant detection
Databases	Include databases used and version
Code	Include essential custom-written code for analysis or data processing as supplemental material or link to code repository such as GitHub
Data	Raw sequencing data: stored in a public repository such as SRA (27), ENA (36), or dbGaP (62) WGS assemblies: stored in a public repository such as GenBank Metadata: follow MIMARKS (26) or MixS guidelines; upload with raw data

^aSRA, Sequence Read Archive; ENA, European Nucleotide Archive; dbGaP, Database of Genotypes and Phenotypes.

these methods are often not suitable because urobiome data are compositional, multivariate, nonnormal, highly skewed, and zero inflated. Therefore, we encourage the use of statistical methods tailored to microbiome data (33). Multiple-test correction is important for controlling for false positives in statistical analyses; however, these efforts may diminish real scientific findings. Thus, we recommend that investigators report raw and corrected *P* values and provide scientific justification for results that should be subject to further investigation and validation. Furthermore, it is important to realize that the exploratory nature of most urobiome projects (at least at the initial phase) makes defining a meaningful “effect size” *a priori* required for sample size calculation challenging.

To ensure the reproducibility of an analysis, documentation of computational steps, software, and data used is essential (34). For example, analysis performed in the R statistical programming language can be documented in RMarkdown (35). This documentation can be shared as supplemental material or stored on a code repository such as GitHub. Both raw data and the associated metadata should be deposited in public repositories for reanalysis (26). In the manuscript methods, software details should be appropriately mentioned and referenced (Table 3).

CONCLUDING COMMENTS

Urobiome research has the potential to advance our understanding of human health and a wide range of diseases, including lower urinary tract symptoms and kidney disease. Many scientific areas have benefited from early consensus on research methods by allowing investigators to more appropriately compare their findings with those of their colleagues, optimizing transparency and communication and facilitating research for the greater, common good. We offer this first consensus document with every expectation of subsequent revision as our field progresses.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TABLE S1, XLSX file, 0.04 MB.

REFERENCES

1. Wolfe AJ, Brubaker L. 2019. Urobiome updates: advances in urinary microbiome research. *Nat Rev Urol* 16:73–74. <https://doi.org/10.1038/s41585-018-0127-5>.
2. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. *J Clin Microbiol* 50:1376–1383. <https://doi.org/10.1128/JCM.05852-11>.
3. Chen YB, Hochstedler B, Pham TT, Alvarez MA, Mueller ER, Wolfe AJ. 2020. The urethral microbiota: a missing link in the female urinary microbiota. *J Urol* 204:303–309. <https://doi.org/10.1097/JU.0000000000000910>.
4. Southworth E, Hochstedler B, Price TK, Joyce C, Wolfe AJ, Mueller ER. 2019. A cross-sectional pilot cohort study comparing standard urine collection to the Peezy midstream device for research studies involving women. *J Female Pelvic Med Reconstr Surg* 25:e28–e33. <https://doi.org/10.1097/SPV.0000000000000693>.
5. Price TK, Wolff B, Halverson T, Limeira R, Brubaker L, Dong Q, Mueller ER, Wolfe AJ. 2020. Temporal dynamics of the adult female lower urinary tract microbiota. *mBio* 11:e00475-20. <https://doi.org/10.1128/mBio.00475-20>.
6. Dong Q, Nelson DE, Toh E, Diao L, Gao X, Fortenberry JD, Van der Pol B. 2011. The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. *PLoS One* 6:e19709. <https://doi.org/10.1371/journal.pone.0019709>.
7. Bajic P, Van Kuiken ME, Burge BK, Kirshenbaum EJ, Joyce CJ, Wolfe AJ, Branch JD, Bresler L, Farooq AV. 2020. Male bladder microbiome relates to lower urinary tract symptoms. *Eur Urol Focus* 6:376–382. <https://doi.org/10.1016/j.euf.2018.08.001>.

8. Jung CE, Chopyk J, Shin JH, Lukacz ES, Brubaker L, Schwanemann LK, Knight R, Wolfe AJ, Pride DT. 2019. Benchmarking urine storage and collection conditions for evaluating the female urinary microbiome. *Sci Rep* 9:13409. <https://doi.org/10.1038/s41598-019-49823-5>.
9. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol* 52:871–878. <https://doi.org/10.1128/JCM.02876-13>.
10. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger P. 2016. The clinical urine culture: enhanced techniques improve detection of clinically relevant microorganisms. *J Clin Microbiol* 54:1216–1222. <https://doi.org/10.1128/JCM.00044-16>.
11. Kass EH. 1956. Asymptomatic infections of the urinary tract. *Trans Assoc Am Physicians* 69:56–64.
12. Khasriya R, Sathiananthamoorthy S, Ismail S, Kelsey M, Wilson M, Rohn JL, Malone-Lee J. 2013. Spectrum of bacterial colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. *J Clin Microbiol* 51:2054–2062. <https://doi.org/10.1128/JCM.03314-12>.
13. Coorevits L, Heytens S, Boelens J, Claeys G. 2017. The resident microflora of voided midstream urine of healthy controls: standard versus expanded urine culture protocols. *Eur J Clin Microbiol Infect Dis* 36:635–639. <https://doi.org/10.1007/s10096-016-2839-x>.
14. Ackerman AL, Underhill DM. 2017. The mycobiome of the human urinary tract: potential roles for fungi in urology. *Ann Transl Med* 5:31. <https://doi.org/10.21037/atm.2016.12.69>.
15. Karstens L, Siddiqui NY, Zaza T, Barstad A, Amundsen CL, Sysoeva TA. 2020. Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *bioRxiv* <https://doi.org/10.1101/2020.11.10.375279>.
16. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 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>.
17. Yuan S, Cohen DB, Ravel J, Abdo Z, Fomey LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7:e33865. <https://doi.org/10.1371/journal.pone.0033865>.
18. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques* 62:290–293. <https://doi.org/10.2144/000114559>.
19. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif* 3:1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>.
20. Moher D, Schulz KF, Altman DG, CONSORT Group. 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357:1191–1194. [https://doi.org/10.1016/S0140-6736\(00\)04337-3](https://doi.org/10.1016/S0140-6736(00)04337-3).
21. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. 2007. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370:1453–1457. [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X).
22. Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). 2001. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 285:1992–1995. <https://doi.org/10.1001/jama.285.15.1992>.
23. Dundore-Arias JP, Eloie-Fadros EA, Schriml LM, Beattie GA, Brennan FP, Busby PE, Calderon RB, Castle SC, Emerson JB, Everhart SE, Eversole K, Frost KE, Herr JR, Huerta AI, Iyer-Pascuzzi AS, Kalil AK, Leach JE, Leonard J, Maul JE, Prithiviraj B, Potrykus M, Redekar NR, Rojas JA, Silverstein KAT, Tomso DJ, Tringe SG, Vinatzer BA, Kinkel LL. 2020. Community-driven metadata standards for agricultural microbiome research. *Phytobiomes J* 4:115–121. <https://doi.org/10.1094/PBIOMES-09-19-0051-P>.
24. Cole JR, Myrold DD, Nakatsu CH, Owens PR, Kowalchuk GA, Tiedje CC, Tiedje TM. 2010. Development of soil metadata standards for international DNA sequence databases, p 5–8. *In* 19th World Congress of Soil Science, soil solutions for a changing world, 1–6 August, Brisbane, Australia. International Union of Soil Sciences, Vienna, Austria.
25. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolk T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–463. <https://doi.org/10.1038/nature24621>.
26. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PSG, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, et al. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat Biotechnol* 29:415–420. <https://doi.org/10.1038/nbt.1823>.
27. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The Sequence Read Archive. *Nucleic Acids Res* 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019>.
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42:377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>.
29. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth J, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>.
30. Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, Gregory WT, Nardos R, McWeeney SK. 2018. Community profiling of the urinary microbiota: considerations for low-biomass samples. *Nat Rev Urol* 15:735–749. <https://doi.org/10.1038/s41585-018-0104-z>.
31. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolk T, McCall L-I, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422. <https://doi.org/10.1038/s41579-018-0029-9>.
32. Ramette A. 2007. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62:142–160. <https://doi.org/10.1111/j.1574-6941.2007.00375.x>.
33. Calle ML. 2019. Statistical analysis of metagenomics data. *Genomics Inform* 17:e6. <https://doi.org/10.5808/GI.2019.17.1.e6>.
34. Schloss PD. 2018. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio* 9:e00525-18. <https://doi.org/10.1128/mBio.00525-18>.
35. Xie Y, Allaire JJ, Grolemund G. 2020. *R Markdown: the definitive guide*. CRC Press, Boca Raton, FL.
36. Amid C, Alako BTF, Balavenkataraman Kadhivelu V, Burdett T, Burgin J, Fan J, Harrison PW, Holt S, Hussein A, Ivanov E, Jayathilaka S, Kay S, Keane T, Leinonen R, Liu X, Martinez-Villacorta J, Milano A, Pakseresht A, Rahman N, Rajan J, Reddy K, Richards E, Smirnov D, Sokolov A, Vijayaraja S, Cochrane G. 2020. The European Nucleotide Archive in 2019. *Nucleic Acids Res* 48:D70–D76. <https://doi.org/10.1093/nar/gkz1063>.
37. Callahan B, McMurdie P, Holmes S. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 11:2639–2643. <https://doi.org/10.1038/ismej.2017.119>.
38. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
39. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2:e00191-16. <https://doi.org/10.1128/mSystems.00191-16>.
40. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.

41. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
42. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42: D643–D648. <https://doi.org/10.1093/nar/gkt1209>.
43. Hoffman C, Siddiqui NY, Fields I, Gregory WT, Simon H, Mooney MA, Wolfe AJ, Karstens L. 2020. Species level resolution of female bladder microbiota from marker gene surveys. *bioRxiv* <https://doi.org/10.1101/2020.10.27.358408>.
44. Federhen S. 2012. The NCBI Taxonomy database. *Nucleic Acids Res* 40: D136–D143. <https://doi.org/10.1093/nar/gkr1178>.
45. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Human Microbiome Consortium, Petrosino JF, Knight R, Birren BW. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21:494–504. <https://doi.org/10.1101/gr.112730.110>.
46. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
47. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, 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>.
48. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
49. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
50. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. <https://doi.org/10.1186/2049-2618-2-26>.
51. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9:811–814. <https://doi.org/10.1038/nmeth.2066>.
52. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
53. Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res* 46:7542–7553. <https://doi.org/10.1093/nar/gky537>.
54. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
55. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
56. Keegan KP, Glass EM, Meyer F. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods Mol Biol* 1399:207–233. https://doi.org/10.1007/978-1-4939-3369-3_13.
57. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar GA, Pesseat S, Boland MA, Hunter FMI, Ten Hoopen P, Alako B, Amid C, Wilkinson DJ, Curtis TP, Cochrane G, Finn RD. 2018. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res* 46:D726–D735. <https://doi.org/10.1093/nar/gkx967>.
58. Chen I-MA, Chu K, Palaniappan K, Pillay R, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpidis NC. 2019. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* 47:D666–D677. <https://doi.org/10.1093/nar/gky901>.
59. Zheng T, Li J, Ni Y, Kang K, Misiakou M-A, Imamovic L, Chow BKC, Rode AA, Bytzer P, Sommer M, Panagiotou G. 2019. Mining, analyzing, and integrating viral signals from metagenomic data. *Microbiome* 7:42. <https://doi.org/10.1186/s40168-019-0657-y>.
60. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.
61. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung N, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenkov T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
62. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 39:1181–1186. <https://doi.org/10.1038/ng1007-1181>.