Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2019 December ; 28(12): 1949–1954. doi: 10.1158/1055-9965.EPI-19-0557.

Perspectives for Consideration in the Development of Microbial Cell Reference Materials

Emma Allen-Vercoe[1,*], **J. Russell Carmical**[2], **Samuel P. Forry**[3], **Mitchell H. Gail**[4], **Rashmi Sinha**[4]

[1]Department of Molecular and Cellular Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, CANADA

[2]Molecular Virology and Microbiology Core, Alkek Center for Metagenomics & Microbiome Research, Baylor College of Medicine, Houston, TX, United States of America

[3]Biosystems and Biomaterials Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899 United States of America

[4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Bethesda, MD, United States of America

Abstract

Microbiome measurement and analyses benefit greatly from incorporation of reference materials as controls. However, there are many points to consider in defining an ideal whole cell reference material standard. Such a standard would embody all the diversity and measurement challenges present in real samples, would be completely characterized to provide ‘ground truth’ data, and would be inexpensive and widely available. This ideal is, unfortunately, not readily attainable because of the diverse nature of different sequencing projects. Some applications may benefit most from highly complex reference materials, while others will value characterization or low expense more highly. The selection of appropriate microbial whole cell reference materials to benchmark and validate microbial measurements should be considered carefully and may vary among specific applications. In this commentary, we describe a perspective on the development of whole cell microbial reference materials for use in metagenomics analyses.

[*]Corresponding author: eav@uoguelph.ca.

Authors’ contributions

All authors contributed equally to the conception, writing, and editing of this manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study

Competing interests

EAV is co-founder of NuBiyota, a company that aims to commercialize ‘microbial ecosystem therapeutics’, and that would therefore benefit from the development of an ideal standard for metagenomic analysis. JMC, SF, MG, RS declare that they have no competing interests. Commercial material suppliers are identified in this paper to foster understanding; such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials identified are necessarily the best available for the purpose.

Keywords

Microbiome; Reference Materials; Measurement Confidence; Whole Cell; Prospective cohort

Background

As evidenced by recent publication trends in microbiome research, researchers are moving the field of metagenomics toward large epidemiologic studies and commercial development at a rapid pace. If these analyses are to ever evolve into reliable assays (e.g. for etiologic studies or clinical diagnostics), the measurement process must be regularly assessed to ensure *measurement quality*. A key aspect of this validation is the routine analysis of reference materials as both positive and negative controls. A reference material is any stable, broadly available, and well characterized specimen that is used to assess the quantitative and/or qualitative validity of a measurement process. Many sorts of reference materials are available, and we focus here on the use of microbial Whole Cell Reference Materials (WCRMs) for characterizing metagenomic analyses.

Main text

Kinds of microbial reference materials

In order to make the most accurate comparisons, a reference standard should closely resemble the physical, chemical, and biological characteristics of the sample of interest. However, it's rarely if ever possible to identify WCRMs which are identical to the sample source and that have been characterized to the extent necessary. Therefore, compromises must be made, and a universal WCRM is not feasible. Further, the specific questions being addressed in an investigation will dictate the most appropriate WCRM (balancing, for example, taxonomic similarity to test samples, depth of characterization, availability and reproducibility). Many diverse WCRMs have been utilized by different research efforts and commercial interests (Table 1).

Microbe-based reference materials can be roughly divided into 3 key categories (Figure 1). Each of these is associated with benefits and drawbacks that need to be assessed in the context of experimental design for a given project.

Category 1: Naturally-occurring microbial ecosystems—Samples from existing microbial environments may be suitable as WCRMs if they can be (i) obtained in large enough quantities, (ii) homogenized to remove variability between aliquots, and (iii) stored for long periods of time with either negligible or immeasurable deterioration. The most obvious benefit of this kind of WCRM is that its composition will closely match the sample of interest that will be collected from that environment.

There are, however, several drawbacks to using environmental samples as WCRMs. One of the major limitations of using such a sample is that it is challenging to define (i.e., measure) the exact composition and abundance of species. This is particularly difficult for ecosystems that are complex and contain many low-abundance taxa. Also, the ecosystem source for the WCRM may change over time; thus the WCRM cannot easily be reproduced if more

material is needed later. For example, although an individual's fecal microbial ecosystem may remain generally stable over long periods, variations (e.g., in abundance profiles) may be observed, reflecting changes in host diet, immune status, physical fitness and environment, etc. over time [1–3].

Category 2: in vitro expansion of microbial ecosystems—The abundance of a given WCRM sourced directly from the environment is a major limitation to its usefulness. To overcome this limitation, several groups have developed laboratory systems that mimic many parameters that shape ecosystems, whilst at the same time allowing these parameters to be measurable and controlled (reviewed in [4]). For example, bioreactors support and maintain the environmental pH and oxygen tension matched to the source environment and provide a constant supply of nutrients and removal of waste material. They can be seeded with environmental innocula and allowed to attain an equilibrium (steady state) under user-defined conditions. The steady state ecosystem may be maintained for several weeks or more, depending on the environment being modeled, and the system can also be scalable. Since these kinds of model systems attempt to recapitulate the natural environment, an added benefit is that any reference material obtained from them will be supported within a matrix that will at least partially reflect that of the natural source. Another advantage is that a scalable quantity of material can be produced at one time. Taken together, these factors partially overcome the limitations of reference material availability described for the first category, above.

When seeded with complex microbial ecosystems from a given natural source, however, *in vitro* models remain subject to batch-to-batch variations that need to be completely characterized each time. This problem may be mitigated by storing multiple identical innocula to re-start the bioreactor on multiple occasions. However, seed stocks are limited and may suffer from deterioration over time even under optimal storage conditions. As with the environmental sample WCRMs discussed above, complete characterization of these *in vitro* expansions remains a significant technical challenge.

Category 3 – pure microbial isolates—Many representative organisms from diverse ecosystems have been isolated and cultured as pure isolates. Defined mixtures of these pure strains can be used to generate WCRMs in two ways: a) a subset of a natural community may be combined to inoculate a defined ecosystem in an *in vitro* model, as described above; and b) pure isolates may be grown separately and subsequently mixed in defined ratios.

Defined subset *in vitro* communities are cultured in systems that potentially restore some metabolic functionality and matrix components of the ecosystem. Although inherently artificial, this modality offers a balance of a defined set of input strains while allowing for those strains to adopt a semi-natural state (e.g., compositional profile, metabolic activity) to mimic some aspects of the *in vivo* situation.

In contrast to this approach, pure microbial isolates have several key benefits as components of microbial standards. Using adequately banked strains and well-characterized propagation procedures, it is possible to obtain an almost limitless supply. In addition, pure isolates can be subjected to thorough phenotypic and genotypic characterization, which adds value to

any resulting microbial standard. Furthermore, aspects of experimental design can be accommodated, such as the need to balance e.g. gram-positive and gram-negative bacterial species in a useful way. Microbial isolates may also be grouped into a standard at specified abundances to mimic natural ecosystem, i.e. a 'mock community' (MC); however, the resulting ecosystem will be inherently artificial and will lack properties, such as metabolite profiles, that can only be gained through culture of an ecosystem as a unit.

Although many banked microorganisms are in principal available for fit-for-purpose assembly of MCs, they often lack the level of characterization required for use as an appropriate reference material. Many consist of incomplete genomes and/or are poorly annotated. For example, the copy number of the 16S rRNA gene in databased bacterial sequences can be unreliable. Whole genome sequencing and complete genome assembly of candidate strains could resolve these issues but requires a large upfront expense. Furthermore, a repository for characterization data would aid the utility and uptake of appropriate WCRMs.

Synthetic mock community WCRMs provide an opportunity to design mixtures that are particularly well suited to test known microbial measurement challenges. For example, bacterial WCRMs should include both gram-positive and gram-negative organisms. A range of relative abundances should be present with at least a 100-fold difference between high- and low-abundance organisms. The genetic diversity of the WCRM should also be considered. Some naturally occurring microbiomes are comprised of >1000 phylogenetically diverse species, while other microbial communities are much simpler. Ideal WCRMs will encompass the expected diversity and phylogenetic breadth of actual samples. The additional inclusion of groups of closely related organisms, including strains within the same species, would be useful for assessing the sensitivity of measurement pipelines to genotypes exhibiting high genetic similarity. Finally, many microbial samples include matrices that complicate analysis (e.g., PCR inhibitors are common in fecal samples). WCRMs can be designed to include matrix elements that are known to be problematic. Alternatively, assembled WCRMs spiked into relevant matrices may be used.

Microbial whole cell reference materials (WCRM) utilization

WCRMs can be used at the beginning of a project to define, evaluate and assess the protocols to be used before going on to work with perhaps expensive or precious samples. During a project, WCRMs evaluated regularly alongside unknown samples will increase confidence that the measurement process is consistently performing as intended. WCRMs can be evaluated when introducing different protocols and platforms to an existing workflow (e.g., use of a new instrument, or a different sequencing center), and this provides an opportunity to increase the value and portability of resulting datasets. An important caveat here is that WCRMs are likely to differ – sometimes substantially – from the actual samples of interest. It remains possible that a measurement process that appears well characterized and predictable based on WCRM analysis will be impacted by particular attributes of actual samples in unanticipated ways. Thus, while WCRMs are very useful for identifying measurement failures, they cannot guarantee valid measurement results. Some common questions and considerations about the measurement process that WCRMs can help address

are provided in Table 2. It is recommended that when microbial measurement results are published, the data from routine analysis of well characterized WCRMs are also made available (i.e., as data critical in supporting the research conclusions). Some studies of mock communities with known composition have been used to demonstrate biases in observed relative abundances that can occur at any stage in the process from DNA extraction, to PCR amplification, to bioinformatics [9, 10]. Posted work by McLaren et al. describes how to estimate taxon-specific biases in mock samples [11]. While such research is promising, it remains unclear if WCRM data could be used to adjust the relative abundances measured in complex experimental samples; perhaps more formal methods for using the WCRM data will become available.

WCRM characterization challenges

Some common characteristics to be considered for metagenomic WCRMs are; taxonomic composition, known and accessible sample source, growth conditions (if cultivable), readily available reference sequence, genome size, GC content, percent repeats in the genome, known copy number of target sequences (e.g. 16S or 18S rRNA genes), and functional information.

For example, one potential use for MCs is to measure the differential in nucleic acid extraction efficiencies among community constituents. To make such assessments, we must first quantify the constituents of the MC and be able to add each in a known quantity; by processing a known quantity one should be able to calculate an expected result. There are various means to counting cells such as plating and colony forming units (CFU) counting, flow cytometry and spectrophotometry but none of these offer the accuracy required to make downstream measurement on the molecular level. Once the MC has been subjected to the extraction process the resulting DNA must be measured, and quite often this is found to be at a concentration of <1ng/ul. Spectrophotometry (NanoDrop) and fluorescent dyes (picogreen) lack the accuracy to measure concentrations typical of metagenomic samples, which leaves qPCR as the choice of measurement. In order for qPCR to be used to measure the individual constituents in a mixed population one must design primers to a sequence that is both unique to a given component strain and occurs at a known copy; this often proves to be problematic. Furthermore, a target that is common to a constituent and occurs in equal copies should be chosen to serve a calibrator for making copy number calculations, which can also be challenging. Without an accurate means of measuring both input (cells) and output (DNA) it becomes very difficult to assess the efficiency of the extraction process.

As all samples of interest are compared to a common reference standard, a poorly suited standard may result in a systemic bias. For example, assembling MCs from single isolates to assimilate metagenomic samples is a logical approach but has proven difficult to achieve.

Development of WCRMs for epidemiologic studies: A Case Study

Epidemiologic microbiome studies benefit greatly from the routine inclusion of comparable standard samples in sequencing experiments to allow laboratories to assess their own performance over time and facilitate data pooling or meta-analyses that are anticipated. As

no complex WCRMs for positive controls for microbiome analysis are currently readily available, researchers at the United States National Cancer Institute (NCI) have recently developed limited quantities of three types of standard reference material for future etiologic studies (Table 1). Samples have been divided into multiple, identical one-use aliquots for distribution to the cancer epidemiologic studies. Cancer epidemiologists interested in obtaining these samples should get in touch with Rashmi Sinha (sinhar@nih.gov) at NCI. Due to limited availability of samples, requests will be reviewed by a committee and a material transfer agreement will need to be established.

The first type of WCRM is a community of 45 purified bacterial strains, grown to known biomass quantities and mixed together in known ratios which will serve as the quantitative "gold standard" artificial community. These bacterial strains were similar to fecal-derived species but considered rare within most donors, provided in known ratios as a mock community in a saline solution. The second of these reference materials is fecal material. Researchers collected >200g fecal samples each from five individual donors (healthy adult, adult on low carbohydrate diet, adult with high body mass index, adult with inflammatory bowel disease, and healthy infant) in order to optimize the ability to observe microbial differences. These five samples were spiked with the artificial community mentioned above. The third is material from a chemostat grown to steady state equilibrium under conditions mimicking those found within the mouth, and seeded with samples of saliva, tongue scraping, and plaque and oral bacterial communities grown as biofilms. These WCRM samples will be extracted and the 16S ribosomal RNA amplicons as well as the metagenomes will be sequenced in multiple laboratories to provide reference data.

The primary use of the WCRM samples will be to see if a particular batch in a given lab meets quality control specifications. If not, data from that batch might need to be discarded. So long as cases and controls are balanced within batches, however, some batch aberrations might be acceptable. Each WCRM needs to be characterized (preferably in more than one expert laboratory) to determine relative abundances of taxa at several taxonomic levels and corresponding measures of alpha diversity, such as Shannon index. For each parameter to be used for quality control (relative abundances, alpha diversity), at least two expert labs should make the measurement repeatedly on different days to estimate within laboratory and between laboratory components of variance. These can be used to determine tolerance limits for quality control testing; values outside such limits indicate that the measurement is bad.

Then, to determine whether an experimental batch is acceptable in a new study, the following quantities might be tested to see if they are within tolerance limits to what was expected:

1. Alpha diversity
2. Phyla relative abundances; and ranks of phyla relative abundances (other taxonomic levels could be examined also)
3. Ability to detect some rare genus or species that had been intentionally spiked in
4. If contaminants not in the WCRM are appreciable, issue a warning. Perhaps eliminate that batch if the relative abundance of the contaminant is large.

5. Negative controls should also be tested for contaminants and the batch eliminated if large quantities are detected.

In a similar manner, data from WCRM could be used to eliminate bad batches for pooling data among different study centers or for meta-analyses. In each study center, the investigator could discard bad batches and only use the good batches to compute summary statistics like the logistic regression slope of case-control status on alpha diversity. These summary statistics could be submitted for meta-analysis. But the WCRM quality control would still be based on assessment of a given batch. For measures of beta-diversity, investigators could pool data only from batches that passed WCRM quality control criteria.

Negative controls

Another aspect that is critical is the need to incorporate negative controls into a study design. Studies should routinely include collection and extraction buffers alone for sequence analysis as they may contain low-level contamination that, despite being minimal, may lead to incorrect detection of non-biologically relevant bacteria. This is especially important in the context of microbiome studies that propose to identify rare taxa in their biological samples. Recently, the Microbiome Quality Control Study found that even simple buffer blank controls had significant contamination [6]. Contamination in buffer blanks is normally a result of trace amounts of DNA present in the DNA isolation, barcoding and sequencing kits [12], or as a result of cross-contamination of samples that are processed together [13]. In general, researchers should be aware that reagents found within most molecular biology kits can introduce contaminants that can mask true biological signal from low-biomass samples, and that this can be a particularly thorny problem when trying to assess the presence of low abundance taxa.

Conclusions

A variety of WCRMs have been developed to provide ‘ground truth’ benchmark data for characterizing metagenomic measurements. The authors are left with the following thoughts:

- Measurement quality is a critical part of any experiment and is particularly important for analysis of complex samples as is commonly carried out in microbiome studies.
- The use of WCRMs represents a logical approach for characterizing and validating experimental procedures and measurement processes.
- WCRMs can increase confidence in experimental design and resulting measurements but cannot guarantee accurate or reliable results.
- To be useful a given WCRM needs to be consistent (i.e., unchanging in time or between labs), well characterized, widely available (ideally in unlimited supply) and inexpensive.
  - A one-size-fits-all WCRM is thus an unattainable goal
- WCRMs need to be chosen that best fit a given experiment or question, taking into account, for example, diversity (both phylogenetic and in terms of microbial

characteristics such as GC ratios or cell wall thickness), abundance, and source environment.

  - A research question may require the use of multiple microbial reference materials throughout the entirety of a research investigation.
  - A modular design (addition of sub-groups of samples) may address some challenges in the application of WCRMs to different projects.

- Large prospective cohort studies being planned now should include standard WCRMs to facilitate future data pooling and meta-analyses by excluding batches where the quality control standards were not within the limit of tolerance.

Acknowledgements

The authors would like to acknowledge and thank participants at the 2016 Standards for Microbiome Measurements Workshop at NIST where the ideas contained herein were nucleated. The authors would also like to acknowledge Dr. Rob Knight (University of California, San Diego, California, USA), and Drs. Deirdre Devine, Phil Marsh, and Thuy Do (University of Leeds, UK) for their help in developing the new WCRMs for prospective cohort studies.

Funding

Funding for developing WCRM for prospective studies was supported by the Intramural Research Program of the NCI at the NIH. Funding for the 2016 Standards for Microbiome Measurements Workshop was provided by NIST, NIAID, and multiple corporate sponsors (including Zymo Research and ATCC). The workshop proceedings are publicly available online [14].

List of abbreviations

WCRM	Whole cell reference material
MC	Mock community
CFU	Colony forming units
qPCR	Quantitative polymerase chain reaction
PCR	Polymerase chain reaction
ATCC	American Type Culture Collection
HMP	Human Microbiome Project

References

1. Costello EK, Bacterial community variation in human body habitats across space and time. Science, 2009 326: p. 1694–1697. [PubMed: 19892944]
2. Schloissnig S, et al., Genomic variation landscape of the human gut microbiome. Nature, 2013 493(7430): p. 45–50. [PubMed: 23222524]
3. Flores GE, C. J, Henley JBet al. , Temporal variability is a personalized feature of the human microbiome. Genome Biol, 2014 15(531).
4. Venema K and van den Abbeele P, Experimental models of the gut microbiome. Best Pract Res Clin Gastroenterol, 2013 27(1): p. 115–26. [PubMed: 23768557]
5. Sinha R, et al., The microbiome quality control project: baseline study design and future directions. Genome Biol, 2015 16: p. 276. [PubMed: 26653756]

6. Sinha R, et al., Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol, 2017 35(11): p. 1077–1086. [PubMed: 28967885]
7. ZymoBIOMICS Microbial Community Standard. 2018 [cited 2018 February 28]; Available from: https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards.
8. Microbiome Standards and Research Solutions. 2018 [cited 2018 February 28]; Available from: https://www.atcc.org/~/media/A36E4E940E3F40B1A5E14D6851E9348F.ashx.
9. Brooks JP, Challenges for case-control studies with microbiome data. Ann Epidemiol, 2016 26(5): p. 336–341.e1. [PubMed: 27255739]
10. Brooks JP, et al., The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol, 2015 15: p. 66. [PubMed: 25880246]
11. McLaren MR, Willis AD, and Callahan BJ, Consistent and correctable bias in metagenomic sequencing measurements. bioRxiv, 2019: p. 559831.
12. Salter SJ, et al., Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 2014 12(1): p. 87. [PubMed: 25387460]
13. Walker AW, A Lot on Your Plate? Well-to-Well Contamination as an Additional Confounder in Microbiome Sequence Analyses. mSystems, 2019 4(4): p. e00362–19. [PubMed: 31239398]
14. Standards for Microbiome Measurements, Day 1 Part 1. 2018 [cited 2018 February 28]; Available from: https://www.nist.gov/video/standards-microbiome-measurements-day-1-part-1.

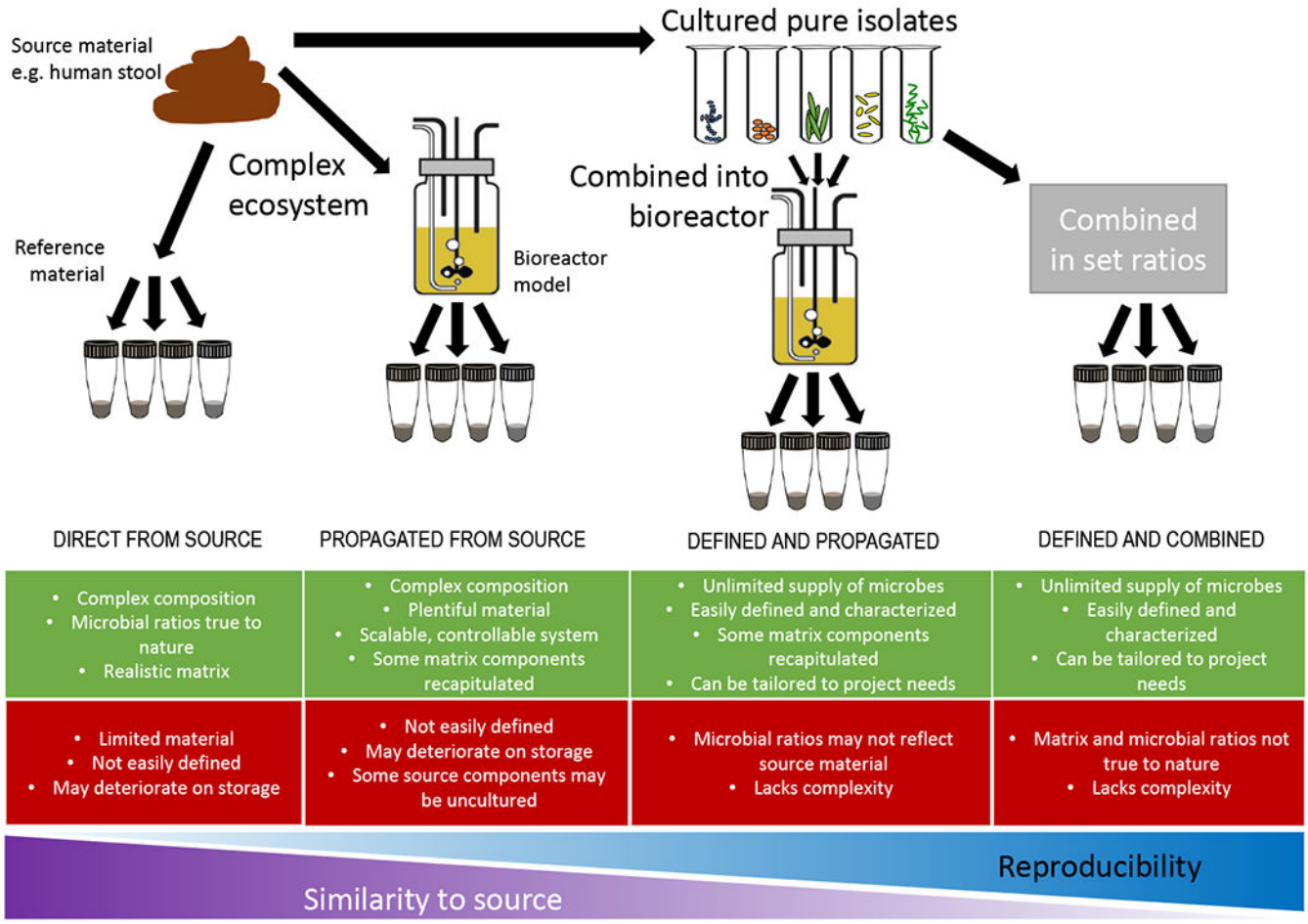

Figure 1.
Three key categories of microbe-based reference materials

Table 1.

Sources for Microbial Cell Reference Materials

Microbial Cell Reference Materials	Details
(✓) Previously analyzed samples	Many laboratories routinely re-analyze banked samples from previous investigations. These are useful for monitoring the run-to-run stability of the measurement process.
Focused community efforts: Microbiome Quality Control project: ✖BEI-ATCC HMP control [5, 6] ✖In vitro 'Robogut' model [5, 6] QC samples for NCI prospective cohort studies: (✓) Mock community (✓) Fecal samples spiked with mock-community (✓) Oral chemostat	Mixture of 20 pure cultures isolated from fecal and oral samples A bioreactor designed to propagate gut microbes was inoculated with feces-derived communities Mixture of 45 pure culture isolates, similar to fecal-derived species but considered rare within most donors, provided in known ratios as a mock community in a saline solution Fecal samples from five individuals with different phenotypes - spiked with mock community (mentioned above) that are considered rare within most donors to control for factors related to stool matrix composition A bioreactor designed to propagate oral microbes was inoculated with oral-derived communities
Purchased mock communities: ✓ ZymoBiomics™ by Zymo Research [7] ✓ MSA-2003™ by ATCC [8] ✓ MSA-2002™ by ATCC [8]	10-organism whole cell mixture, comprised of 3 gram positive bacteria, 5 gram negative bacteria, and 2 yeasts at even abundances 10-species whole cell mixture of 7 gram positive and 3 negative bacteria at even abundances 20-species whole cell mixture of 11 gram positive and 9 gram negative bacteria at even abundances (including the 10 organisms in MSA-2003)

✖ No longer available; (✓) Limited availability; ✓ Currently available.

Table 2.

Utilization of microbial WCRMs

Questions WCRMs can help address	Considerations
Does the measurement process exhibit bias (e.g., systematic underestimation of gram-positive abundances)?	• Measurement results should reflect the known WCRM composition. • Well characterized biases may be accounted for during subsequent data analysis
Is the measurement process performing within specifications (e.g., CLIA certification)?	• Measurement results from analysis of WCRMs must consistently fall within a predetermined range. • Deviations could result in discarding data from batched samples that are out of specification on the WCRM.
How should datasets be compared (e.g., data pooling for cohort studies, integrating future data collection as protocols evolve)?	• A WCRM should be widely available, and routine analysis of WCRMs under multiple protocols should inform the interpretation of results. • Future cohorts should ideally use the same WCRM to be able to compare their results.
Can a protocol be improved; what parameters are most important (e.g., duration and intensity of bead beating)?	• A WCRM should provide an inexpensive sample for systematically varying protocol parameters for optimization and sensitivity analyses.
How sensitive is the measurement process? What are the limits of quantitation (e.g., discriminate genetically similar strains, detect low-abundance taxa)?	• Well-designed WCRMs should be able to test the performance of measurement processes. • Ideally, WCRMs should include characteristics that challenge the measurement process to failure to help identify reasonable ranges for quantitative assessments.