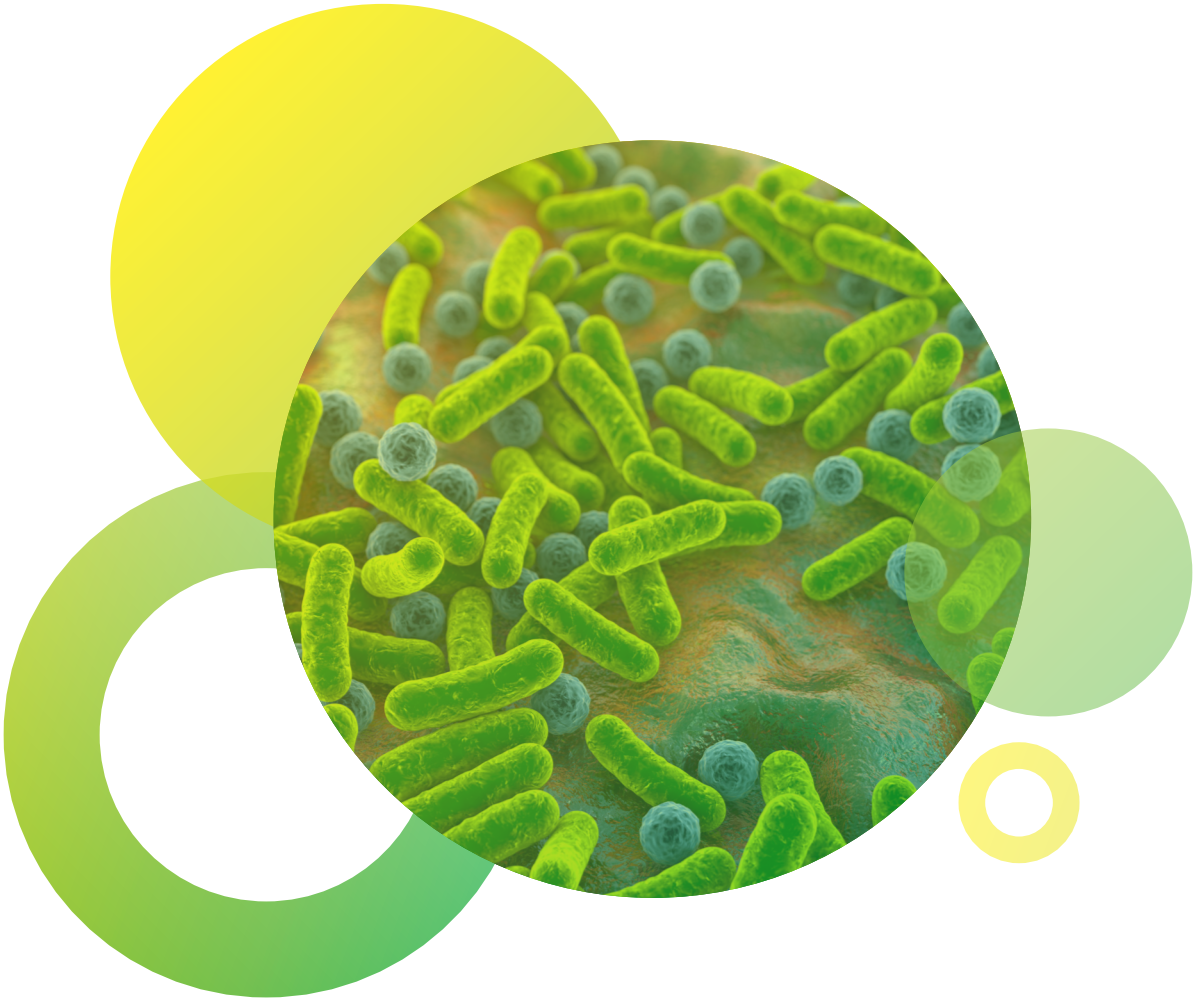
ZYMO RESEARCH

The Beauty of Science is to Make Things Simple

How to Choose a Microbiome Standard



Controls and Standards in Microbiome Research

The advancement of NGS based technologies has led to a rapid growth in the field of microbiome research and deciphering microbial community composition, function, and interactions. Many studies conclude that technical variability in microbiome processing methods leads to significant variations in results[1-3]. Most of the discrepancies in reporting are explained by differences among the methods for nucleic acid extraction, NGS library preparation, bioinformatic data processing, and the choice of reference databases. Despite the complexity and variation introduced by varying protocols and methods for each step of the microbiomics workflow, data is being generated at an unprecedented pace. In many cases, a lack of proper controls or comparison to microbiome reference materials means that important and high-impact conclusions cannot be reproduced or reliably compared to similar data sets.

Commonly used and accepted controls or reference reagents are often called 'standards' because their inclusion and consideration allow for comparisons of methods, equipment, and protocols. Microbiome standards are imperative for microbial community profiling and analysis. Whereas the microbial compositions of experimental samples are variable and often unknown, microbiome standards provide a common, accurate, and consistent measurement as a basis for comparison. By providing a common control to measure and evaluate performance, microbiome standards indicate biases allowing users to verify and optimize methods, enable inter-lab comparisons, and ensure reproducibility.

How to Select the Appropriate Microbiome Controls

The principle of a microbiome standard is simple: use a well characterized, quantified, and known microbial input to perform experimental procedures and evaluate consistency of the output. Standards can then be run as a parallel quality control to experimental samples to evaluate the consistency of the method. The resulting profile provides a basis to calibrate and when needed, begin troubleshooting. Several different types of NGS Microbiome controls are available, each detecting different and sometimes overlapping parts of the complex microbiome processing workflow. This article is meant to aid in selecting the appropriate reference reagents and controls for your microbiome experiments.

Mock Communities, True Diversity Reference, and Spike-in Controls

Several categories of microbiome reference reagents are available including microbial mock communities, true diversity reference material, and spike-in controls. Each category has overlapping characteristics, such as the use as positive controls, and each detects different biases throughout the microbiome analysis workflow.

Table 1 – Microbiome Standards and Controls Suggested Use

Standards	Suggested Applications
Mock Community Standards (Cellular)	
ZymoBIOMICS™ Microbial Community Standard	• General optimization and benchmarking • Positive control for microbial lysis
ZymoBIOMICS™ Gut Microbiome Standard	• General optimization and benchmarking for gut microbiome workflows • Assess cross-kingdom, strain-level resolution, and pathogen detection
ZymoBIOMICS™ Microbial Community Standard II (Log Distribution)	• Assessing detection limit of whole workflows beginning with DNA extractions
Mock Community Standards (DNA)	
ZymoBIOMICS™ Microbial Community DNA Standard	• Optimization and positive control for library preparation and bioinformatics
ZymoBIOMICS™ HMW DNA Standard	• Optimization and positive control for long-read sequencing library preparation and bioinformatics
ZymoBIOMICS™ Microbial Community DNA Standard II (Log Distribution)	• Assessing detection limits of library preparation and bioinformatics
True Diversity Reference	
ZymoBIOMICS™ Fecal Reference with TruMatrix™ Technology	• Assessing taxonomic assignment and bioinformatic processing parameters • Enable inter-lab and inter-study data comparisons
Spike-In Controls	
ZymoBIOMICS™ Spike-in Control I (High Microbial Load)	• *In situ* extraction control and absolute quantification for high biomass samples
ZymoBIOMICS™ Spike-in Control II (Low Microbial Load)	• *In situ* extraction control and absolute quantification for low biomass samples

The categories of microbiome standards and suggested applications are listed in Table 1.

Mock communities are accurately quantified and well-defined artificial microbial communities that act as ground truths of known composition and abundance. On the other hand, a true diversity reference is created from a specified natural source, such as human stool, stabilized and homogenized to be a common and consistent control material containing a true-to-to life microbial profile and diversity. Finally, while mock communities and true diversity references are meant to be used in parallel to experimental samples, spike-in controls are added directly to experimental samples and processed within each sample. The defined abundance of the spike-ins' unique species allows for absolute cell number quantification and quality control for each individual sample.

Cellular Mock Community Standards

Mock communities generated from whole cells are the most commonly used microbiome standard because they function as positive controls for the entire workflow. But perhaps more importantly, cellular mock communities such as the ZymoBIOMICS™ Microbial Community Standard are used to optimize and compare microbial lysis methods[4-5] because they contain equal abundances of species with a wide range of cell wall recalcitrance and cell size. By comparing the resulting profile to the theoretical profile, the ability of the lysis method can be assessed. For example, if the Gram-negative bacteria in the mock community profile are observed to be in excess while the Gram-positive bacteria are deficient compared to the theoretical abundance, the lysis method may struggle to break open thicker cell walls.

Additionally, site-specific microbial standards are another type of mock communities with their own uses. For example, the ZymoBIOMICS™ Gut Microbiome Standard contains 21 microbial strains from 3 kingdoms to allow for the evaluation of methods analyzing the gut microbiome and to act as a general positive control[6-7].

Finally, log-distributed mock community standards, such as the ZymoBIOMICS™ Microbial Community Standard II (Log Distribution), contain species at different abundances ranging from 10^2 – 10^8 cells per prep. This logarithmic

Table 2 – ZymoBIOMICS™ Standards, References, and Controls

Application	Mock Community (Cellular)			Mock Community (DNA)			True Diversity Reference	Spike-In Controls	
	ZymoBIOMICS™ Microbial Community Standard	ZymoBIOMICS™ Microbial Community Standard II (Log Distribution)	ZymoBIOMICS™ Gut Microbiome Standard	ZymoBIOMICS™ Microbial Community DNA Standard	ZymoBIOMICS™ Microbial Community DNA Standard II (Log Distribution)	ZymoBIOMICS™ HMW DNA Standard	ZymoBIOMICS™ Genuine Fecal Standard with TruMatrix™ Technology	ZymoBIOMICS™ Spike-in Control I (High Microbial Load)	ZymoBIOMICS™ Spike-in Control II (Low Microbial Load)
	D6300	D6310	D6331	D6305	D6311	D6322	D6323	D6320	D6321
General Microbiome Samples	✓			✓					
Fecal Samples	✓		✓	✓			✓	✓	
Assessing Detection Limit		✓			✓				
Long Read Sequencing	✓		✓			✓			
High Diversity							✓		
Internal Spike-ins								✓	✓
Targeted (16S, ITS) Sequencing	✓	✓	✓	✓	✓	✓	✓	✓	✓
Metagenomic (Shotgun) Sequencing	✓	✓	✓	✓	✓	✓	✓	✓	✓

distribution of species enables users to evaluate the detection limits of their microbiome analysis workflow[8].

DNA Mock Community Standards

Mock community standards made with purified microbial genomic DNA are more often used to detect biases and as optimization tools because they are utilized as input for library preparation rather than at the beginning of the workflow. DNA mock community standards such as the ZymoBIOMICS™ Microbial Community DNA Standard can be utilized to control biases associated with library prep and bioinformatics[9-10]. The optimization can be focused on library prep by first aligning NGS reads generated from the standard only to the genomes within the standard. After library prep has been optimized, the bioinformatics pipeline can be evaluated by aligning NGS reads against an entire reference database.

Similar to the cellular version, log distributed DNA standards, such as the ZymoBIOMICS™ Microbial Community DNA Standard II (Log Distribution), are used to assess detection limits but for library prep and bioinformatics pipelines.

Furthermore, an emerging technology for metagenomic analysis and genome assembly is long-read sequencing, often referred to as 3rd gen sequencing. Critical to long-read sequencing library prep and bioinformatics is high molecular weight DNA. The ZymoBIOMICS™ HMW DNA Standard is the only commercially available high molecular weight mock community, and has been used to evaluate sequencing chemistries and bioinformatic tools for long read sequencing[11-12].

True Diversity Reference

A true diversity reference is control material from a specified natural source that contains a complete, unchanging microbiome. In contrast to mock communities which have a quantified, known, and defined composition, the microbial composition of a true diversity reference is naturally derived. The ZymoBIOMICS™ Fecal Reference with TruMatrix™ Technology* is the first commercially available true diversity reference stabilized for long-term and lot-to-lot consistency. This reference features the high microbial diversity of a real fecal sample as well as a wide range of abundance.

Run-to-run and user-to-user consistency can be assessed on the same sample for each experiment. Reference materials can also be used to test system suitability by challenging experimental methods with actual source material. Bioinformatic analysis and taxonomy assignment are challenged with the added complexity of an unchanging true diversity sample. Since the microbial composition is static, the abundance and composition are stable and therefore allow users to assess method and analysis consistency.

Spike-in Controls

Unlike mock communities and true diversity references, spike-in controls offer different functions when added directly to experimental samples. The ZymoBIOMICS™ Spike-in Controls are composed of very unique species, alien to the human microbiome as well as many others. This enables them to be spiked into samples without interfering with the native microbiome. The defined composition of these species enables the quantification of the absolute cell number within the unknown sample, when analyzed with NGS-based microbiome methods. Furthermore, an emerging use of these spike-in controls is as *in situ* quality controls, meaning that it can be used as a positive control for every sample rather than a positive control for a whole run. This is very useful for NGS-based pathogen diagnosis.

Two spike-in controls are available for different sample types. The ZymoBIOMICS™ Spike-in Control I (High Microbial Load) is meant for high biomass samples such as stool. The ZymoBIOMICS™ Spike-in Control II (Low Microbial Load) is meant for low microbial biomass samples such as sputum and bronchoalveolar lavage (BAL) fluid.

Choosing a Microbiome Standard

The past several years has seen an explosion in the demand for microbiome standards, controls, and references that provide different and specific utilities. The scientists at Zymo Research share a passion for creating and providing the world with tools to improve microbiome data accuracy and reproducibility. As a result, the ZymoBIOMICS™ line of standards, references, and controls provides a range of utility for various microbiome applications. Additional information about the standards and applications can be found in Table 2.

*TruMatrix™ is a trademark of The BioCollective.

References

1. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S. Microbiome Quality Control Project C et al: Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017; 35(11): 1077–86.
2. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017; 35(11): 1069–76.
3. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Frontiers in Microbiology. 2016; 7:459.
4. Bartolomaeus TUP, Birkner T, Bartolomaeus H, Löber U, Avery EG, Mähler A, Weber D, Kochlik B, Balogh A, Wilck N, Boschmann M, Müller DN, Markó L, Forslund SK. Quantifying technical confounders in microbiome studies. Cardiovascular Research. 2021;17(3): 863-875.
5. Ojo-Okunola A, Claassen-Weitz S, Mwaikono KS, Gardner-Lubbe S, Zar HJ, Nicol MP, du Toit E. The Influence of DNA Extraction and Lipid Removal on Human Milk Bacterial Profiles. MDPI Methods and Protocols. 2020; 3(2): 39
6. Zhang B, Brock M, Arana C, Dende C, van Oers NS, Hooper LV, Raj P. Impact of bead-beating intensity on the genus and species level characterization of gut microbiome using amplicon and complete 16S rRNA gene sequencing. Frontiers in Cellular and Infection Microbiology. 2021; 11: 678522
7. Palkova L, Tomova A, Repiska G, Babinska K, Bokor B, Mikula I, Minarik G, Ostatnikova D, Soltys K. Evaluation of 16S rRNA primer sets for characterisation of microbiota in paediatric patients with autism spectrum disorder. Nature Scientific Reports. 2021; 11: 6781
8. Nicholls SM, Quick JC,Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. GigaScience. 2019; 8(5): giz043
9. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R, Albertsen M. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nature Methods. 2021; 18: 165-169.
10. Holm JB, Humphrys MS, Robinson CK, Settles ML, Ott S, Fu L, Yang H, Gajer P, He X, McComb E, Gravitt PE, Ghanem KG, Brotman RM, Ravel J. Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform. mSystems. 2019; 4(1): e00029-19
11. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M. Oxford Nanopore R10.4 long-read sequencing enables near-perfect bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. bioRxiv. 2021
12. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nature Biotechnology. 2021; 39: 442-450

Learn More About the ZymoBIOMICS™ Microbiome Standards at
www.zymoresearch.com/pages/microbiome-standards

ZYMO RESEARCH
The Beauty of Science is to Make Things Simple®

www.zymoresearch.com info@zymoresearch.com Toll Free: (888) 882-9682