Standardizing Microbiomics: Removing Bias and Errors in Collection, Purification and Analyses

Ryan Kemp, Shuiquan Tang, Elinne Becket, Larry Jia, Marc Van Eden, Standa Forman, Eric Kircher, John Sherman, Luigi Basillio, Michelle Thai, Steven Wong, and Mikayla Mager
Zymo Research Corp., Irvine, CA, USA

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

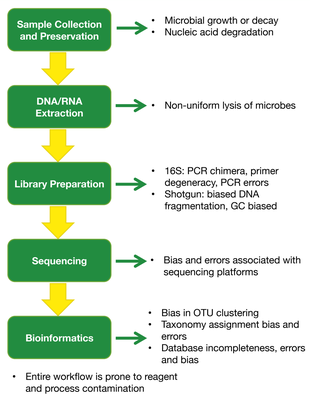
OZYME
Des femmes et des hommes au service de vos recherches

Introduction

The Need For Microbiomics Reference Materials and Standards

Microbial composition profiling techniques powered by Next-Generation sequencing (NGS) are becoming routine studies. However, it is well known that results are prone to bias and errors in every step of the workflow, including sample collection, DNA/RNA extraction, library preparation, sequencing and bioinformatics analysis. Therefore, standardization of these NGS workflows is necessary to minimize bias and improve data quality and reproducibility in microbiomics studies. The figure on the right shows challenges associated with each step.

At Zymo Research, we have constructed a comprehensive product line, coined ZymoBIOMICS™, designed for unbiased, accurate microbiomics and metagenomics analyses and validated using the ZymoBIOMICS™ Microbial Community/DNA Standards.



ZymoBIOMICS™ is a Total Solution for your Microbiomics Workflows

Methods & Results

Construction and Characterization of the ZymoBIOMICS™ Microbial Community Standards

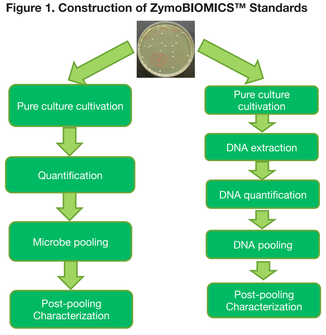
In order to assess bias and errors in NGS-based microbial composition profiling workflows, reference materials that mimic a mixed microbial community of well-defined composition are necessary.

The ZymoBIOMICS™ Microbial Community/DNA Standards contain both tough-to-lyse and easy-to-lyse microbes – both prokaryotes and eukaryotes – that span a wide GC range (15% - 85%; Table 1) and contain negligible impurities (<0.01%, Table 2) making it ideal for optimizing 16S rRNA gene seq. and metegenomic seq. workflows. The standards are constructed by pure culture cultivation (Figure 1), cell counting/quantification via Qubit™, pooling of the organisms according to the predefined ratios and post pooling characterization using shotgun sequencing (Figure 2). Sequencing results of each lot are reported in a certificate of analysis.

Table 1. Microbial Composition of ZymoBIOMICS™ Standards

Species	Avg GC (%)	Gram Stain	gDNA Abun. (%)
Pseudomonas aeruginosa	66.2	-	12
Escherichia coli	56.8	-	12
Salmonella enterica	52.2	-	12
Lactobacillus fermentum	52.8	+	12
Enterococcus faecalis	37.5	+	12
Staphylococcus aureus	32.7	+	12
Listeria monocytogenes	38.0	+	12
Bacillus subtilis	43.8	+	12
Saccharomyces cerevisiae	38.4	Yeast	2
Cryptococcus neoformans	48.2	Yeast	2

Figure 1. Construction of ZymoBIOMICS™ Standards

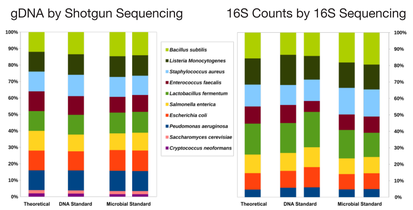


ZymoBIOMICS™

Negligible Impurities (<0.01%)

Table 2. Assessing impurities in ZymoBIOMICS™ Microbial Community Standard using shotgun metagenomic sequencing. Genomic DNA was extracted using ZymoBIOMICS™ DNA Mini kit. Shotgun library was prepared with an internal method. The library was sequenced on HiSeq, resulting in ~178 million reads (100 bp). Taxonomy identification was performed with mOTU (http://www.bork.embl.de/software/mOTU).

Species	mOTU Counts	mOTU Abun. (%)
Bacillus subtilis	9048	11.86
Listeria monocytogenes	11454	15.01
Staphylococcus aureus	7960	10.43
Enterococcus faecalis	11322	14.84
Lactobacillus fermentum	17081	22.39
Escherichia coli	6994	9.17
Salmonella enterica	7939	10.41
Pseudomonas aeruginosa	4484	5.88
Propionibacterium acnes	1	0.0013

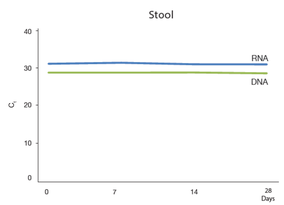


Accurate composition

Figure 2. Assessing microbial composition of ZymoBIOMICS™ standards using shotgun metagenomic sequencing and 16S rRNA gene targeted sequencing. Genomic DNA from the microbial standard was extracted using ZymoBIOMICS™ DNA Mini Kit. Shotgun library was prepared with an internal method and sequenced on MiSeq. Microbial composition was determined by mapping raw reads to the 10 microbial genomes contained in the standard. The library of 16S rRNA targeted sequencing was prepared with primers that amplify 16S v3-4 region and sequenced on MiSeq (2x250 bp). Paired-end reads were joined to construct complete amplicon sequences, which were mapped to the 8 bacterial 16S sequences contained in the standard to determine relative abundance.

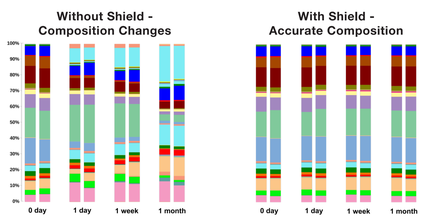
Eliminating Bias in Sample Collection

Sample collection can cause significant bias or lead to erroneous results due to organism growth/decay, nucleic acid degradation, and reagent removal. Ethanol, RNAlater, cryopreservatives and freeze-thawing have all been shown to lead to composition bias. At collection, DNA/RNA Shield™ stabilizes nucleic acids (Figure 3) and inactivates all organisms — including pathogens (Figure 4) — in your sample, allowing for accurate community profiling (Figure 5). There is no degradation of nucleic acids, creating a perfect molecular snapshot of your sample at the time of collection.



Nucleic Acid Stabilization at Ambient Temperatures

Figure 3. DNA and RNA in stool is effectively stabilized in DNA/RNA Shield™ at ambient temperature. DNA and RNA spike-in controls were added to stool, purified at the indicated time points, and were analyzed by (RT)qPCR. Controls: HSV-1 and HIV (AcroMetrix, Life Technologies).

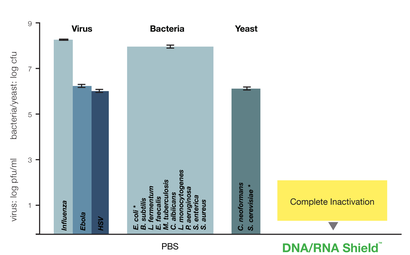


Preserve Microbial Composition

Figure 4. Microbial composition of stool is unchanged after one month at ambient temperature with DNA/RNA Shield™. Stool samples suspended in DNA/RNA Shield™ and stored at room temperature were compared to stool without preservative for one month. They were sampled at the indicated time points and processed with ZymoBIOMICS™ DNA Mini Kit. The extracted DNA was then subjected to microbial composition profiling via 16S rRNA gene targeted sequencing. Samples stored with DNA/RNA Shield™ had a constant microbial composition while the samples stored without shifted dramatically.

Microbial Inactivation

Figure 5. Viruses, bacteria and yeast are effectively inactivated by DNA/RNA Shield™. Samples containing the infectious agent (virus, bacteria, yeast) were treated with DNA/RNA Shield™ or mock (PBS) treated for 5 minutes. Titer (PFU) was subsequently determined by plaque assay. Validated by: Influenza A - D. Poole and Prof. A. Mehle, Department of Medical Microbiology and Immunology, University of Wisconsin, Madison; Ebola (Kikwit) - L. Avena and Dr. A. Griffiths, Department of Virology and Immunology, Texas Biomedical Research Institute; HSV-1/2 - H. Oh, F. Diaz and Prof. D. Knipe, Virology Program, Harvard Medical School; E. coli, L. fermentum, B. subtilis, S. cerevisiae – Zymo Research Corporation).



DNA Extraction

Eliminating Bias in DNA Extraction Using the ZymoBIOMICS™ DNA Mini Kit and Standards

A significant contributor to bias in the microbiomic workflow is incomplete lysis of a sample. Chemical, enzymatic, and many lysis matrices lead to inferior lysis and an unrealistic representation of the microbial community. Mechanical lysis offers the most accurate representation of a microbial community. However, not all mechanical lysis protocols are created equal. Figure 6 depicts how four different extraction methods, including some of the most cited, can lead to four different profiles, with only one method representing the true community profile. A large source of error for mechanical lysis arises from the inability to lyse a wide array of organisms that range in size and hardiness, including bacteria, yeast, and spores. The ZymoBIOMICS™ DNA Mini Kit enables accurate community profiling by providing unbiased extraction of DNA from any sample (e.g. feces, soil, water, biofilms) that is ultra-pure and inhibitor-free.

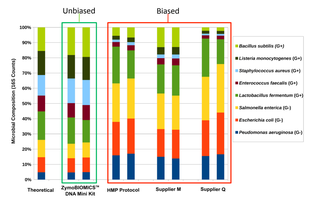


Figure 6. Assessing bias in DNA extraction protocols using ZymoBIOMICS™ Microbial Community Standard. Genomic DNA was extracted from 75 µL of the microbial standard by strictly following each protocol. The microbial composition was profiled using 16S rRNA gene targeted sequencing in the same way as described in Figure 2.

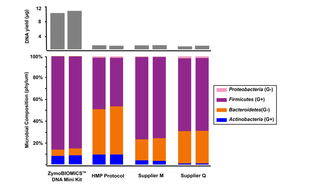


Figure 7. Comparing DNA extraction protocol s using human fecal samples. Genomic DNA was extracted from 200 µL 10% (w/v) fecal sample suspended in PBS by strictly following each protocol. The microbial composition was profiled using 16S rRNA gene targeted sequencing in the same way as described in Figure 4. DNA concentration was determined with Qubit.

Library Prep and Sequencing

Detecting and Eliminating Artifacts in Library Preparation

Bioinformatics and Analysis

Eliminate artifacts in 16S rRNA gene targeted sequencing

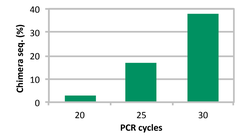


Figure 8. Assessing PCR chimera formation in library preparation of 16S rRNA gene targeted sequencing. Each PCR contained 10 ng genomic DNA from ZymoBIOMICS™ Microbial Community DNA Standard as template. Primers targeting 16S v3-4 region were used. The libraries were sequenced on MiSeq (2x250bp). Paired-end reads were assembled into complete amplicons. Chimeric sequences were identified with Uchime and using the rRNA genes sequences present in the standard as references.

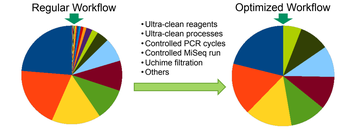


Figure 9. Comparing the performance of two different workflows of 16S rRNA gene targeted sequencing using ZymoBIOMICS™ Microbial Community DNA Standard. The pie chart on the left shows the genus-level microbial composition of the standard sequenced and analyzed using a regular workflow. The pie chart on the right shows the corresponding profile sequenced and analyzed using an optimized workflow. While the pie chart on the left shows the presence of 43 bacterial genera, most of which we believe were caused by artifacts like PCR chimera, process contaminations, the pie chart on the right only shows the presence of 8 bacterial genera, which agrees with the real composition of the standard.

Eliminate GC bias in shotgun metagenomic sequencing

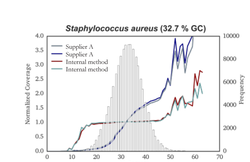


Figure 10. Our internal library preparation process has no GC bias. Shotgun metagenomics sequencing with two different library preparation process, Supplier A and an internal method, was conducted. Barplots depict the GC content distribution in each genome. Normalized coverage was defined as the sequencing coverage divided by the average sequencing coverage of the genome. The two genomes from the ZymoBIOMICS™ standard was chosen to present the cases of low GC content and high GC content. Sequencing was performed on MiSeq (2x150 bp).

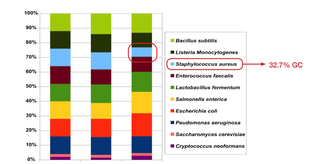


Figure 11. Assessing bias of two different library preparation processes in shotgun metagenomic sequencing using ZymoBIOMICS™ Microbial Community Standard. Compared to our internal method, the Supplier A kit from Illumina has some bias due to GC content variation. Sequencing was performed on MiSeq (2x150 bp).

Conclusion

- The use of a reference material is necessary to develop a workflow that is unbiased, reproducable, and accurate from sample collection to analyses.
- The ZymoBIOMICS™ product line offers a complete and unbiased solution to microbiomics and metagenomics studies that was validated using the ZymoBIOMICS™ Standards.