ZYMO RESEARCH

Microbiomics
Made Simple™

ZymoBIOMICS® Microbial Community Standard

Assess bias and errors in NGS-based microbial composition profiling workflows

Highlights

- **Accurate composition:** allows for benchmarking and validation of NGS microbiome workflows.
- **Assess bias in DNA isolation:** contains microbes of varying size and cell wall recalcitrance (8 bacteria and 2 yeasts).
- **Quality control:** ideal control for microbiome profiling quality.

Catalog Number:
D6300

Scan with your smart-phone camera to view the online protocol/video.

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

Table of Contents

Product Contents

ZymoBIOMICS® Microbial Community Standard	D6300 (10 preps.)	Storage Temperature[1]
ZymoBIOMICS® Microbial Community Standard	0.75 ml	-80°C

Specifications

- **Source –** eight bacteria (three Gram-negative and five Gram-positive) and two yeasts.
- **Biosafety –** this product is not biohazardous as microbes have been fully inactivated.
- **Reference Genomes and 16S & 18S rRNA Genes[2] –** https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip
- **Storage Solution –** DNA/RNA Shield™ (R1100-50)
- **Total Cell Concentration –** ~1.4 x 10^{10} cells/ml
- **Impurity Level –** <0.01% foreign microbial DNA
- **Relative Abundance Deviation in Average –** <15%
- **Microbial Composition –** Table 1 shows the theoretical microbial composition of the standard.

The microbial composition of each lot was measured by shotgun metagenomic sequencing post mixing. The results (including the composition, impurities and abundance deviation) can be accessed through the Certificate of Analysis based on the lot number (printed on tube level) by the following link: https://www.zymoresearch.com/pages/certificate-of-analysis.

1 For short-term storage or regular use, -20°C may be used.
2 Several strains within the standard were replaced with similar strains beginning from Lot ZRC190633. This update will not affect the species composition of the standard. Refer to Appendix C to check if your product is from an older lot and find the correct reference database to use accordingly if needed.

Table 1. Microbial Composition

Species	Theoretical Composition (%)				
	Genomic DNA	16S Only[1]	16S & 18S[1]	Genome Copy[2]	Cell Number[3]
Pseudomonas aeruginosa	12	4.2	3.6	6.1	6.1
Escherichia coli	12	10.1	8.9	8.5	8.5
Salmonella enterica	12	10.4	9.1	8.7	8.8
Lactobacillus fermentum	12	18.4	16.1	21.6	21.9
Enterococcus faecalis	12	9.9	8.7	14.6	14.6
Staphylococcus aureus	12	15.5	13.6	15.2	15.3
Listeria monocytogenes	12	14.1	12.4	13.9	13.9
Bacillus subtilis	12	17.4	15.3	10.3	10.3
Saccharomyces cerevisiae	2	NA	9.3	0.57	0.29
Cryptococcus neoformans	2	NA	3.3	0.37	0.18

[1] The theoretical composition in terms of 16S (or 16S & 18S) rRNA gene abundance was calculated from theoretical genomic DNA composition with the following formula: 16S/18S copy number = total genomic DNA (g) × unit conversion constant (bp/g) / genome size (bp) × 16S/18S copy number per genome. Use this as reference when performing 16S targeted sequencing.

[2] The theoretical composition in terms of genome copy number was calculated from theoretical genomic DNA composition with the following formula: genome copy number = total genomic DNA (g) × unit conversion constant (bp/g) / genome size (bp). Use this as reference when inferring microbial abundance from shotgun sequencing data based on read depth/coverage.

[3] The theoretical composition in terms of cell number was calculated from theoretical genomic DNA composition with the following formula: cell number = total genomic DNA (g) × unit conversion constant (bp/g) / genome size (bp)/ploidy.

Product Description

ZymoBIOMICS® Microbial Community Standard is a mock microbial community consisting of eight bacterial and two fungal strains. It includes three easy-to-lyse Gram-negative bacteria (e.g. *Escherichia coli*), five tough-to-lyse Gram-positive bacteria (e.g. *Listeria monocytogenes*), and two tough-to-lyse yeasts (e.g. *Cryptococcus neoformans*) (Table 1). Seven of these strains are known human pathogens and have been fully inactivated with DNA/RNA Shield™ (R1100-50). The GC content[1] of the contained genomes covers a wide range from 15% to 85%. The standard was constructed by pooling pure cultures of the ten microbial strains. The cells and DNA content of each pure culture were quantified before pooling. Cultures were mixed based on a predetermined composition (Table 1).

The microbial standard is accurately characterized and is guaranteed to contain <0.01% impurity. This enables it to be used to expose artifacts, errors, and bias in microbiomics or metagenomics workflows. Serving as a defined input from the beginning, this standard can be used to guide construction and optimization of entire workflows or as a quality control tool for inter-lab studies. Benchmarking with this standard, we found that most cited DNA extraction methods currently used in the field, including the Human Microbiome Project fecal DNA extraction protocol, are dramatically biased (Figure 1).

Details regarding the ten microbial strains (including species name, genome size, ploidy, average GC content, 16S/18S copy number, phylogeny) can be found in Table 2. The 16S/18S rRNA sequences (FASTA format) and genomes (FASTA format) of these strains[2] are available from the link below. Feel free to contact us if we can help to analyze sequencing data generated from this standard.

Reference Genome Download:
https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip.

1 GC content can cause bias of sequencing coverage in PCR-based library prep workflows of shotgun sequencing.
2 Several strains within the standard were replaced with similar strains beginning from Lot ZRC190633. This update will not affect the species composition of the standard. Refer to Appendix C to check if your product is from an older lot and find the correct reference database to use accordingly if needed.

Background on the Need for Microbiome Standards

Microbial composition profiling techniques powered by next-generation sequencing are becoming routine in microbiomics and metagenomics studies. It is well known that these analytical techniques can suffer from bias and errors in every step of the workflow, including DNA extraction, library preparation, sequencing, and bioinformatics analysis. To assess the performance of different microbiomics workflows, there is an urgent need in the field for reliable reference materials, *e.g.* a mock microbial community with defined composition.

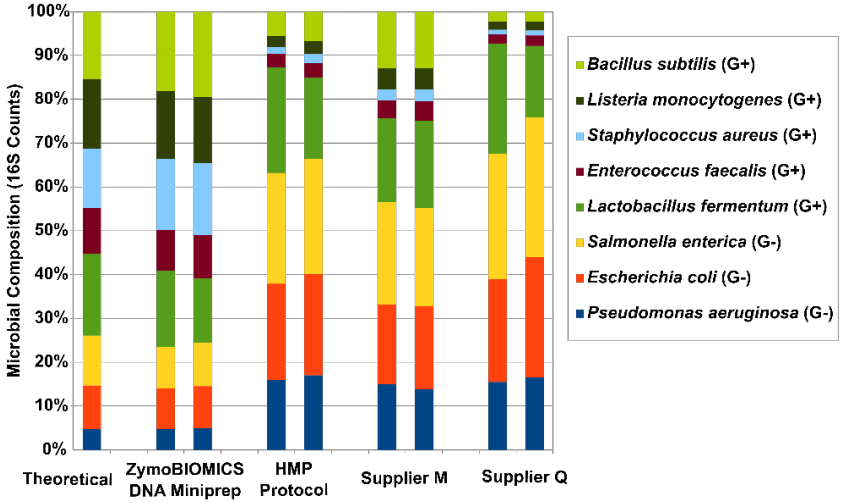

Figure 1. Comparing DNA extraction processes with ZymoBIOMICS® Microbial Community Standard. DNA was extracted from the ZymoBIOMICS® standard using four different DNA extraction methods (ZymoBIOMICS® DNA Kit, Human Microbiome Project fecal DNA extraction protocol, a DNA extraction kit from Supplier M, and a fecal DNA extraction kit from Supplier Q) and analyzed using 16S rRNA gene sequencing. 16S rRNA genes were amplified with primers targeting V3-4 region and the amplicons were sequenced on Illumina® MiSeq™ (2x250bp). Overlapping paired-end reads were assembled into complete amplicon sequences. The composition profile was determined based on sequence counts after mapping amplicon sequences to the known 16S rRNA genes of the eight different bacterial strains contained in the standard. Only the ZymoBIOMICS® DNA kit provides unbiased profiles in this study.

Table 2. Strain Information

Species	NRRL Accession NO.	Genome Size (Mb)	Ploidy	GC Content (%)	16/18S Copy Number	Gram Stain
Pseudomonas aeruginosa	B-3509	6.792	1	66.2	4	-
Escherichia coli	B-1109	4.875	1	46.7	7	-
Salmonella enterica	B-4212	4.760	1	52.2	7	-
Lactobacillus fermentum	B-1840	1.905	1	52.4	5	+
Enterococcus faecalis	B-537	2.845	1	37.5	4	+
Staphylococcus aureus	B-41012	2.730	1	32.9	6	+
Listeria monocytogenes	B-33116	2.992	1	38.0	6	+
Bacillus subtilis	B-354	4.045	1	43.9	10	+
Saccharomyces cerevisiae	Y-567	12.1	2	38.3	109[3]	Yeast
Cryptococcus neoformans	Y-2534	18.9	2	48.3	60[3]	Yeast

Table 2, *continued*

Species	NCBI Phylogeny Database
Pseudomonas aeruginosa	*Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas; Pseudomonas aeruginosa* group
Escherichia coli	*Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia*
Salmonella enterica	*Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella*
Lactobacillus fermentum	*Bacteria; Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus*
Enterococcus faecalis	*Bacteria; Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus*
Staphylococcus aureus	*Bacteria; Firmicutes; Bacilli; Bacillales; Staphylococcaceae; Staphylococcus*
Listeria monocytogenes	*Bacteria; Firmicutes; Bacilli; Bacillales; Listeriaceae; Listeria*
Bacillus subtilis	*Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus subtilis* group
Saccharomyces cerevisiae	*Eukaryota; Opisthokonta; Fungi; Dikarya; Ascomycota; saccharomyceta; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces*
Cryptococcus neoformans	*Eukaryota; Opisthokonta; Fungi; Dikarya; Basidiomycota; Agaricomycotina; Tremellomycetes; Tremellales; Tremellaceae; Filobasidiella; Filobasidiella/Cryptococcus neoformans* species complex

3 18S rRNA gene copy numbers in a haploid genome of the two strains of *Saccharomyces cerevisiae* and *Cryptococcus neoformans* were estimated based on read depth information from mapping shotgun sequencing data.

Protocol

1. Thaw the standard on ice. Once thawed, vortex the standard for at least 30 seconds, then spin down briefly.

***Note**: Cells might aggregate due to freeze-thaw cycling; therefore, it is critical to mix the cellular standard thoroughly before use.*

2. For DNA extraction of the standard, use 75 µl per prep[1]. We recommend using mechanical lysis featured in Zymo Research's microbial DNA isolation kits[2] for unbiased and efficient isolation. Expected yield is approximately 2 µg[3] DNA per prep when using **ZymoBIOMICS® DNA Miniprep Kit** (D4300).

***Note**: The duration of homogenization (bead beating) will vary depending on the homogenization device, and may require optimized by the end-user. Zymo Research has validated optimized lysis parameters for many common homogenization devices, which can be found here:* https://files.zymoresearch.com/documents/bead_beating_short_protocol_tables.pdf

1. For use with kits that are incompatible with DNA/RNA Shield and other sample preservation reagents, such as Qiagen's PowerFecal and PowerSoil kits, reducing the input volume can improve compatibility. For kits such as these, it is recommend to reduce the input of this standard to ≤ 20 µl.
2 This microbial standard contains several tough-to-lyse microbes; therefore, to extract DNA from this standard, we strongly recommend using ZymoBIOMICS® DNA Miniprep Kit (#D4300), Quick-DNA™ Fungal/Bacteria DNA Miniprep Kit (#D6005), Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (# D6010). These kits feature a unique lysis matrix that contains our ultra-high-density BashingBeads™, which provides unbiased lysis of bacteria and fungi for accurate microbial composition profiling.
3 The expected yield for one prep (75 µl) of the standard is approximately 2 µg. Yields significantly lower than 2 µg may suggest inefficient lysis during DNA extraction.

Appendices

Appendix A: Bioinformatics Analysis Recommendations

Assessing accuracy of taxonomy identification

A fundamental goal in microbiome studies is to identify what microbes are present in a sample. After analyzing this microbiome standard using a workflow that includes wet-lab processing and dry-lab interpretation, the taxa identified can be compared with the taxonomy information of the ten strains included in the standard (Table 2). This allows a performance assessment of a workflow regarding the limit of the taxonomy resolution, false positives, and false negatives. False positives can be caused by contaminations from wet-lab processes, chimeric sequences during library prep, sequencing errors, demultiplexing errors and defects in bioinformatics analysis. We certify that the impurity level of the standard is <0.01% (by DNA abundance). Therefore, it can be concluded that any alien taxa present at >0.01% (by DNA abundance) in the standard is introduced artificially by the user's workflow. The detection limit of a workflow can be easily determined by checking what strains are detected in the microbiome standard as their abundance follows log distribution.

Assessing bias in composition profiling

To assess composition bias, compare the composition profile determined by the user's workflow to the defined composition shown in Table 1. Both wet-lab and dry-lab processes can introduce bias. To determine the quality of a wet-lab process, an accurate/unbiased dry-lab analysis method is needed to interpret the sequencing data from the standard. A straightforward and accurate method to infer the microbial composition from sequencing data of our microbiome standard is through direct read-mapping against reference genomes (or against reference 16S & 18S sequences in the case of targeted sequencing). The reference sequences of this microbiome standards can be found in the Specifications.

Note: *Bacterial strains that are phylogenetically distant can potentially share highly similar sequences in their genomes, e.g. ribosomal RNA sequences and conserved single-copy genes. In the process of direct read mapping, the presence of these highly homologous regions can cause reads that are derived from high-abundance microbes to be assigned to low-abundance microbes, resulting in the overestimation of low-abundance microbes in the standard. One way to overcome this issue is to use a mapping tool that can choose to ignore reads that map to more than one genome. Another way to address this problem is to filter these highly conserved sequences from the reference genomes. Please contact us if you need assistance.*

Appendix B: Additional Strain Information

Species	NRRL Accession NO.	Strain Name[1]
Bacillus subtilis	B-354	*Bacillus subtilis* (Ehrenberg 1835) Cohn 1872 ATCC 6633=NRRL B-209=NRS-231=PCI 219
Cryptococcus neoformans	Y-2534	*Cryptococcus deneoformans* T. Boekout & F. Hagen (2014) 32045=ATCC 32719=CBS 132=CCRC 20528=CCY 17-1-2=DBVPG 6010=IFO 0608=IGC 3957=NRRL Y-8347=PYCC 3957
Enterococcus faecalis	B-537	*Enterococcus faecalis* (Andrewes and Horder 1906) Schleifer and Kilpper-Bälz 1984 ATCC 7080
Escherichia coli	B-1109	Castellani and Chalmers 1919, 01485cm
Lactobacillus fermentum	B-1840	*Lactobacillus fermentum* Beijerinck 1901 19lc3=ATCC 14931=BCRC 12190=CCUG 30138=CECT 4007=CIP 102980=DSM 20052=IFO 15885=JCM 1173=KCTC 3112=LMG 6902=NBRC 15885=NCDO 1750=NCIMB 11840=NRIC 1752=NRRL B-4524.
Listeria monocytogenes	B-33116	*Listeria monocytogenes* (Murray et al. 1926) Pirie 1940 2847=ATCC 19117
Pseudomonas aeruginosa	B-3509	*Pseudomonas aeruginosa* (Schroeter 1872) Migula 1900 ATCC 15442=NCIB 10421=Pdd-10
Saccharomyces cerevisiae	Y-567	*Saccharomyces cerevisiae* Meyen ex E. C. Hansen (1883) ATCC 9763=CBS 2978=CBS 5900=CCY 21-4-48=CCY 21-4-54=NCTC 10716=NCTC 7239=NCYC 87=Pattee 6=PCI M-50
Salmonella enterica	B-4212	*Salmonella enterica* subspecies *enterica*, Castellani and Chalmers 1919, TA1536
Staphylococcus aureus	B-41012	*Staphylococcus aureus* Rosenbach 1884

1 The strain information was extracted from the website of the Agricultural Research Service Culture Collection (NRRL, https://nrrl.ncaur.usda.gov/).

Appendix C: Reference Sequences

We replaced five strains in the ZymoBIOMICS® standards (D6300, D6305 and D6306) with similar strains beginning with Lot ZRC190633 (Table 3 and Table 4). We apologize for any inconvenience that this update may cause.

Key Points:

- No further organism changes will occur; hence the strains will remain constant.
- The updated standards include 8 complete bacterial genomes and 2 draft yeast genomes.
- Species-level composition of the standards is unchanged.
- For analyses that require the reference genomes or sequences of the strains, please use the correct references as listed in the table below.

Table 3. Products Containing New Strains

Cat. #	Lot #	Product Name	Reference Genome and 16S/18S sequences
D6300	All current lots	ZymoBIOMICS® Microbial Community Standard	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip
D6305	All current lots	ZymoBIOMICS® Microbial Community DNA Standard (200ng)	
D6306	All current lots	ZymoBIOMICS® Microbial Community DNA Standard (2000ng)	

Table 4. Products Containing Old Strains

Cat. #	Lot #	Product Name	Reference Genome and 16S/18S sequences
D6300	ZRC183430 ZRC187326	ZymoBIOMICS® Microbial Community Standard	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.genomes.ZR160406.zip
D6305	ZRC183430	ZymoBIOMICS® Microbial Community DNA Standard (200ng)	
D6306	ZRC183430	ZymoBIOMICS® Microbial Community DNA Standard (2000ng)	

Ordering Information

Product Description	Catalog No.	Size
ZymoBIOMICS® Microbial Community Standard	D6300	10 preps.

Related Products	Catalog No.	Amount
ZymoBIOMICS® Microbial Community DNA Standard (200ng)	D6305	200 ng
ZymoBIOMICS® Microbial Community DNA Standard (2000ng)	D6306	2000 ng
ZymoBIOMICS® Microbial Community Standard II (Log Distribution)	D6310	10 preps.
ZymoBIOMICS® Microbial Community DNA Standard II (Log Distribution)	D6311	220 ng
ZymoBIOMICS® Spike-in Control I (High Microbial Load)	D6320 D6320-10	25 preps. 250 preps.
ZymoBIOMICS® Spike-in Control II (Low Microbial Load)	D6321 D6321-10	25 preps. 250 preps.
ZymoBIOMICS® HMW DNA Standard	D6322	5000 ng
ZymoBIOMICS® Gut Microbiome Standard	D6331	10 preps.

Complete Your Workflow

✓ To collect and transport microbiome samples at ambient temperatures:

DNA/RNA Shield™ and Collection Devices	
1X Reagent (R1100)	For sample lysis and stabilization of DNA/RNA
2X Concentrate (R1200)	Reagent concentrate (2X) for use with liquids at 1:1 ratio
Fecal Collection Tube (R1101)	15 mL container (prefilled with 9 mL DNA/RNA Shield™). Direct collection of up to 1g or 1 mL stool
Collection Tube w/ Swab (R1106)	12 x 80 mm screwcap container filled with 1 mL DNA/RNA Shield™ and sterile swab for specimen collection

✓ Unbiased and inhibitor-free DNA and RNA extraction (high-throughput and automatable) for microbiome profiling:

ZymoBIOMICS® DNA and RNA Kits	
DNA Miniprep (D4300)	Up to 25 µg DNA
DNA Microprep (D4301)	Up to 5 µg DNA
MagBead DNA (D4302)	Automatable (Tecan, Hamilton, Kingfisher, etc.)
96-Well DNA (D4309)	Spin-plate
DNA/RNA Miniprep Kit (R2002)	Up to 100 µg DNA/RNA

✓ Streamlined workflows with comprehensive bioinformatics analysis and publication-ready plots and figures:

ZymoBIOMICS® Services	
Targeted Sequencing Service 16S (Q2001)	With DNA Extraction
Targeted Sequencing Service 16S (Q2012)	Without DNA Extraction
Targeted Sequencing Service ITS (Q2003)	With DNA Extraction
Targeted Sequencing Service ITS (Q2003)	Without DNA Extraction

Notes

100% satisfaction guarantee on all Zymo Research products, or your money back.

Zymo Research is committed to simplifying your research with quality products and services. If you are dissatisfied with this product for any reason, please call 1(888) 882-9682.

Integrity of kit components is guaranteed for up to one year from date of purchase.
Reagents are routinely tested on a lot-to-lot basis to ensure they provide the highest performance and reliability.

This product is for research use only and should only be used by trained professionals. It is not for use in diagnostic procedures. Some reagents included with this kit are irritants. Wear protective gloves and eye protection. Follow the safety guidelines and rules enacted by your research institution or facility.

™ Trademarks of Zymo Research Corporation, ® Registered Trademarks of Zymo Research Corporation
Other trademarks: Disruptor Genie® (Scientific Industries, Inc.), FastPrep® (MP Biomedicals), Illumina® MiSeq™, (Illumina Inc.)

ZYMO RESEARCH

The **BEAUTY** *of* **SCIENCE** *is to Make Things* **SIMPLE**®

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