ZYMO RESEARCH

Microbiomics
Made Simple™

ZymoBIOMICS® Microbial Community DNA Standard

Assess bias and errors in NGS-based microbial profiling workflows

Highlights

- **Accurate composition:** allows for benchmarking and validation of NGS microbiome workflows.
- **Negligible impurity:** guaranteed to contain <0.01% foreign microbial DNA.
- **Wide range of GC content:** 15%-85%, for assessing bias caused by GC content variation.

Catalog Number:
D6305, D6306

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

Product Name	D6305	D6306	Storage Temperature[1]
ZymoBIOMICS® Microbial Community DNA Standard	200ng/20µl	2000ng/20µl	-20°C

Specifications

- **Source –** eight bacteria (three Gram-negative and five Gram-positive) and two yeasts.
- **Reference Genomes and 16S & 18S rRNA Genes[2] –** https://zymo-files.s3.amazonaws.com/BioPool/ZymoBIOMICS.STD.refseq.v3.zip
- **Storage Solution –** 10 mM Tris-HCl and 0.1 mM EDTA, pH 8.0
- **DNA Concentration –** 10 ng/µl (D6305); 100 ng/µl (D6306)
- **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 DNA Standard is a mixture of genomic DNA isolated from pure cultures of eight bacterial and two fungal strains. Genomic DNA from each pure culture was isolated and quantified before mixing[1]. The GC content[2] of the containing genomes covers a range from 15% to 85%. The microbial standard is accurately characterized and contains negligible impurities (< 0.01%). This enables it to be used to expose artifacts, errors, and bias in microbiomics or metagenomics workflows. This product is ideal for assessing biases and errors associated with library preparation, sequencing and bioinformatics analyses. It serves perfectly as a microbial standard for benchmarking the performance of microbiomics or metagenomics analyses or as a quality control tool for inter-lab studies. This standard is also ideal to help users construct and optimize workflows, e.g. assessing PCR chimera rate (Figure 1), and removing false positives (Figure 2) in 16S rRNA gene targeted sequencing, and assessing GC bias in sequencing coverage of shotgun metagenomic sequencing (Figure 3).

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[3] 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://zymo-files.s3.amazonaws.com/BioPool/ZymoBIOMICS.STD.refseq.v3.zip.

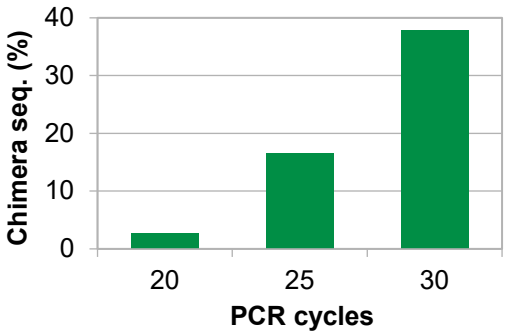

Figure 1. PCR chimera increases with increasing PCR cycle number in the library preparation process of 16S rRNA gene targeted sequencing. 20 ng ZymoBIOMICS® Microbial Community DNA Standard was used as a template. The PCR reaction was performed with primers that target V3-V4 region of the 16S rRNA gene. Chimera sequences were identified with Uchime (http://drive5.com/usearch) and using the 16S rRNA gene of the 8 bacterial strains contained in the standard as reference.

1 Genomic DNA from each culture was extracted and quantified before mixing so this DNA standard was independent and not a direct derivative of the microbial version, ZymoBIOMICS® Microbial Community Standard.
2 GC content can cause bias of sequencing coverage in PCR-based library prep workflows of shotgun sequencing.
3 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.

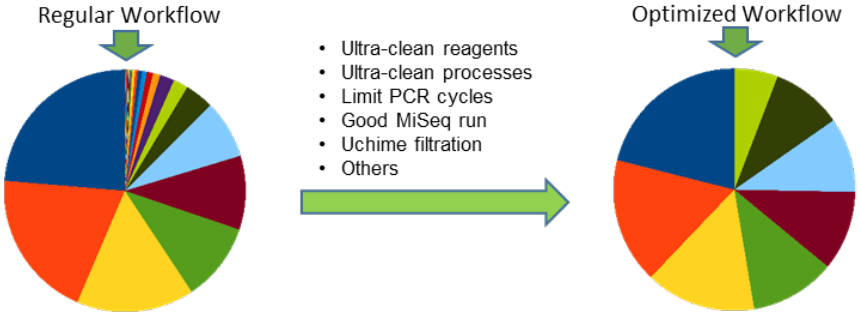

Figure 2. Eliminating noise/false positives in 16S rRNA gene targeted sequencing guided with ZymoBIOMICS® Microbial Community DNA Standard. The pie chart on the left is the microbial composition profile of the standard determined by a regular workflow of 16S sequencing using primers targeting 16S V3-V4 region. The pie chart on the right is the profile of the same standard determined using the same primer sets, but with an optimized in-house 16S sequencing workflow. Noise observed on the left panel were mainly caused by PCR chimera, process contamination, and reagent contamination, which were controlled in the optimized workflow. The accuracy of the standard's microbial composition is critical for revealing the presence of composition bias and false positives when optimizing a workflow.

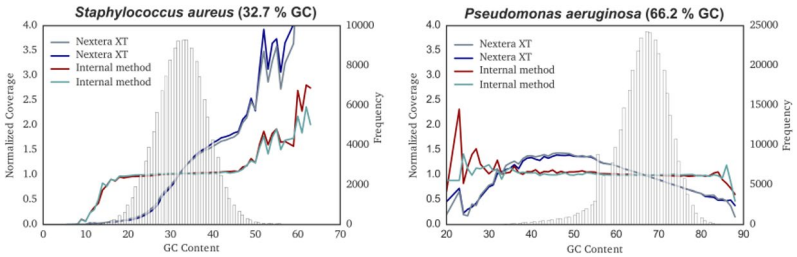

Figure 3. Assessing GC bias of two different library preparation methods in shotgun metagenomic sequencing. Library preparation for shotgun metagenomic sequencing was performed in two different ways: one by Illumina Nextera® XT kit and one by an in-house method. Shotgun sequencing was performed on MiSeq™ with paired-end sequencing (2x150 bp). Raw reads were mapped to the 10 microbial genomes to evaluate the potential effect of GC content on sequencing coverage. Normalized coverage was calculated by normalization with the average sequencing coverage of each genome. The coverage profiles of two selected genomes, *Staphylococcus aureus* and *Pseudomonas aeruginosa*, were picked to demonstrate as they cover a wide range of G/C content, 15%-85%. While the in-house method shows little or no G/C-bias, the Nextera® XT kit has reduced representation for both low GC and high GC regions.

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.

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[1]	Yeast
Cryptococcus neoformans	Y-2534	18.9	2	48.3	60[1]	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

1 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 from mapping shotgun sequencing data.

Protocol

1. Thaw the standard on ice. Once thawed, pulse vortex the standard, then centrifuge briefly to settle the liquid.

2. The amount of DNA used depends on the library preparation process being evaluated. Example quantities are shown below.

Table 3. Typical DNA Input for Different Library Prep Processes

Library Prep Method	DNA Input (ng)
16S Amplicon Libraries	10
Illumina DNA Prep	1
Illumina TruSeq® Nano	>200
Illumina TruSeq® PCR-free	2000
KAPA® HyperPlus	1 - 2000

3. To assess potential bias and inaccuracies in NGS pipelines, analyze raw sequencing data from ZymoBIOMICS® Microbial Community DNA Standard using the MIQ Score utility, which can be accessed via the web portal at www.miqscore.com[1].

1 Full documentation and docker version of the MIQ Score utility can be found at https://github.com/Zymo-Research/miqScore16SPublic and https://github.com/Zymo-Research/miqScoreShotgunPublic

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 standard can be found in the Specifications section.

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.	NCBI Reference Accession No.	Strain Name[1]
Bacillus subtilis	B-354	CP118021 CP118022	*Bacillus subtilis* (Ehrenberg 1835) Cohn 1872 ATCC 6633=NRRL B-209=NRS-231=PCI 219
Cryptococcus neoformans	Y-2534	JAQZRY000000000	*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	CP117970	*Enterococcus faecalis* (Andrewes and Horder 1906) Schleifer and Kilpper-Bälz 1984 ATCC 7080
Escherichia coli	B-1109	CP117971 CP117972	Castellani and Chalmers 1919, 01485cm
Lactobacillus fermentum	B-1840	CP132481	*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	CP117973	*Listeria monocytogenes* (Murray et al. 1926) Pirie 1940 2847=ATCC 19117
Pseudomonas aeruginosa	B-3509	CP117974 CP117975	*Pseudomonas aeruginosa* (Schroeter 1872) Migula 1900 ATCC 15442=NCIB 10421=Pdd-10
Saccharomyces cerevisiae	Y-567	JAQZRZ000000000	*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	CP117976 CP117977 CP117978	*Salmonella enterica* subspecies *enterica*, Castellani and Chalmers 1919, TA1536
Staphylococcus aureus	B-41012	CP117979 CP117981	*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; 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 4. Products Containing New Strains

Cat. #	Lot #	Product Name	Reference Genome and 16S/18S sequences
D6300	All current lots	ZymoBIOMICS® Microbial Community Standard	https://zymo-files.s3.amazonaws.com/BioPool/ZymoBIOMICS.STD.refseq.v3.zip
D6305	All current lots	ZymoBIOMICS® Microbial Community DNA Standard (200ng)	
D6306	All current lots	ZymoBIOMICS® Microbial Community DNA Standard (2000ng)	

Table 5. 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 DNA Standard	D6305 D6306	200ng/20µl 2000ng/20µl

Related Products	Catalog No.	Amount
ZymoBIOMICS® Microbial Community Standard	D6300	10 preps
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
ZymoBIOMICS® Oral Microbiome Standard	D6332	10 preps

Complete Your Workflow

✓ To collect and transport microbiome samples at ambient temperatures:

DNA/RNA Shield™ and Collection Devices		
R1100	DNA/RNA Shield™ Reagent	50 ml, 250 ml
R1200	DNA/RNA Shield™ Reagent (2x Concentrate)	25 ml, 125 ml
R1101	DNA/RNA Shield™ Fecal Collection Tube	10 pack
R1150	DNA/RNA Shield™ Blood Collection Tube	50 pack
R1160	DNA/RNA Shield™ SafeCollect Swab Collection Kit	1 ml, 2 ml
R1211	DNA/RNA Shield™ SafeCollect Saliva Collection Kit	2 ml

✓ Unbiased and inhibitor-free DNA and RNA extraction for microbiome profiling, available in a variety of formats to suit your needs:

ZymoBIOMICS® DNA and RNA Kits		
D4300	ZymoBIOMICS® DNA Miniprep Kit	50 preps
D4301	ZymoBIOMICS® DNA Microprep Kit	50 preps
D4302	ZymoBIOMICS® 96 MagBead DNA Kit	2 x 96 preps
R2001	ZymoBIOMICS® RNA Miniprep Kit	50 preps
R2137	ZymoBIOMICS® MagBead RNA Kit	96 preps
R2002	ZymoBIOMICS® DNA/RNA Miniprep Kit	50 preps
R2135	ZymoBIOMICS® MagBead DNA/RNA Kit	96 preps

✓ Ultra-streamlined workflows offering the fastest targeted 16S library prep available, with only 30 minutes of hands-on time and no tedious normalization required:

Quick-16S/ITS™ Plus NGS Library Prep Kits		
D6421	*Quick*-16S® Plus NGS Library Prep Kit (V3-V4)	768 Indexes
D6424	*Quick*-ITS® Plus NGS Library Prep Kit (ITS2)	384 Indexes
D6430	*Quick*-16S® Plus NGS Library Prep Kit (V4)	384 Indexes
D6434	*Quick*-16S® Plus NGS Library Prep Kit (V1-V2)	96 Indexes
D6440	*Quick*-16S® Plus NGS Library Prep Kit (V1-V3)	96 Indexes

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.

ZYMO RESEARCH

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

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