ZYMO RESEARCH

The Beauty of Science is to Make Things Simple

INSTRUCTION MANUAL

ZymoBIOMICS™ Microbial Community DNA Standard

Catalog Nos. D6305 (200ng) and D6306 (2000ng)

Highlights

- **Accurate composition:** composition cross-validated with multiple types of measurements.
- **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.

Contents

Product Contents 1
Specifications 1
Product Description 2-3
Strain Information 4
Protocol 5
Bioinformatics Analysis Recommendations 5
Appendix A: Additional Strain Information 6
Appendix B: Reference Sequences 7
Ordering Information 8

ZYMO RESEARCH CORP.

Phone: (949) 679-1190 ▪ Toll Free: (888) 882-9682 ▪ Fax: (949) 266-9452 ▪ info@zymoresearch.com ▪ www.zymoresearch.com

Notes: Satisfaction of all Zymo Research products is guaranteed. If you are dissatisfied with this product, please call 1-888-882-9682.

Product Contents

Product Name	D6305	D6306	Storage Temp.
ZymoBIOMICS™ Microbial Community DNA Standard	200ng/20µl	2000ng/20µl	-20°C

Specifications

Source: eight bacteria (3 gram-negative and 5 gram-positive) and 2 yeasts.

Reference genomes and 16S&18S rRNA genes[1]: https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip.

Storage solution: 10 mM Tris-HCl and 0.1 mM EDTA, pH 8.0.

DNA concentration: 10 ng/µl (D6305) and 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: http://www.zymoresearch.com/microbiomics/microbial-standards/zymobiomics-microbial-community-standards.

[1] 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 B to check if your product is from an older lot, and find the correct reference database to use accordingly.

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.

[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

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.

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, and phylogeny) can be found in Table 2. The 16S/18S rRNA sequences (fasta format) and genomes (fasta format) of these strains[3] are available at: https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip. Feel free to contact us if we can help analyze the sequencing data generated from this standard.

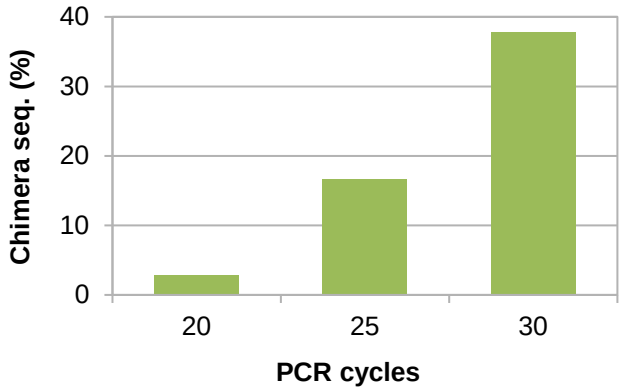

Figure 1. PCR chimera increases with increasing PCR cycle number in the library preparation process of 16S rRNA gene targeted sequencing. 20 ng ZymoBIOMICS™ Microbial Community DNA Standard was used as a template. The PCR reaction was performed with primers that target v3-4 region of 16S rRNA gene. Chimera 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.

Notes:

[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 preparation processes of shotgun sequencing.

[3] Several strains within the standard were replaced with similar strains beginning from Lot #190633. This update will not affect the species composition of the standard. Refer to Appendix B to check if your product is from an older lot, and find the correct reference database to use accordingly if needed.

Notes:

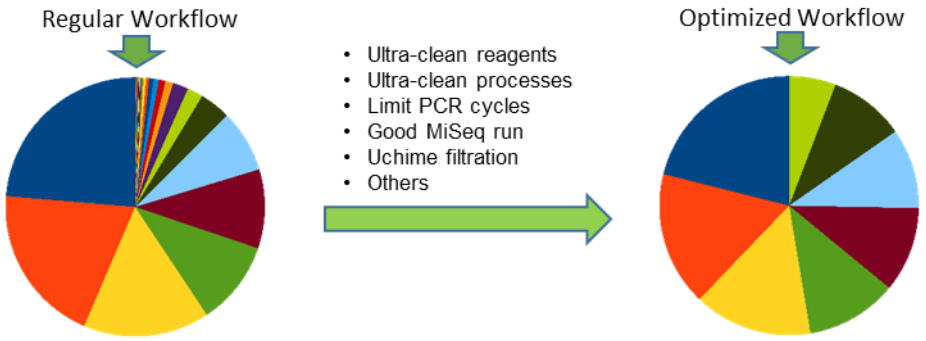

Figure 2. Eliminating noise/false positives in 16S rRNA gene targeted sequencing guided with ZymoBIOMICS™ Microbial Community DNA Standard. The pie chart on the left is the microbial composition profile of the standard determined by a regular workflow of 16S sequencing using primers targeting 16S v3-4 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.

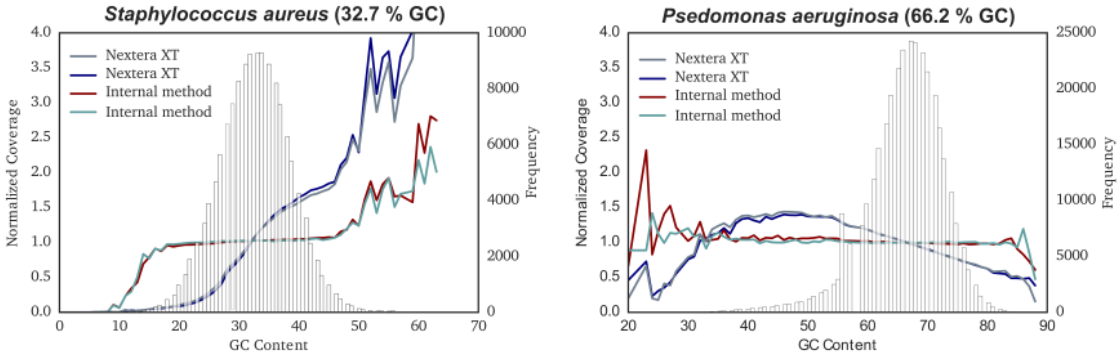

Figure 3. Assessing GC bias of two different library preparation methods in shotgun metagenomic sequencing. Library preparation for shotgun metagenomic sequencing was performed in two different ways: one by Illumina Nextera® XT kit and one by an in-house method. Shotgun sequencing was performed on MiSeq™ 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 GC content, 15%-85%. While the in-house method shows little/no GC-bias, the Nextera® XT kit has reduced representation for both low GC and high GC regions.

Note – Nextera® and MiSeq™ are trademarks of Illumina, Inc.

Table 2: Strain Information

Species	NRRL Accession NO.[1]	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[2]	Yeast
Cryptococcus neoformans	Y-2534	18.9	2	48.3	60[2]	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

Notes:

[1] 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 B to check if your product is from an older lot, and find the correct reference database to use accordingly.

[2] 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.

Notes:

Protocol

1. Thaw the standard on ice. After thawing, vortex and spin down quickly.

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 preparation processes[1]

Lib. Prep	16S Library	Illumina Nextera® XT	Illumina TruSeq® Nano	TruSeq® PCR-free	Kapa HyperPlus
DNA input (ng)	10	1	>200	2000	1-2000

[1] The table was prepared in April of 2016

Bioinformatics Analysis Recommendations

1. Assessing accuracy of taxonomy identification

A fundamental goal in microbiome studies is to identify what microbes are present in a sample. After you analyze this microbiome standard using a workflow that include wet-lab processing and dry-lab interpretation, you can compare the taxa identified with the taxonomy information of the ten strains included in the standard (Table 2). This allows you to assess the performance of your workflow regarding the limit of the taxonomy resolution, false positives, and false negatives. The issue of false positives is an important factor. False positives can be caused by contaminations in the dry-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, as long as you observe any alien taxa present at >0.01% (by DNA abundance) in the analysis results of the standard, you can conclude that they are introduced artificially by your workflow.

2. Assessing bias in composition profiling

Another important goal in microbiome studies is to accurately determine the microbial composition of a sample. You can compare the composition profile determined by your workflow to the data shown in Table 1. Both wet-lab and dry-lab processes can introduce bias into the final results. When you want to focus on the question about whether or not the wet-lab process causes bias, you will need an accurate/unbiased dry-lab analysis method to interpret the sequencing data from the standard. We found that direct read-mapping against reference genomes (or against reference16S&18S sequences in the case of targeted sequencing) of the ten strains is a straightforward and accurate way to infer the microbial composition from sequencing data. The reference sequences of this microbiome standards can be found in the section of “Specifications” (Page 1) or from Appendix B.

Appendix A

Table 3. Additional Strain Information

Species	NRRL Accession NO.	Strain Name[1]
Bacillus subtilus	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
Psuedomonas 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 B: Reference Sequences

Different lots of the ZymoBIOMICS Microbial Community DNA Standard contain DNA from different strains of the microbial species. Table 3 provides the reference sequences corresponding to the appropriate lot numbers.

Key Points:

- Species-level composition of the standards is unchanged lot to lot.
- 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: ZymoBIOMICS Microbial Community DNA Standard (D6305) Lot #s and Corresponding Reference Sequences

Lot #	Reference Genome and 16S/18S Sequences
ZRC183430	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.genomes.ZR160406.zip
195708	https://zymo-files.s3.amazonaws.com/BioPool/ZymoBIOMICS.STD.refseq.SAL.zip
ZRC190812	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip
Future lots	

Table 4: ZymoBIOMICS Microbial Community DNA Standard (D6306) Lot #s and Corresponding Reference Sequences

Lot #	Reference Genome and 16S/18S Sequences
ZRC183430	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.genomes.ZR160406.zip
195709	https://zymo-files.s3.amazonaws.com/BioPool/ZymoBIOMICS.STD.refseq.SAL.zip
210348	
210618	
210792	
ZRC190811	https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip
Future lots	

Ordering Information

Product Description	Catalog No.	Size
ZymoBIOMICS™ Microbial Community DNA Standard (200ng)	D6305	200 ng
ZymoBIOMICS™ Microbial Community DNA Standard (2000ng)	D6306	2000 ng

Related Products

Product Description	Catalog No.	Size
ZymoBIOMICS™ Microbial Community Standard	D6300	10 preps
ZymoBIOMICS™ DNA Miniprep Kit	D4300	50 preps
ZymoBIOMICS™ Microbial Community Standard II (Log Distribution)	D6310	10 preps
ZymoBIOMICS™ Microbial Community DNA Standard II (Log Distribution)	D6311	220ng

Sample Collection	Catalog No.	Size
DNA/RNA Shield™ – Swab Collection Tube	R1106	10 preps
DNA/RNA Shield™ – Fecal Collection Tube	R1101	10 preps
DNA/RNA Shield™ – Lysis Tube	R1103	50 preps
DNA/RNA Shield™	R1100-50 R1100-250	50 ml 250 ml
DNA/RNA Shield™ (2X concentrate)	R1200-25 R1200-125	25 ml 125 ml

ZymoBIOMICS™ is a trademark of Zymo Research Corporation. ZymoBIOMICS™ Microbial Community DNA Standard is for research use only. It is not sold for use in diagnostic procedures. Reagents included with this kit are irritants. Follow the safety guidelines and rules enacted by your research institution or facility including the wearing of protective gloves and eye protection when using this kit.

Nextera®, MiSeq™ and TruSeq® are trademarks of Illumina, Inc. Kapa HyperPlus is a trademark of Kapa Biosystems, Inc.