Evaluation of 16S rRNA and Shotgun Metagenomic Analytical Methods for Community Profiling using ATCC® Microbiome Standards

Juan Lopera, Ph.D., Ray-Yuan Chuang, Ph.D., Stephen King, M.S., and Dev Mittar, Ph.D.
ATCC, Manassas, VA

ATCC® | THE ESSENTIALS OF LIFE SCIENCE RESEARCH GLOBALLY DELIVERED™

Background and summary

Complexities involved in 16S rRNA and shotgun metagenomic analysis methods pose significant challenges in microbiome research as various biases can be introduced during PCR amplification, library preparation, sequencing, and analysis. One of the primary obstacles in assay standardization is the limited availability of reference materials. To support this need, we developed microbiome reference standards from fully sequenced and characterized ATCC® strains and evaluated their use in a proof-of-concept study. Here, four standards were created comprising mixtures of 10 or 20 genomic DNAs in equal or staggered quantities prepared from a diverse set of bacteria that were selected based on relevant attributes such as Gram stain, genome size, GC content, and other special characteristics. Initially, we performed an inter-laboratory comparison of the 16S rRNA V4 region from three different commercial laboratories. Analysis of the resulting sequencing data using One Codex revealed variability in the number of true positives and false positives as well as the relative abundances. A subsequent comparison of three different regions of the 16S rRNA gene—V1V2 (27f-YM+3 338R), V3V4 (341F & 806R), and V4 (515F & 806R)—revealed that only the analysis of the V1/V2 region using One Codex was able to profile the bacteria to the species level. Following this analysis, we evaluated a shotgun metagenomics approach and compared it to the 16S rRNA V1/V2 results. Here, the shotgun metagenomics approach showed a high correlation between the expected vs. observed ratios as compared to the 16S rRNA results, which had variable correlation for three standards with equal and staggered ratios. These results demonstrate that 16S rRNA community profiling of the V1V2 region and the shotgun metagenomic approach could identify bacterial strains to the species level, with the latter method generating relative abundances statistically close to the expected. Taken together, this proof-of-concept study demonstrates the potential use of ATCC® Microbiome Standards in the identification of potential biases and methodology drawbacks associated with microbiome studies.

ATCC® Microbiome Standards

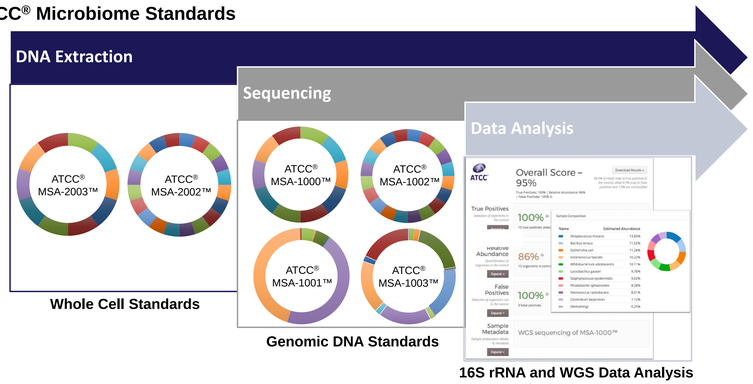


Table 1: Individual bacterial strains within the ATCC® Microbiome Standards

ATCC® No.	Name	Gram Stain	% GC	Genome Size (Mb)	Special Features	Microbiome	16S rRNA Copies	GenBank ID
10987™	*Bacillus cereus*	+	35.2	5.42	Endospores former	Soil	12	NC_003909.8
15703™	*Bifidobacterium adolescentis*	+	59.2	2.09	Anaerobe	Gut	5	NC_008618.1
35702™	*Clostridium beijerinckii*	+	30	6.49	Spores former	Gut/soil	14	NC_009617.1
BAA-816™	*Deinococcus radiodurans*	-	66.7	3.29	Thick cell wall	Gut/environment	7	NC_001263.1
47077™	*Enterococcus faecalis*	+	37.5	3.36	Biofilm producer	Gut	4	NC_017316.1
700926™	*Escherichia coli*	-	50.8	4.64	Facultative anaerobe	Gut	7	NC_000913.3
33323™	*Lactobacillus gasseri*	+	35.3	1.89	Nuclease producer	Vaginal/gut	6	NC_008530.1
17029™	*Rhodobacter sphaeroides*	-	68.8	4.60	Metabolically diverse	Aquatic	3	NZ_AKVW01000001.1
12228™	*Staphylococcus epidermidis*	+	31.9	2.56	Thick cell wall	Skin/mucosa	5	NC_004461.1
700610™	*Streptococcus mutans*	+	36.8	2.03	Facultative anaerobe	Oral	5	NC_004350.2
17978™	*Acinetobacter baumannii*	-	39	4.34	Filaments, capsule	Environment	6	NZ_CP009257.1
17982™	*Actinomyces odontolyticus*	+	65.5	2.39	Type 1 fimbriae	Oral	2	NZ_DS264586.1
8482™	*Bacteroides vulgatus*	-	42.2	5.16	Anaerobe	Gut	7	NC_009614.1
700392™	*Helicobacter pylori*	-	38.9	1.67	Helix shaped	Stomach/gut	2	NC_000915.1
BAA-335™	*Neisseria meningitidis*	-	51.5	2.27	Diplococcus	Respiratory tract	4	NC_003112.2
33277™	*Porphyromonas gingivalis*	-	48.4	2.35	Anaerobe, collagenase	Oral	4	NC_010729.1
11828™	*Propionibacterium acnes*	+	60	2.56	Aerotolerant anaerobe	Skin	4	NC_006085.1
9027™	*Pseudomonas aeruginosa*	-	66.6	6.26	Facultative anaerobe	Skin	4	NC_009656.1
BAA-1556™	*Staphylococcus aureus*	+	32.8	2.82	Thick cell wall	Skin/respiratory	6	NC_007795.1
BAA-611™	*Streptococcus agalactiae*	+	35.6	2.16	Serogroup B	Vaginal/environment	7	NC_004116.1

Development of genomic DNA microbiome standards

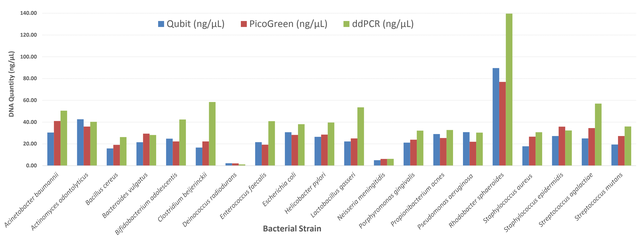


Figure 1. Quantification of DNA for preparation of genomic DNA standards. DNA from each bacteria was quantified via the Qubit® dsDNA BR assay (Thermo Fisher Scientific Inc.), PicoGreen® (Thermo Fisher Scientific Inc.), and Droplet Digital™ PCR (ddPCR™; Bio-Rad) (using primers and probes against a single copy gene, *rpoB*) as per the manufacturers' instructions. Each sample was processed from an equivalent volume of starting material.

Inter-laboratory study using V4 region of 16S rRNA

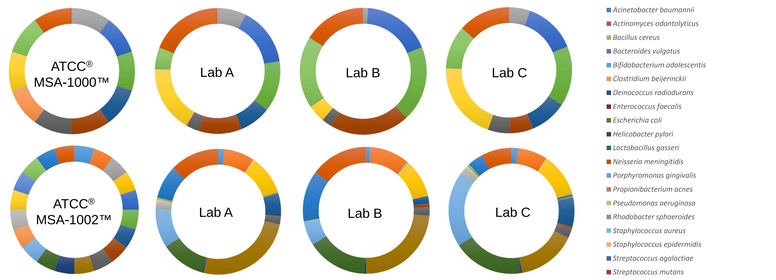


Figure 2. The use of standards during PCR amplification, library preparation, and sequencing. Inter-laboratory variations in identity and relative abundances. 16S rRNA V4 sequences data from three different laboratories. Percent ratios of expected and observed organisms in the even genomic mock community comprising 10 and 20 organisms. The blinded samples were sent to commercial vendors where they used their standard 16S protocol (Earth Microbiome Project) on the Illumina® platform. The FASTQ files were analyzed using One Codex.

Comparison of different regions of 16S rRNA gene

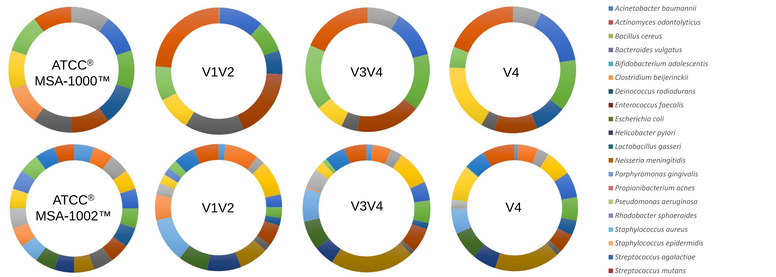


Figure 3. Choice of 16S rRNA primer regions affects identity and relative abundances. 16S rRNA community profiling results from ATCC® MSA-1001™ and ATCC® MSA-1002™ using primer sets covering the V1V2, V3V4, and V4 regions on the Illumina® platform. The FASTQ files were analyzed using One Codex.

Comparison of 16S V1/V2 rRNA and shotgun metagenomic sequencing

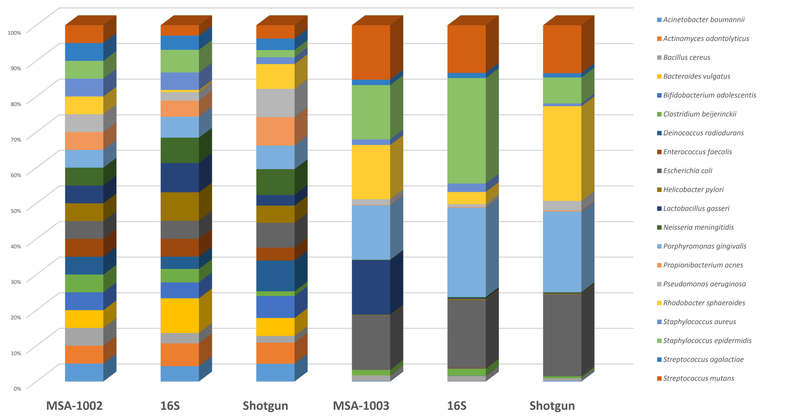


Figure 4. Data analysis platform impacts identification and relative abundances. The ATCC® MSA-1002™ and ATCC® MSA-1003™ genomic DNA standards were tested using the 16S rRNA (V1/V2) and shotgun metagenomic sequencing. Relative abundance data was calculated after performing analysis in One Codex.

Running the control analysis in One Codex

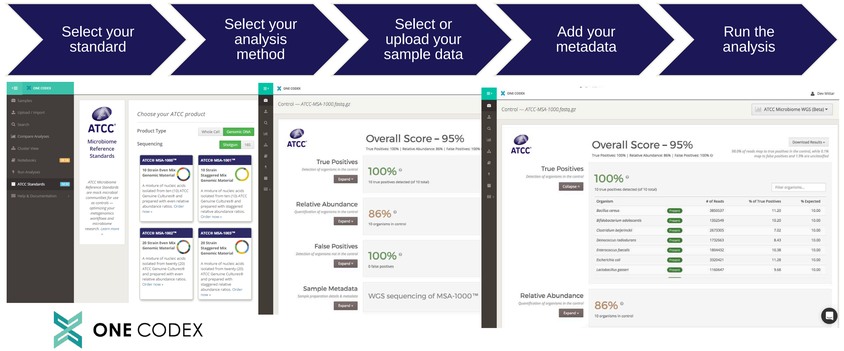
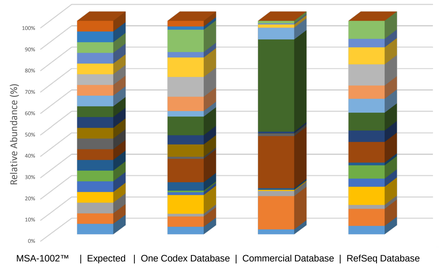


Figure 5. Analyzing data using the One Codex cloud-based web interface. The One Codex platform can be used to evaluate the number of true-positive, relative abundance, and false-positive scores for 16S rRNA and WGS sequencing methods.

Comparison of One Codex with other platforms



	Expected	One Codex Database	Commercial Database	RefSeq Database
Streptococcus agalactiae	0.05	2.63%	0.00%	0.00%
Staphylococcus aureus	0.05	1.51%	0.01%	0.00%
Pseudomonas aeruginosa	0.05	10.44%	1.16%	7.29%
Propionibacterium acnes	0.05	2.54%	0.53%	3.43%
Porphyromonas gingivalis	0.05	5.10%	1.42%	6.88%
Neisseria meningitidis	0.05	9.23%	0.13%	8.56%
Helicobacter pylori	0.05	6.68%	0.08%	5.37%
Bacteroides vulgatus	0.05	2.58%	5.38%	5.61%
Actinomyces odontolyticus	0.05	8.64%	43.01%	7.15%
Acinetobacter baumannii	0.05	4.30%	0.87%	4.63%
Streptococcus mutans	0.05	5.74%	0.08%	0.00%
Staphylococcus epidermidis	0.05	0.97%	1.12%	0.00%
Rhodobacter sphaeroides	0.05	10.93%	24.34%	8.36%
Lactobacillus gasseri	0.05	3.95%	0.79%	1.10%
Escherichia coli	0.05	0.72%	0.00%	5.37%
Enterococcus faecalis	0.05	1.30%	0.01%	3.29%
Deinococcus radiodurans	0.05	6.76%	0.52%	7.36%
Clostridium beijerinckii	0.05	1.29%	2.37%	1.58%
Bifidobacterium adolescentis	0.05	4.89%	15.73%	7.00%
Bacillus cereus	0.05	3.60%	2.34%	3.48%

Figure 6. Performance and overall accuracy of the One Codex platform as compared to other platforms. The theoretical performance of One Codex was compared to other bioinformatics tools and databases by using ATCC® MSA-1002™ standard. The % relative abundance data is presented for the analysis performed using one codex, a commercial and RefSeq databases

One Codex Analysis Overview

1. Detect reference organisms by local alignment with high sensitivity
2. Convert the read counts to relative abundance using genome size or 16S rRNA copy number
3. Detect non-reference organisms (false positives) using the One Codex Database (WGS) or Targeted Loci Database (16S)
4. Display the "scorecard" with high-level accuracy scores and per-organism details

Table 2. Scorecard Results for ATCC® Genomic DNA Microbiome Standards

ATCC® No.	Analysis type	True positives (Detected/input)	Relative ratio (Correlation coefficient)	False positives (100 – false + penalty)	Overall score (Average of three sub-scores)
MSA-1000™ (Lot # 70003364)	Shotgun	100%	52%	100%	84%
	16S (V1/V2)	100%	66%	100%	89%
MSA-1001™ (Lot # 70003365)	Shotgun	100%	91%	99%	97%
	16S (V1/V2)	100%	89%	92%	94%
MSA-1002™ (Lot # 70003364)	Shotgun	100%	59%	100%	86%
	16S (V1/V2)	100%	62%	98%	87%
MSA-1003™ (Lot # 70003365)	Shotgun	100%	95%	95%	97%
	16S (V1/V2)	100%	92%	93%	95%

Summary and Conclusions

The data presented in this poster clearly reveals the need for standardization in microbiome analyses. Here, we demonstrate that bacterial identification and the evaluation of relative abundances in mixed samples can be affected by the following:

- Inter-laboratory variations, even when using the same protocol - 16S rRNA community profiling varied between labs
- Choice of 16S rRNA region for PCR can significantly impact the identification and relative abundances
- Data analysis platforms vary in their ability to calculate the relative abundances and to identify all bacteria in the mix, particularly at the species level

Overall, ATCC® Microbiome Standards combined with the One Codex data analysis module provide a comprehensive solution for assay development and process control monitoring for 16S rRNA community profiling and shotgun metagenomics methods used in microbiome studies.