Authenticated Microbial Reference Genomes for Microbiome Analysis

Nikhita Puthuveetil, MS;[1] Juan Lopera, PhD;[1] Ford Combs, MS;[1] David Yarmosh, MS;[1] Samuel Greenfield, BS;[1] Stephen King, MS;[1] Marco Riojas, PhD;[1] Amanda Pierola, BS;[1] John Bagnoli, BS;[1] Denise B. Lynch, PhD;[2] Austin G. Davis-Richardson, PhD;[2] Briana Benton, BS;[1] Jonathan Jacobs, PhD[1]
[1]ATCC, Manassas VA, 20110
[2]Invitae, San Francisco CA, 94103

ATCC®
Credible leads to Incredible™

Background

Shotgun metagenomic and targeted amplicon sequencing are two widely used methods to study the human and environmental microbiome. ATCC has collaborated with One Codex to provide analysis modules for ATCC Microbiome Standards that enable researchers to analyze shotgun and 16S rRNA sequencing data. These modules come with pre-loaded ATCC Microbiome Standards metadata; sequences for shotgun or 16S comparative analyses; and a set of scores assessing the relative abundance, true positives, and false positives of the microbiome sample.

Previously, these microbiome modules used NCBI sequences to classify reads, but as the ATCC Genome Portal has expanded its collection of high-quality and authenticated genome sequences, the latest versions of these modules have switched to using ATCC-derived sequences due to their high fidelity and traceability. Here, we demonstrate the differences between the One Codex microbiome module versions and highlight the discrepancies between ATCC genomic data sets and publicly available genomes cited as "ATCC" or with identical strain designations.

ATCC Microbiome Standards

Table 1: A sample of the ATCC mock microbial communities used in this analysis.

ATCC® Number	Preparation	Number of Organisms	Importance
MSA-1000™	Genomic DNA	10	Standards for assay development and optimization
MSA-1001™		10	
MSA-1002™		20	
MSA-1003™		20	
MSA-2002™	Whole Cell	10	
MSA-2003™		20	
MSA-3000™	Genomic DNA	6	Environmental studies
MSA-3001™		10	
MSA-3002™		10	

ATCC Genome Portal

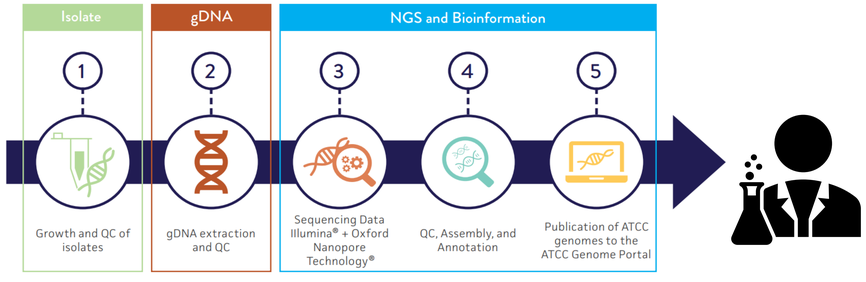

Figure 1: The ATCC genome portal workflow used to produce high-quality, authenticated genomes.

Overall Improvements in Microbiome Measures

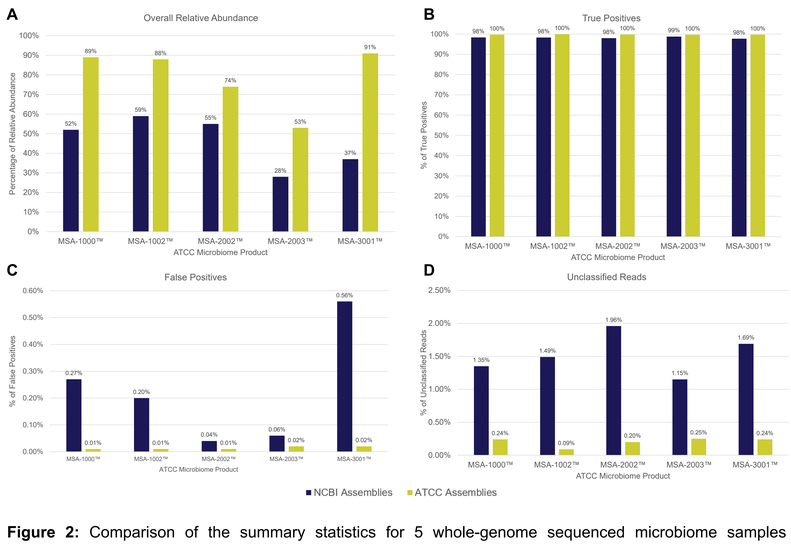

Figure 2: Comparison of the summary statistics for 5 whole-genome sequenced microbiome samples generated using both versions of the ATCC microbiome module (one using NCBI reference assemblies and the other using ATCC Genome Portal assemblies). The data depict (A) Overall relative abundance, (B) number of true positives, (C) number of false positives, and (D) the number of unclassified reads.

Improvements in Reads Classification

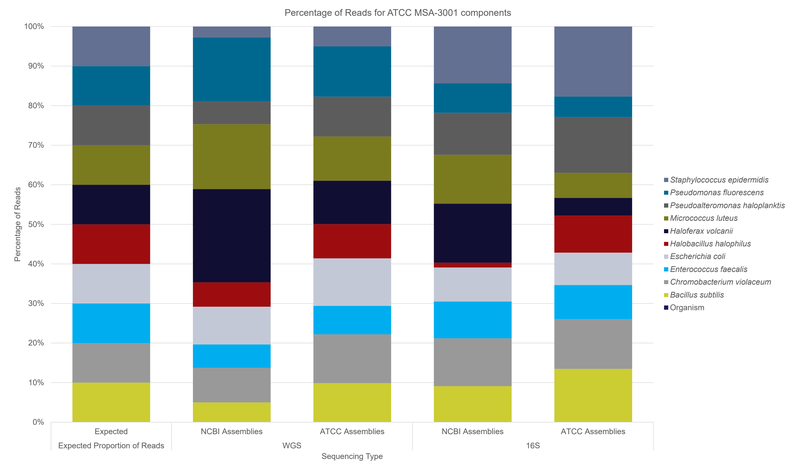

Figure 3: Comparison of read taxonomic classification done by both versions of the ATCC microbiome module (one using NCBI reference assemblies and the other using ATCC Genome Portal assemblies).

Evaluation of ATCC sequences to GenBank assemblies

Table 2: A selection of differences between ATCC Genome Portal sequences and GenBank assemblies.

ATCC® Number	Species	GenBank Accession	Number of Contigs ATCC / GenBank	16S copies ATCC / GenBank	Completeness ATCC / GenBank	# SNPs / Indels
BAA-816™	*Deinococcus radiodurans*	GCA_000687895.1	5 / 51	3 / 0	99.15% / 99.57%	83 / 6
BAA-611™	*Streptococcus agalactiae*	GCA_000007265.1	1 / 1	7 / 7	100.00% / 100.00%	5 / 2
700802™	*Enterococcus faecalis*	GCA_000007785.1	4 / 4	4 / 4	99.53% / 99.53%	16 / 24
25285™	*Bacteriodes fragilis*	GCA_001997325.1	25 / 59	5 / 1	99.26% / 99.26%	15 / 6
14393™	*Pseudoalteromonas haloplanktis*	GCA_000238355.2	8 / 56	8 / 0	100.00% / 100.00%	122,005 / 20,057
35676™	*Halobacillus halophilus*	GCA_000284515.1	3 / 3	7 / 7	99.33% / 99.33%	51 / 13

Taxonomic Profiling of *De Novo* Contigs

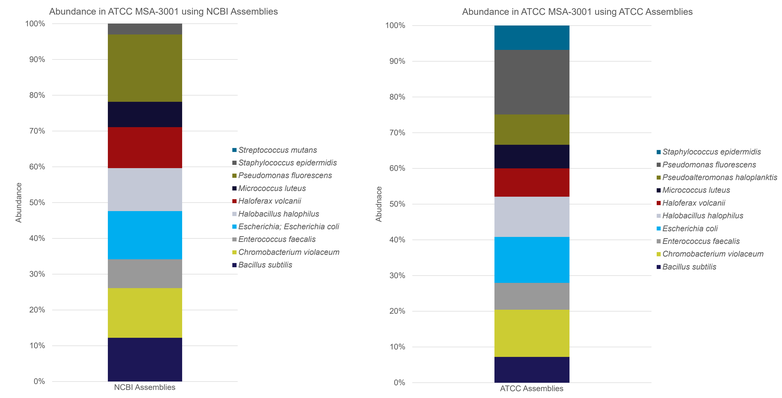

Figure 4: Comparison of *de novo* contig taxonomic classification using NCBI Assemblies and ATCC Assemblies. The classification using NCBI Assemblies detected a component not present in ATCC MSA-3001 (*Streptococcus mutans*).

Conclusion

Availability of high-fidelity reference genomes can lead to improvements in standardized analyses and better reproducibility in microbiome research. Based on our comparisons of the One Codex microbiome analysis modules, we discovered many improvements in both relative abundance and true and false positive scores using the newer version of the module. These differences can be attributed to discrepancies between ATCC genome portal sequences and public assemblies.

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 800.638.6597 **Email:** sales@atcc.org **Web:** www.atcc.org