ATCC Whole Cell- and Genomic DNA-Based Microbiome Reference Standards

Juan Lopera[1], Ph.D., Monique Hunter[1], M.S., Ray-Yuan Chuang[1], Ph.D., Megan Amselle[1], M.T., Brian Chase[1], M.S., Stephen King[1], M.S., Maria Mayda[1], Ph.D., Kevin Zinn[1], B.S., Nicholas B. Greenfield[2], M.A., and Dev Mittar[1], Ph.D.

[1]ATCC, Manassas, VA
[2]One Codex, San Francisco, CA

Background and Significance

Advancement and accessibility of next-generation sequencing technologies have influenced microbiome analyses in tremendous ways, opening up applications in the areas of clinical, diagnostic, therapeutic, industrial, and environmental research. However, due to the complexity of 16S rRNA and metagenomic sequencing analysis, significant challenges can be posed by biases introduced during sample preparation, DNA extraction, PCR amplification, library preparation, sequencing, or data interpretation. One of the primary challenges in assay standardization is the limited availability of reference materials. To address these biases and provide a measure of standardization within microbiome research and applications, ATCC has developed a set of mock microbial communities comprising fully sequenced, characterized strains (Table 1) selected on the basis of phenotypic and genotypic attributes, such as cell wall type (Gram stain classification), GC content, genome size, unique cell wall characteristics, and spore formation (Table 2). These mock communities mimic mixed metagenomic samples and offer a universal control for microbiome analyses and assay development. Moreover, these standards have been developed with different levels of mock community complexity (10 or 20 strains per community) with even or staggered relative abundance, including diverse strains that are relevant to a broad range of applications. In addition, to minimize the bias associated with data interpretation, we have developed a data analysis module in collaboration with One Codex. This module provides a user-friendly output in the form of true-positive, relative abundance, and false-negative scores for 16S rRNA community profiling and shotgun metagenomic sequencing.

ATCC® Microbiome Standards

Table 1. ATCC® Microbiome Standards

ATCC® No.	Preparation	Number of Organisms	Specification Range	Applications
MSA-1000™	Genomic DNA	10	Even amounts, 2x10⁶ genome copies/organism	Assay development for 16S rRNA and shotgun sequencing-based profiling of microorganisms
MSA-1001™		10	Staggered amounts, 2x10⁴ - 2x10⁷ genome copies/organism	
MSA-1002™		20	Even amounts, 2x10⁶ genome copies/organism	
MSA-1003™		20	Staggered amounts, 2x10⁴ - 2x10⁷ genome copies/organism	
MSA-2003™	Whole cells	10	Even amounts, 2x10⁶ cells/organism	Parallel run controls for monitoring the full process from extraction to analysis
MSA-2002™		20	Even amounts, 2x10⁶ cells/organism	
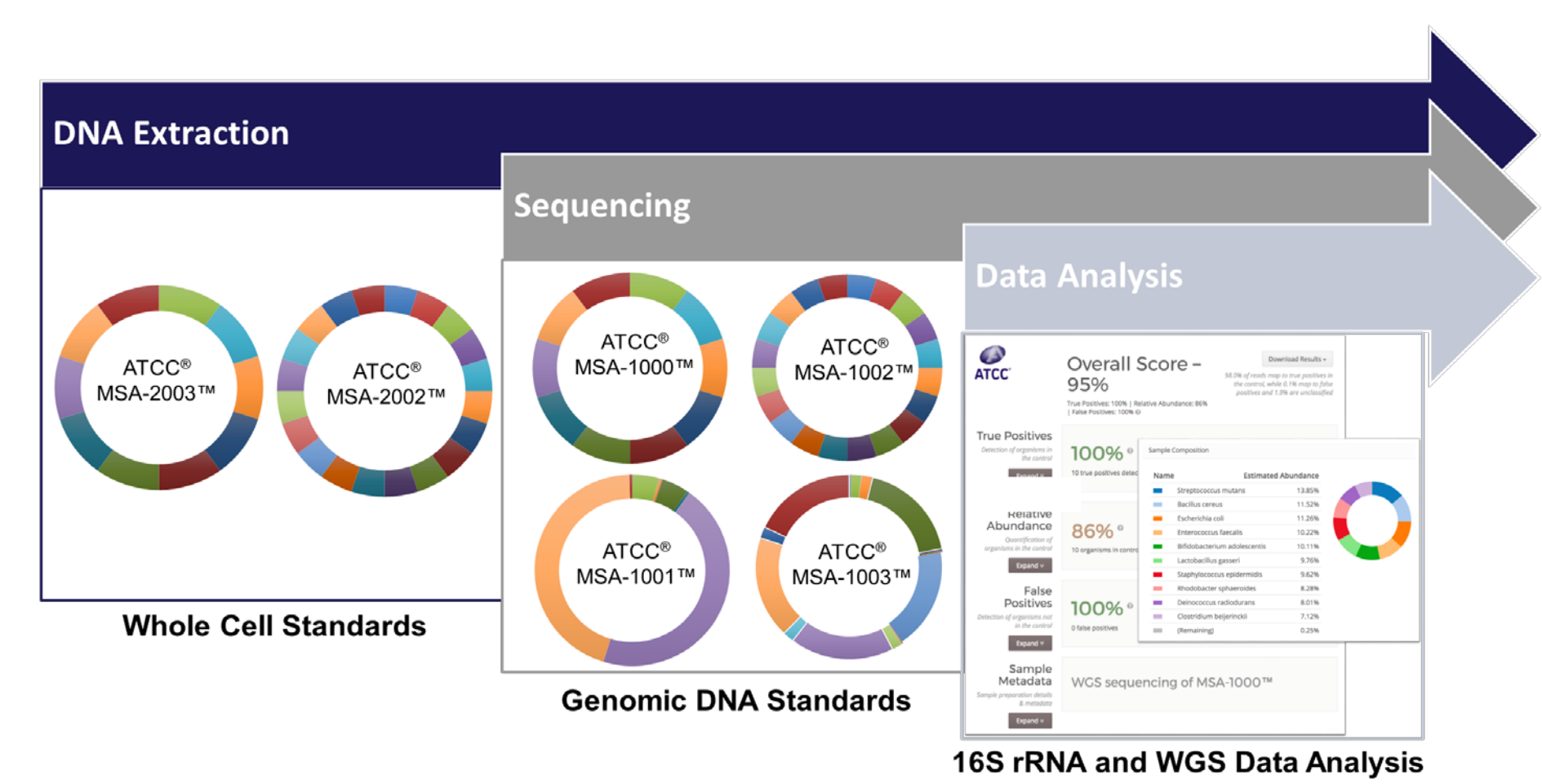
Development of the ATCC® Microbiome Standards

Table 2. Individual bacterial strains within the ATCC® Microbiome Standards

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

Figure 1. Combining the ATCC® Microbiome Standards with the One Codex data analysis platform. A) Work flow for analyzing data using the One Codex cloud-based web interface. B) The One Codex platform can be used to evaluate the number of true-positive, relative abundance, and false-positive scores for 16S rRNA and whole genome sequencing methods.

ONE CODEX

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 (whole genome sequencing) or Targeted Loci Database (16S rRNA)
4. Display the "**scorecard**" with high-level accuracy scores and per-organism details – True-positive, false-positive, and relative abundance scores

Evaluation of ATCC® Microbiome Standards

D.

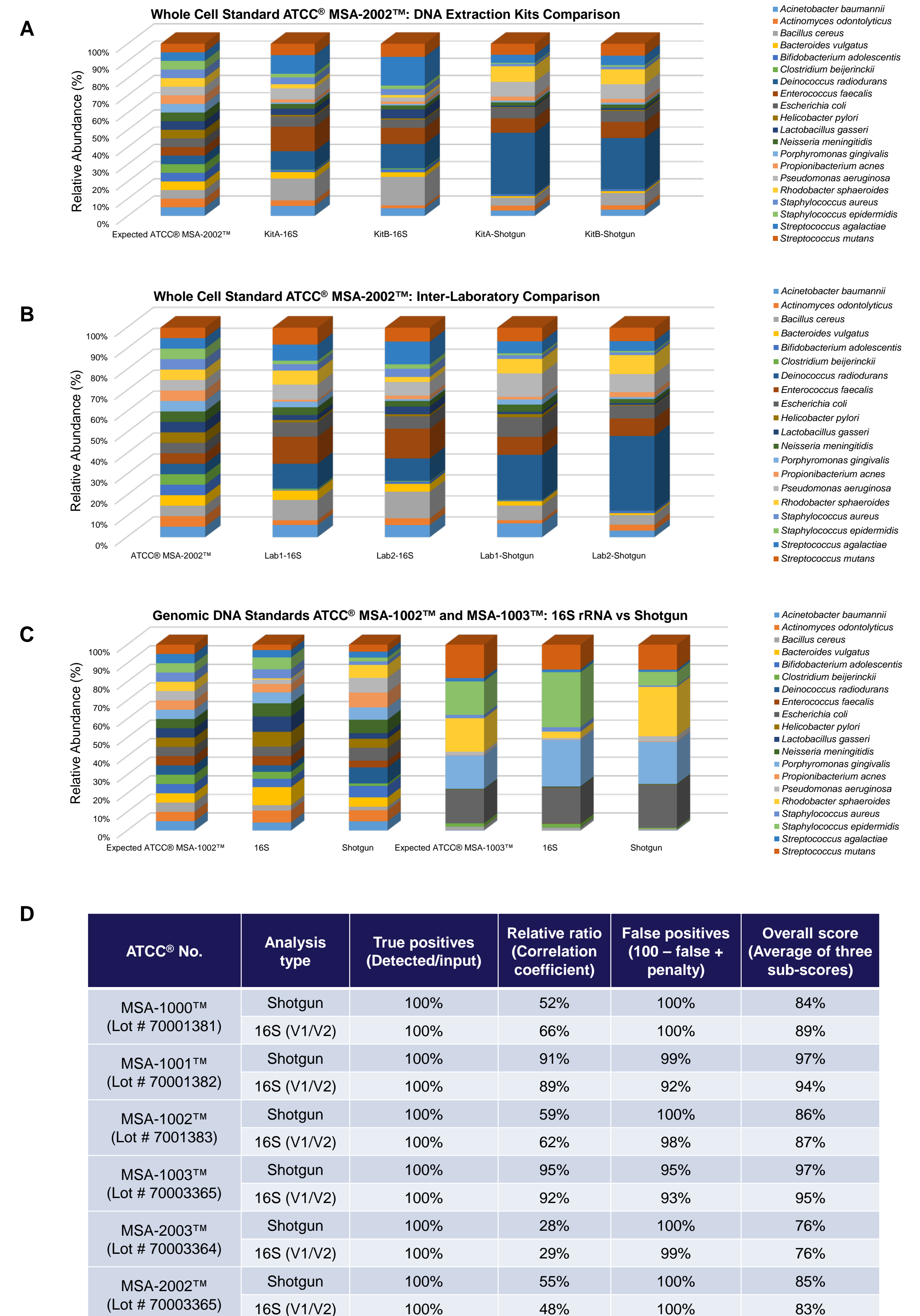| ATCC® No.	Analysis type	True positives (Detected/input)	Relative ratio (Correlation coefficient)	False positives (100 – false + penalty)	Overall score (Average of three sub-scores)
MSA-1000™ (Lot # 70001381)	Shotgun	100%	52%	100%	84%
	16S (V1/V2)	100%	66%	100%	89%
MSA-1001™ (Lot # 70001382)	Shotgun	100%	91%	99%	97%
	16S (V1/V2)	100%	89%	92%	94%
MSA-1002™ (Lot # 7001383)	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%
MSA-2003™ (Lot # 70003364)	Shotgun	100%	28%	100%	76%
	16S (V1/V2)	100%	29%	99%	76%
MSA-2002™ (Lot # 70003365)	Shotgun	100%	55%	100%	85%
	16S (V1/V2)	100%	48%	100%	83%

Figure 2. Evaluation of the ATCC® Microbiome Standards. A) ATCC® MSA-2002™ was extracted using two different DNA extraction kits followed by profiling of the 16S rRNA V1/V2 region and shotgun genomic sequencing on the Illumina® platform. Data analysis was performed in the One Codex module. B) ATCC® MSA-2002™ was profiled by both 16S rRNA and shotgun genomic sequencing methods by two different laboratories (MoBio kit, V1/V2 primers and Illumina® platform). C) ATCC® MSA-1002™ and ATCC® MSA-1003™ were profiled using the 16S rRNA V1/V2 region protocol and shotgun sequencing on the Illumina® platform. D) Scorecard analysis output from the One Codex data analysis module.
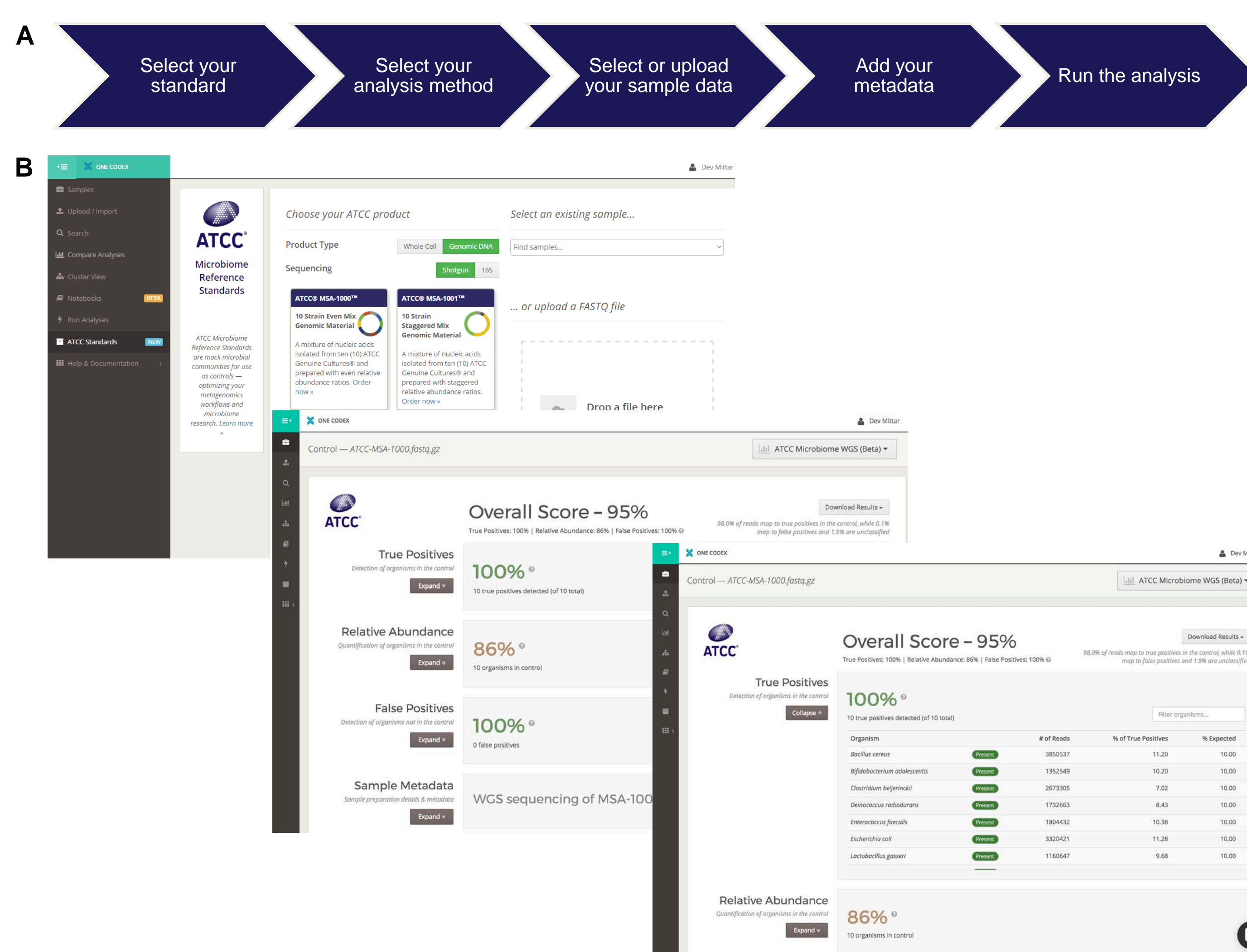
Summary and Conclusions

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.