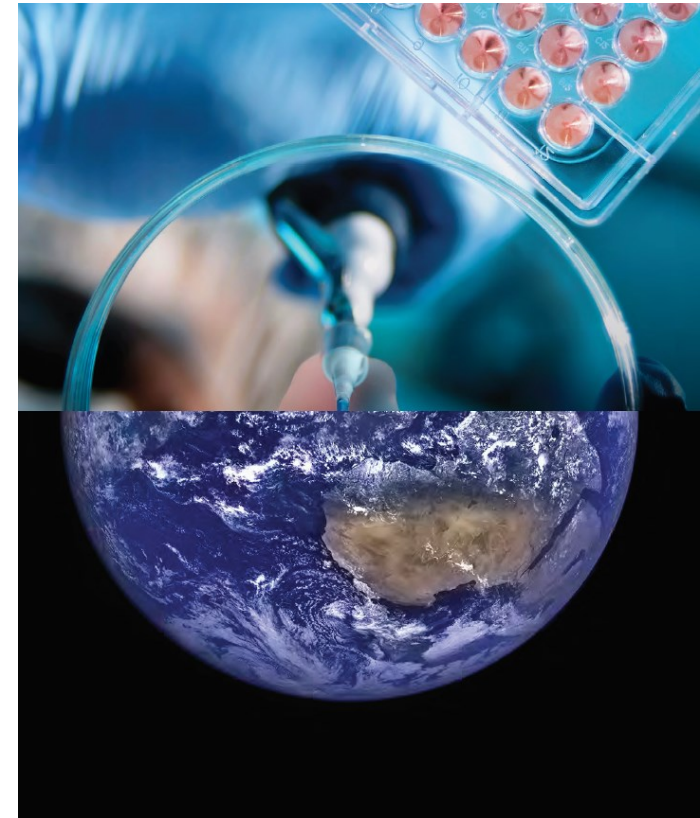
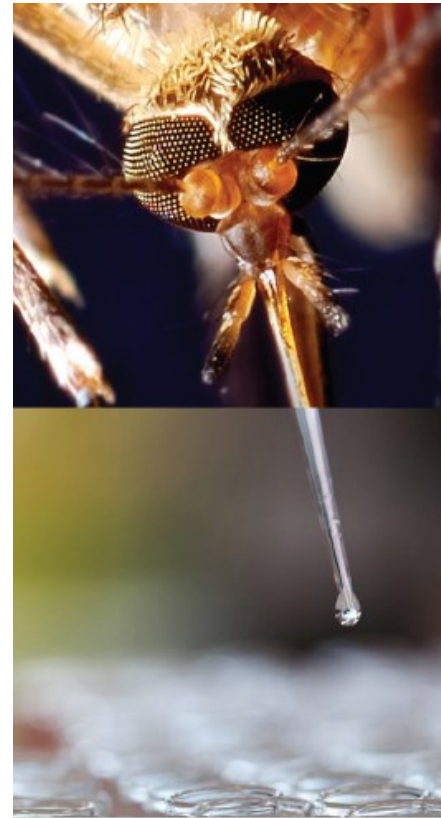
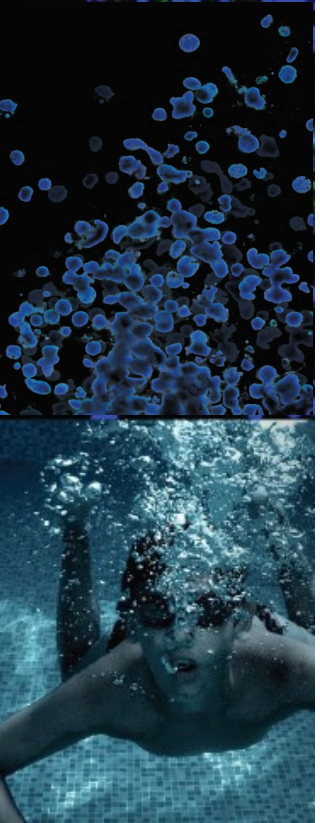




Making Sense Out of Microbiome Data – The Importance of Standards

Briana Benton, BS
Technical Manager, ATCC

Credible Leads to Incredible™



Agenda

- ✓ Background on American Type Culture Collection
- ✓ Challenges in microbiome research
 - ✓ Discuss why standards are essential
- ✓ Development of microbiome standards
- ✓ The ATCC[®] Microbiome Standards portfolio and upcoming new products
- ✓ Applications of standards in microbiome research
 - ✓ Extraction method, assay variability, NGS library preparation, and bioinformatics analysis
- ✗ Microbiome assay development
- ✗ Show the best data
- ✗ Recommend any specific assay, kit, protocol, or instrument



About ATCC

- Founded in 1925, ATCC is a non-profit organization with HQ in Manassas, VA, and an R&D and Services center in Gaithersburg, MD
 - World's largest, most diverse biological materials and information resource for microbes – the “*gold standard*”
 - Innovative R&D company featuring gene editing, microbiome, NGS, advanced models
 - cGMP biorepository
- Partner with government, industry, and academia
 - Leading global supplier of authenticated cell lines, viral and microbial standards
 - Sales and distribution in 150 countries, 18 international distributors
 - Talented team of 450+ employees, over one-third with advanced degrees



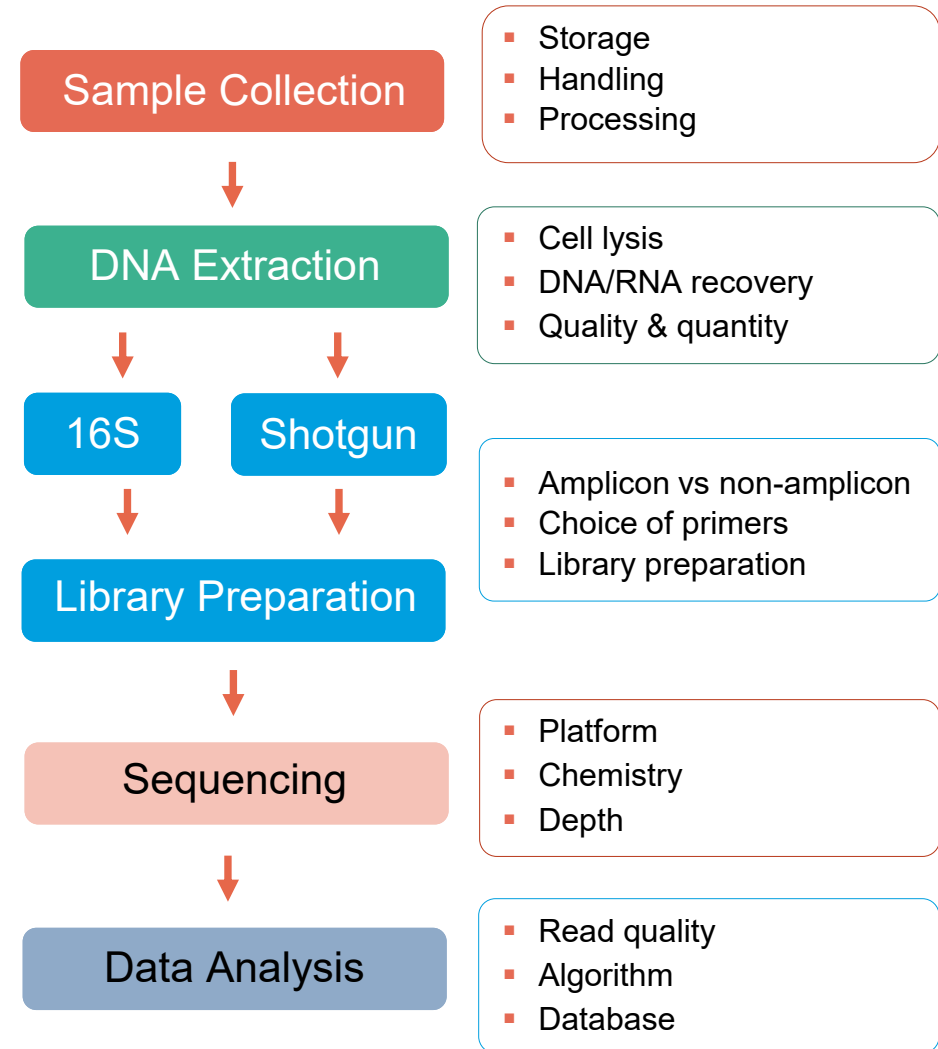
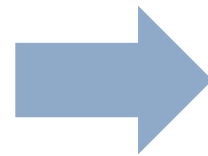
Microbiome Research

The microbiome field is rapidly moving toward translational research pertinent to human health and disease, therapeutics, and personalized medicine

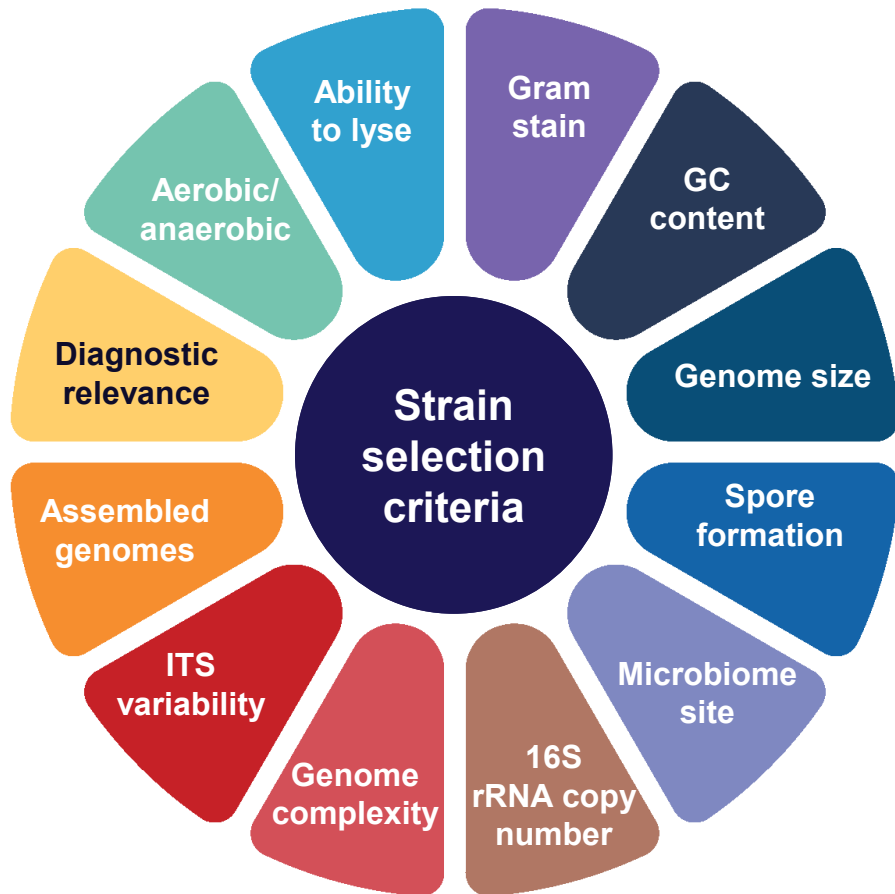
Challenges in Microbiome Research

sequencing viability coverage
extraction technology
amplification
depth bioinformatics

bias



Development of Mock Microbial Communities



Whole Cell Standards

- Authenticated ATCC cultures
- Growth and image cytometry cell counting
- Mixed in even proportions.
- Store at 4°C until ready to use



Genomic DNA Standards

- Authenticated ATCC nucleic acids
- Fluorescent dye-based quantification
- Mixed in even or staggered proportions based on genome copy number
- Store at -20°C

Assay development, optimization, verification, and quality control

ATCC® Microbiome Standards Portfolio



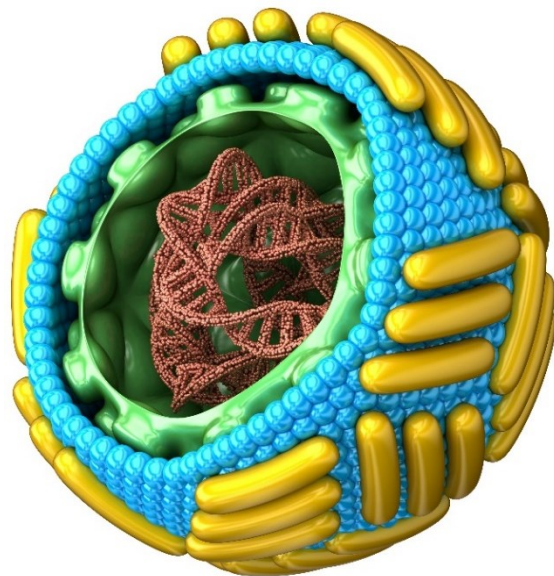
Preparation	ATCC® Catalog No.	Number of Organisms	Composition	Complexity	Importance
Genomic DNA	MSA-1000™	10	Even	Medium	Standards for assay development and optimization
	MSA-1001™	10	Staggered	Medium	
	MSA-1002™	20	Even	High	
	MSA-1003™	20	Staggered	High	
Whole cell	MSA-2003™	10	Even	Medium	
	MSA-2002™	20	Even	High	
Genomic DNA	MSA-4000™	11	Staggered	Medium	NGS-based pathogen detection
Genomic DNA	MSA-3000™	6	Even	Low	Environmental studies
	MSA-3001™	10	Even	Medium	
	MSA-3002™	10	Staggered	Medium	

Site-specific Microbiome Standards



Standard	Preparation	ATCC® Catalog No.	Number of Organisms	Importance
Oral	Whole cell	MSA-2004™	6	<ul style="list-style-type: none"> • Mock microbial communities representing the oral, skin, gut, and vaginal microbiomes • Comprises normal and atypical flora • Anaerobic and aerobic microbial strains • A combination of Gram-positive and -negative bacterial cultures • Even composition
	Genomic DNA	MSA-1004™		
Skin	Whole cell	MSA-2005™	6	
	Genomic DNA	MSA-1005™		
Gut	Whole cell	MSA-2006™	12	
	Genomic DNA	MSA-1006™		
Vaginal	Whole cell	MSA-2007™	6	
	Genomic DNA	MSA-1007™		

ATCC Virome Standards



Composition of Virome Standards

Human herpesvirus 5 strain AD169 (ATCC® VR-538™)

Human mastadenovirus strain F (ATCC® VR-931™)

Influenza B virus strain B/Florida/4/2006 (ATCC® VR-1804™)

Zika virus strain MR 766 (ATCC® VR-1838™)

Reovirus 3 strain Dearing (ATCC® VR-824™)

Human respiratory syncytial virus strain A2 (ATCC® VR-1540™)

Standard	Preparation	ATCC® Catalog No.	Number of Organisms	Specification (ddPCR™)	Applications
Virome	Virus Mix	MSA-2008™	6	2×10^3 genome copies/ μ L per virus	Standards for virome assay development, optimization, verification, and validation; evaluating reproducibility; and use as a daily run quality control
	Nucleic Acid Mix	MSA-1008™	6	2×10^4 genome copies/ μ L per virus	

Spike-in Standards (3 Strain Tagged Mix)

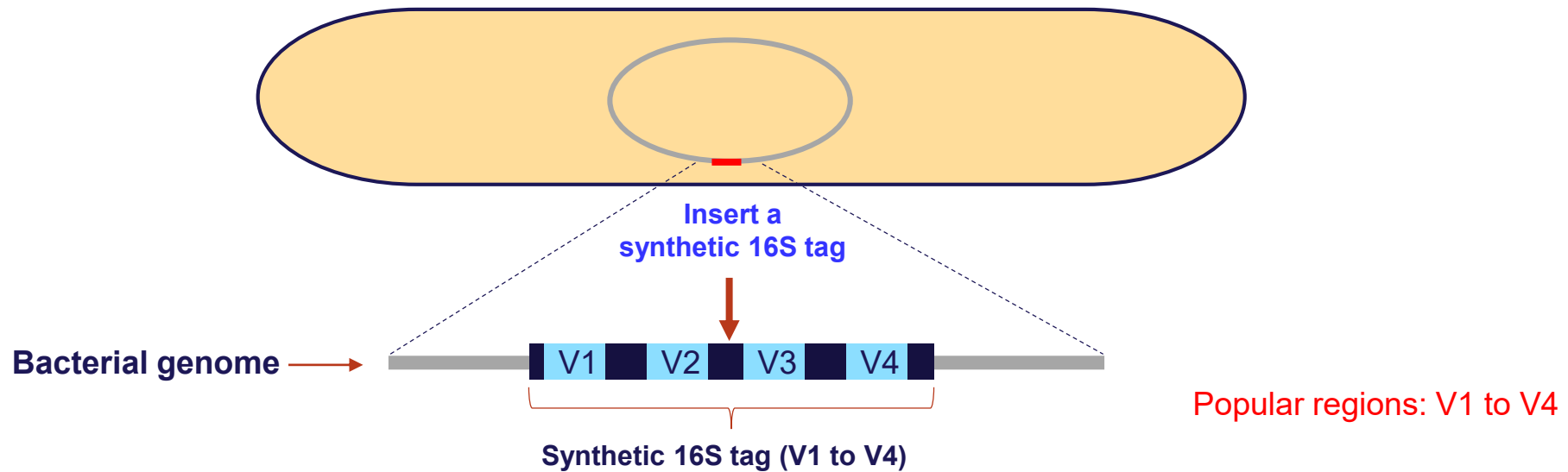


Standard	Preparation	ATCC® Catalog No.	Number of Organisms	Importance
3 Strain Tagged Mix	Whole cell	MSA-2014™	3	<ul style="list-style-type: none"> • Microbiome measurements and data normalization • 16S rRNA and shotgun assay verification, validation, and quality control
	Genomic	MSA-1014™		

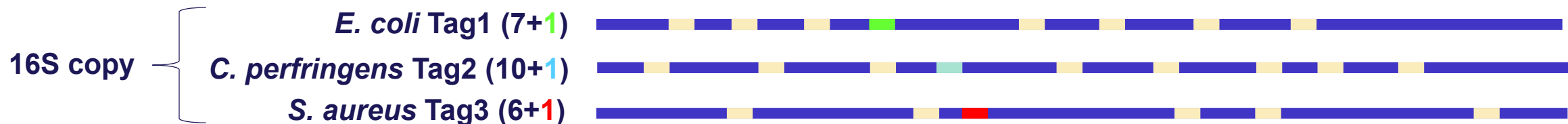
Spike-in Composition

Species	Gram Stain	Genome size (Mb)	Tag size (bp)	G/C Content (%)	16S Copies	Tag copies	Cells per vial
<i>Escherichia coli</i> Tag1	Negative	4.59	829	50.8	7	1	2 x 10 ⁷
<i>Clostridium perfringens</i> Tag2	Positive	3.25	799	29.0	10	1	2 x 10 ⁷
<i>Staphylococcus aureus</i> Tag3	Positive	2.70	833	32.8	6	1	2 x 10 ⁷

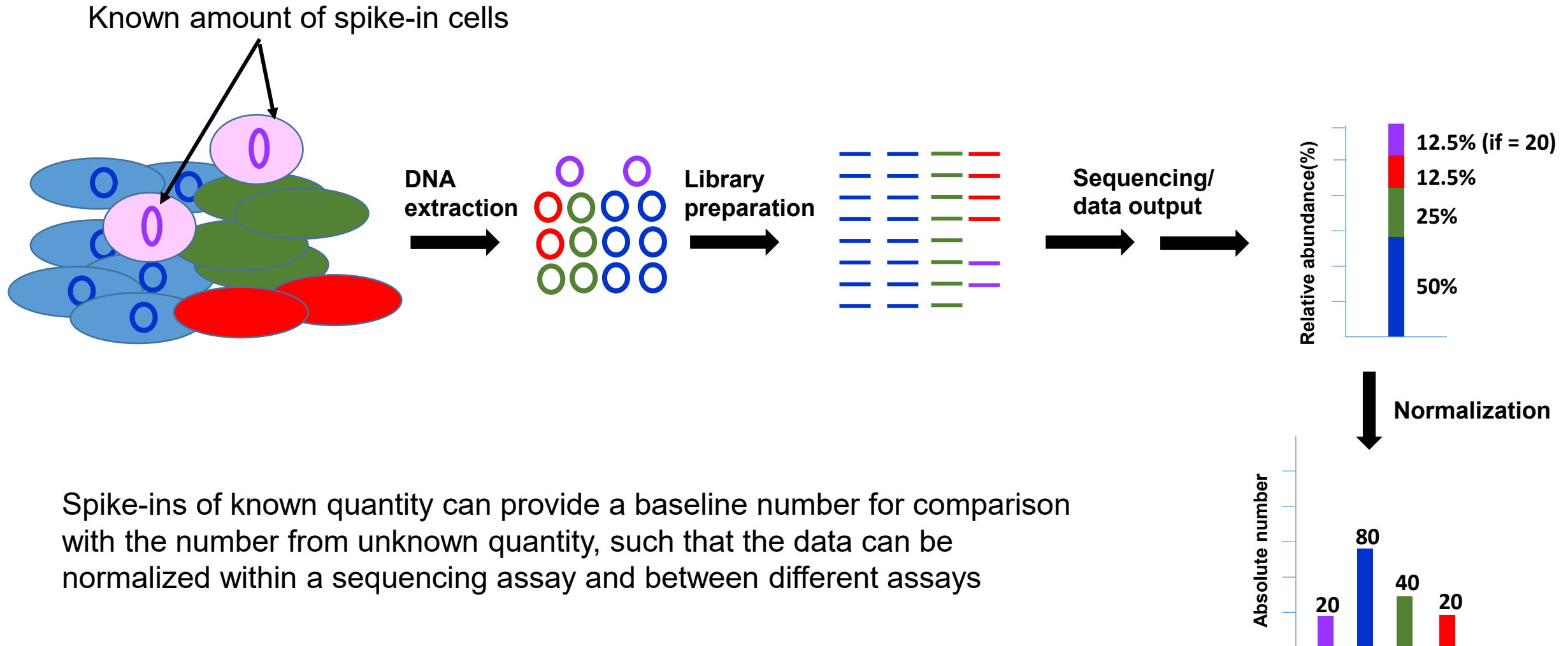
Engineering Synthetic 16S Tag Into Bacterial Genome



Bacterial Strains	Gram Stain	Genome G/C (%)	Insertion loci	BSL	
<i>Escherichia coli</i> Tag1	Negative	50.8	Beta galactosidase	1	
<i>Clostridium perfringens</i> Tag2	Positive	29.0	Theta-toxin	2	
<i>Staphylococcus aureus</i> Tag3	Positive	32.8	O-antigen polymerase	2	



Spike-in Internal Control for Normalization



Spike-ins of known quantity can provide a baseline number for comparison with the number from unknown quantity, such that the data can be normalized within a sequencing assay and between different assays

Spike-in: Calculating Absolute Numbers in a Sample

Microorganisms	Cell# in Sample X	Relative abundance in X	Cell# in Sample Y	Relative abundance in X
A	6,000	60.00%	500	45.45%
B	2,000	20.00%	250	22.73%
C	2,000	20.00%	350	31.82%



Mix with a known amount of spike-in cells (or DNA)

Microorganisms	Cell# in Sample X	Relative abundance in X	Cell# in Sample Y	Relative abundance in Y
A	6,000	59.41%	500	41.67%
B	2,000	19.80%	250	20.83%
C	2,000	19.80%	350	29.17%
Spike-in S	100	0.99%	100	8.33%

In sample X, if 0.99% Spike-in S = 100, then $19.80\% \text{ B} = 19.80\% \div 0.99\% \times 100 = 2,000$

In sample Y, if 8.33% Spike-in S = 100, then $20.83\% \text{ B} = 20.8\% \div 8.33\% \times 100 = 250$

Total cell#: In sample X, if 0.99% Spike-in S = 100, then total cell# of the sample X = $99\% \div 0.99\% \times 100 = 10,000$

Criterion of a spike-in bacterial sequence: Uniqueness

Resources

Technical Data Sheet: 3 Strain Tagged Whole Cell Even Mix

ATCC | Credible leads to Incredible™

Login | Create a Profile | Quick Order | 🛒 (0 item)

Search by Ke

Products Services Standards Resources Support About

Home ▶ Products ▶ All Products ▶ MSA-2014™

3 Strain Tagged Whole Cell Even Mix (ATCC® MSA-2014™)

GENERAL INFORMATION DOCUMENTATION

Permits

These permits may be required for shipping this product:

- Customers located in the state of Hawaii will need to contact the Hawaii Department of Agriculture to determine if an Import Permit is required. A copy of the permit or documentation that a permit is not required must be sent to ATCC in advance of shipment.

Basic Documentation

[Product Sheet](#) [Certificate of Analysis](#) [SDS](#)

Other Documentation

[Microbiome Standards](#) [Technical Data Sheet](#)

3 Strain Tagged Mix
ATCC® MSA-2014™
freeze-dried

Qty:

CUS
SEE

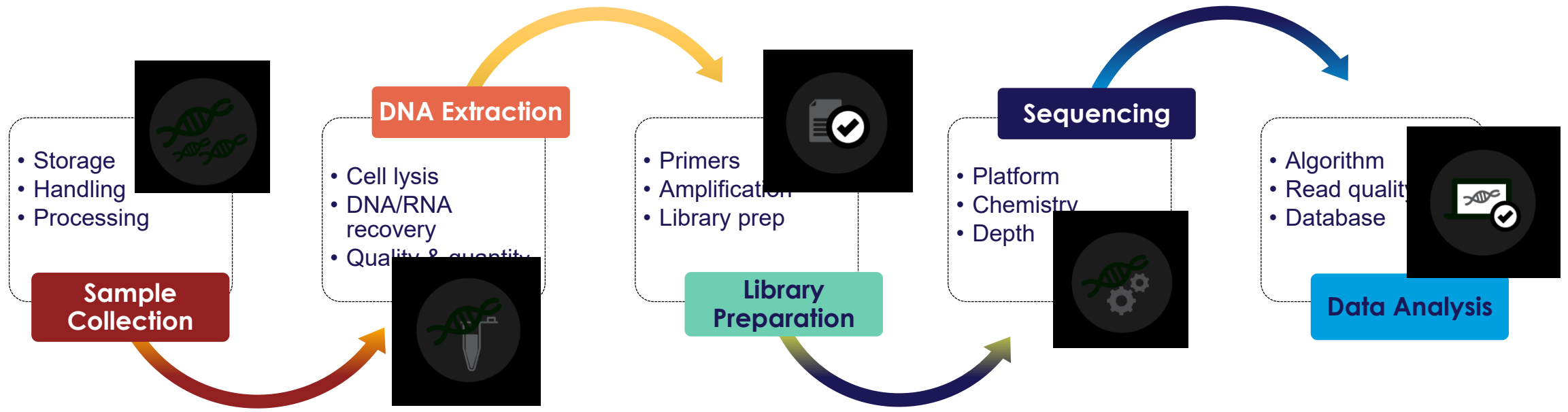
Coming Soon - Mycobiome (MSA-2010™ and MSA-1010™)

MSA-2010™	MSA-1010™
Whole cell	Genomic
In production now	Coming soon
Release date: 23Sep2019	Release: Q4



ATCC® No.	Species Name	Genome size (Mb)	Relevancy
MYA-4609™	<i>Aspergillus fumigatus</i> Af293	28.8	<u>Opportunistic</u> , airborne pathogen that is responsible for 90% of fungal infections in immunocompromised patients
10231™	<i>Candida albicans</i>	17.1	<u>Commensal</u> fungus of the oral cavity that can form biofilms on denture surfaces leading to mucosal infections
2001™	<i>Candida glabrata</i>	12.3	<u>Commensal</u> fungus of the oral cavity and human gut that can acquire resistance to azole antifungals leading to infection
208821™	<i>Cryptococcus neoformans</i> var. <i>grubii</i>	18.9	Responsible for cryptococcal meningitis in <u>immunosuppressed</u> patients
MYA-4612™	<i>Malassezia globosa</i> CBS 7966	9.0	Part of the <u>normal skin flora</u> but can be responsible for skin diseases such as dandruff, dermatitis, and folliculitis
201390™	<i>Saccharomyces cerevisiae</i>	12.2	Eukaryotic model organism <u>used to study gene expression</u> , signal transduction, the cell cycle, and metabolism
9533™	<i>Trichophyton interdigitale</i>	21.9	Can infect skin and nails to cause <u>chronic infections</u> such as athlete's foot and ringworm
204094™	<i>Cutaneotrichosporon dermatis</i> (CBS2043)	23.3	<u>Emerging opportunistic</u> agent of invasive fungal infections, particularly in severely immunocompromised patients
10106™	<i>Penicillium chrysogenum</i> Thom	32.5	<u>Spore-former</u> and source of beta-lactam antibiotics such as penicillin
36031™	<i>Fusarium keratoplasticum</i> (<i>F solani</i>)	48.6	Filamentous, <u>opportunistic</u> pathogen that causes fungal keratitis

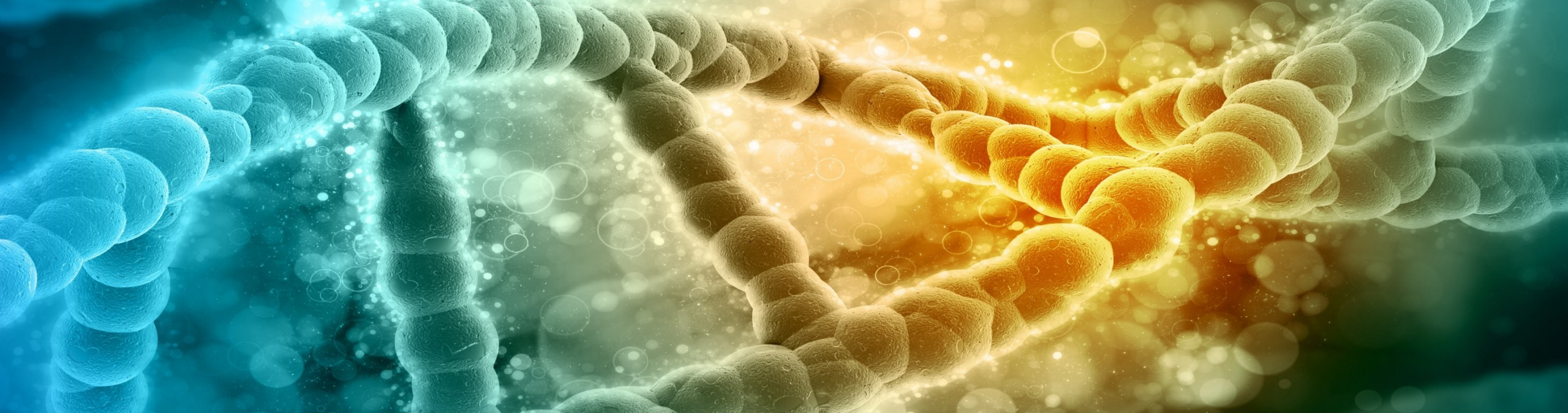
Microbiome Workflow, Biases, and Standardization



Ideal use - Whole Cell Microbiomes (full process control)

Ideal use – Genomic Microbiomes (assay control)

One Codex Modules



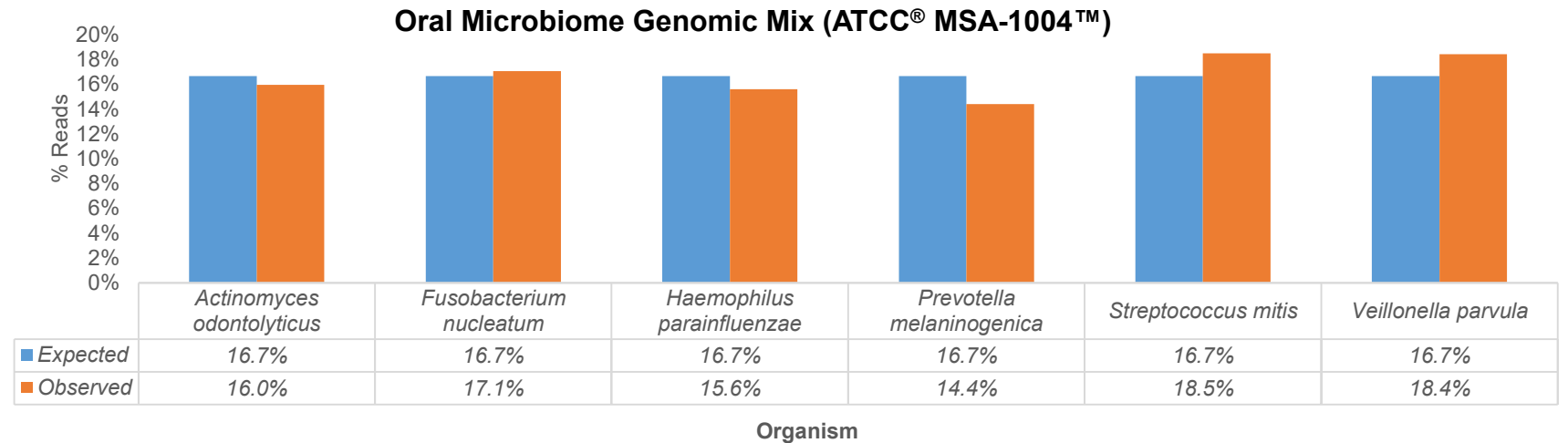
Applications of standards in microbiome research

- DNA extraction
- 16S rRNA amplification and library kits
- NGS platforms
- Bioinformatics and databases

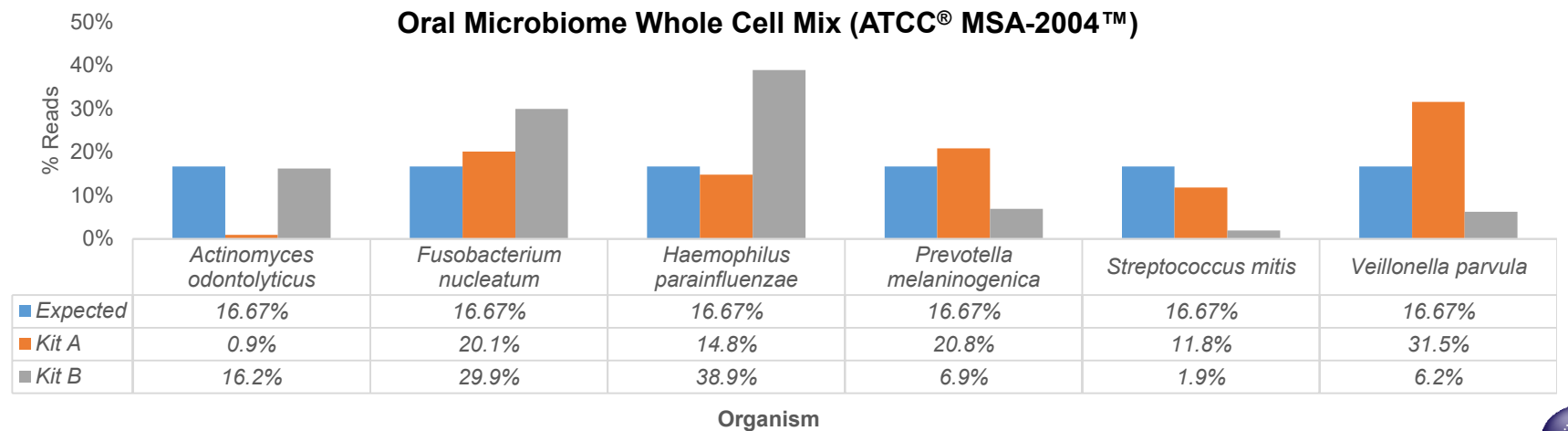
Genomic Versus Whole Cell Standards

DNA extraction methods are not perfect

Shotgun metagenomic analysis of the Oral Microbiome Genomic Mix



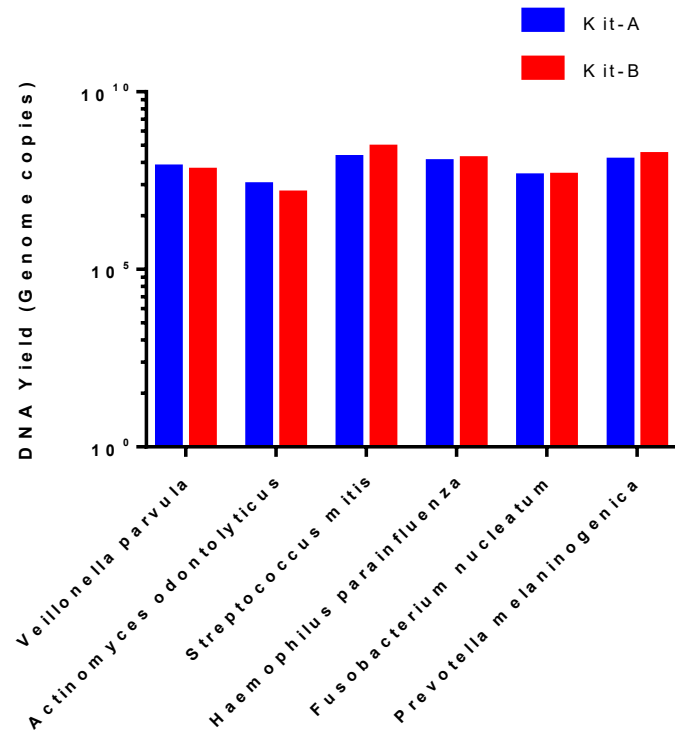
DNA extraction from the Oral Microbiome Whole Cell Mix with two different kits followed by shotgun metagenomic analysis



Assess Biases in DNA Extraction

Compare different pre-treatments and extraction methods, optimize protocols, and validate different kits

DNA extraction from individual components of the Oral Whole Cell Mix



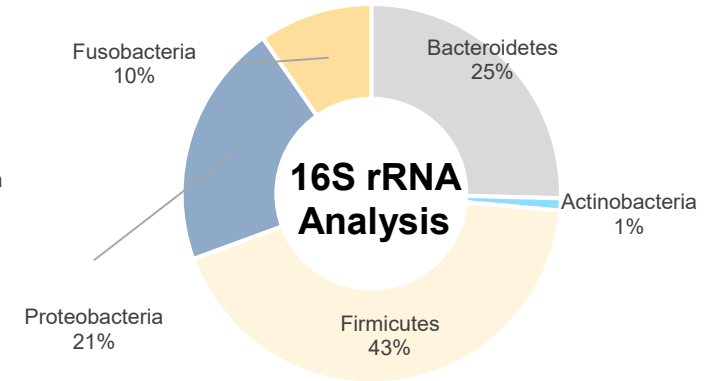
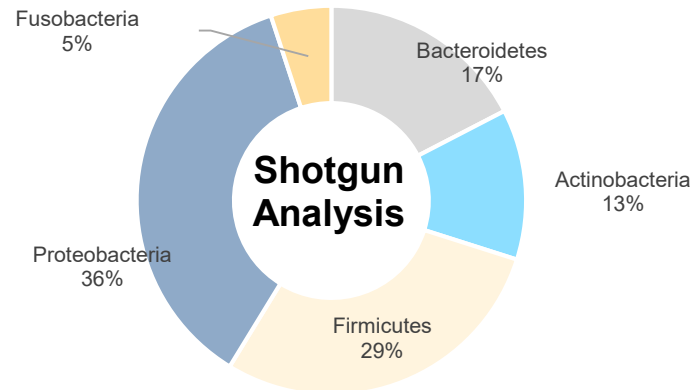
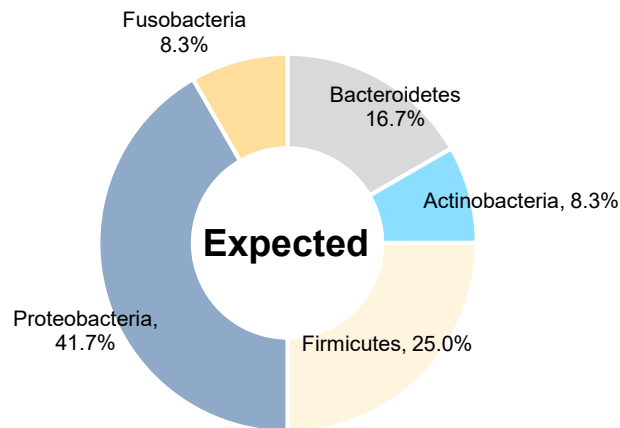
Organism	Number of Cells per Component	Gram Stain	Genome size Mb	%GC
<i>Actinomyces odontolyticus</i>	~2x10 ⁷	+	2.39	65.5
<i>Fusobacterium nucleatum</i>		-	2.17	27.2
<i>Haemophilus parainfluenzae</i>		-	2.12	39.3
<i>Prevotella melaninogenica</i>		-	3.17	35.1
<i>Streptococcus mitis</i>		+	1.83	40.5
<i>Veillonella parvula</i>		-	2.16	38.6

Gut Whole Cell Standard

Profiling of gut microbiome standard at the phylum, genus, and species level

The Gut Whole Cell Microbiome Standard (ATCC® MSA-2006™) can be used as a full process control for shotgun and 16S rRNA assays

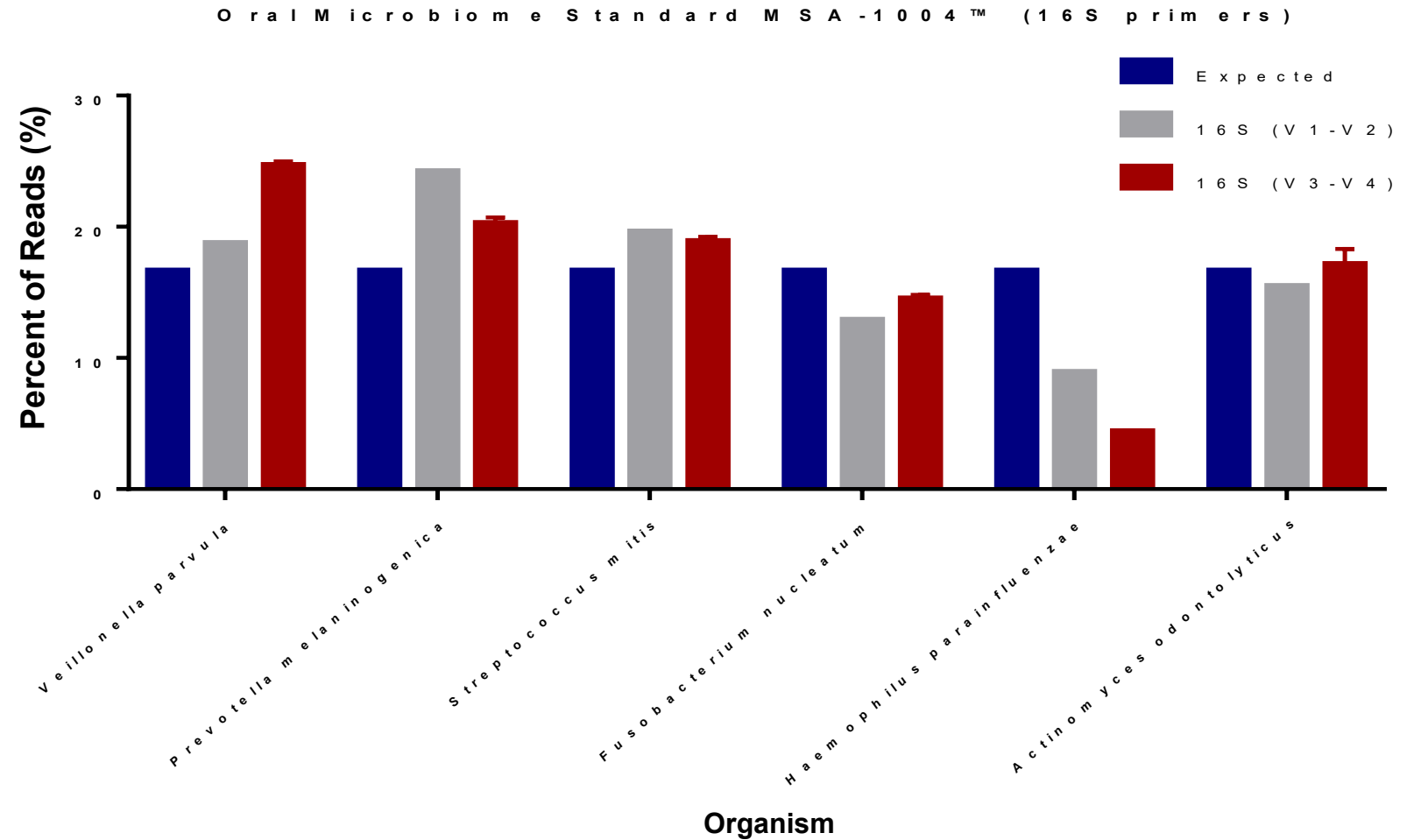
Phylum	Expected	Strain	Expected	Observed-Shotgun	Observed-16S (V1V2)
Bacteroidetes	16.6%	<i>Bacteroides fragilis</i>	8.3%	12.3%	17.0%
		<i>Bacteroides vulgatus</i>	8.3%	8.6%	8.4%
Actinobacteria	8.33%	<i>Bifidobacterium adolescentis</i>	8.3%	12.0%	1.0%
Firmicutes	25.0%	<i>Clostridium difficile</i>	8.3%	16.5%	29.0%
		<i>Enterococcus faecalis</i>	8.3%	3.1%	1.6%
		<i>Lactobacillus plantarum</i>	8.3%	8.1%	12.3%
Proteobacteria	41.7%	<i>Enterobacter cloacae</i>	8.3%	10.6%	4.0%
		<i>Escherichia coli</i>	8.3%	6.6%	3.3%
		<i>Helicobacter pylori</i>	8.3%	3.8%	7.7%
		<i>Salmonella enterica</i>	8.3%	4.9%	2.2%
		<i>Yersinia enterocolitica</i>	8.3%	8.8%	3.6%
Fusobacteria	8.3%	<i>Fusobacterium nucleatum</i>	8.3%	4.8%	9.7%



16S Amplicon-based Analysis: Primer Selection

Compare different primer sets, optimize amplification steps, and validate 16S analysis methods

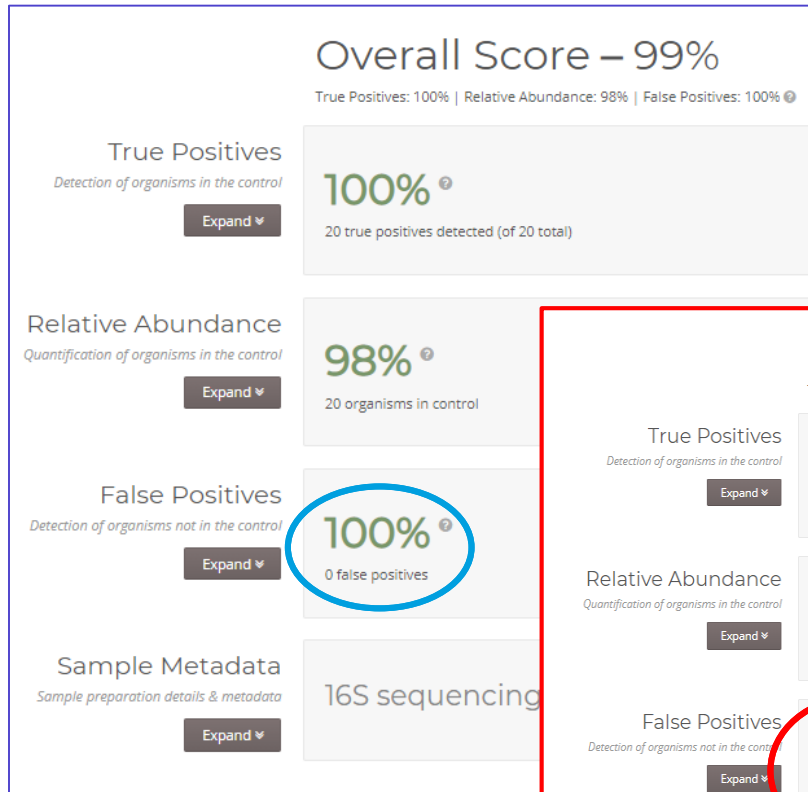
16S rRNA analysis of the Oral Genomic DNA Standard via two primer sets



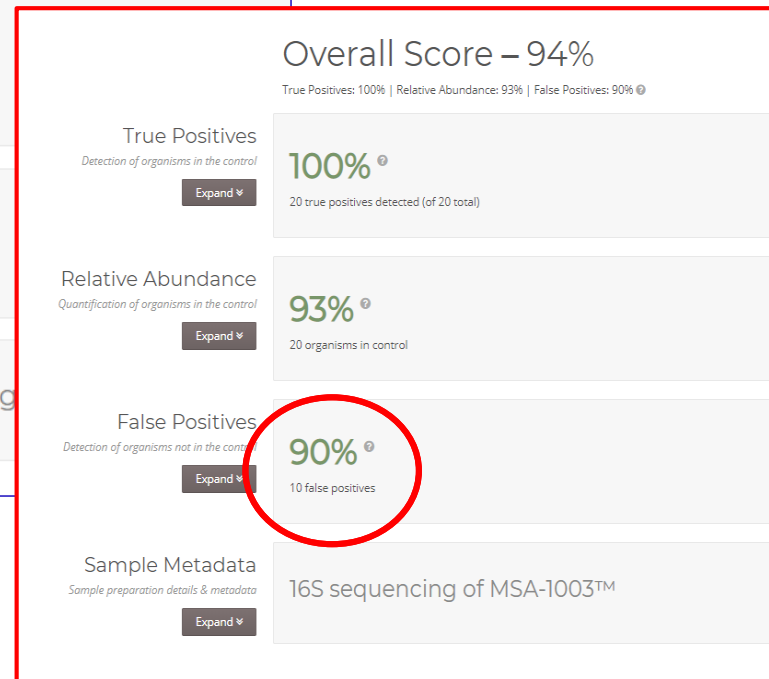
Comparing Library Preparation Kits

The LoopSeq™ 16S rRNA long-read method allows highest sequence accuracy and species-level taxonomy

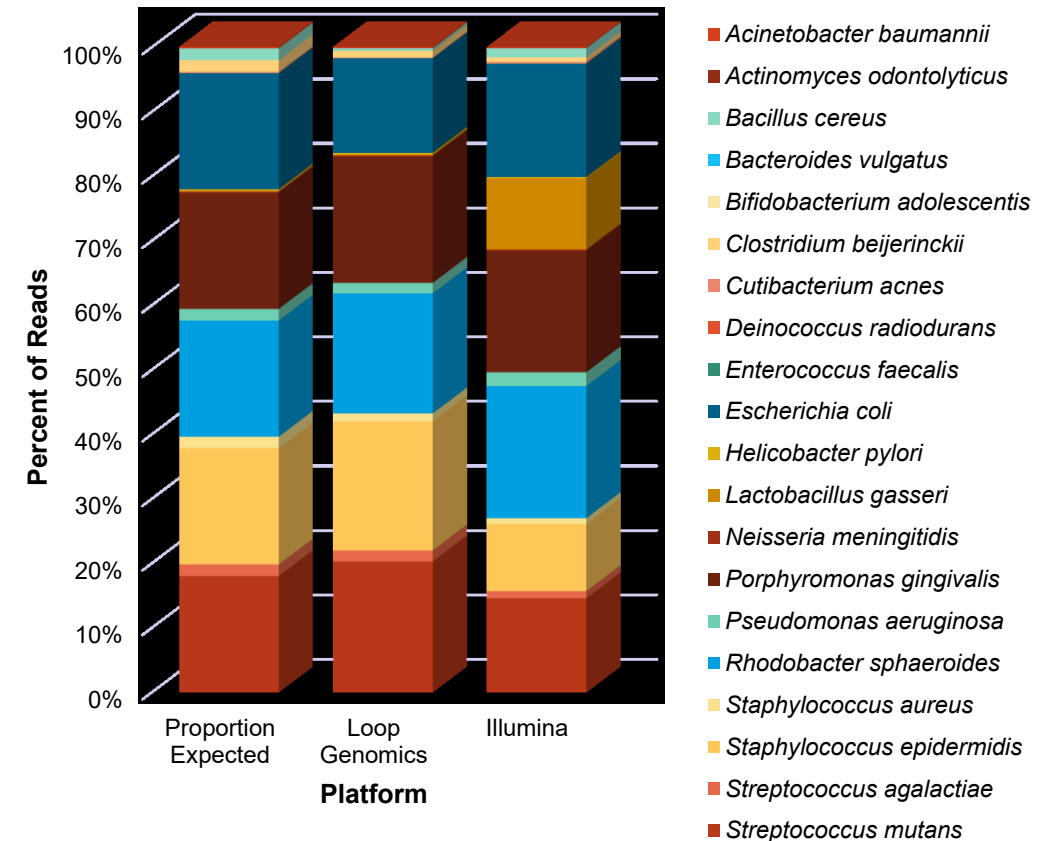
Loop Genomics



Short Reads



Genomic DNA (ATCC® MSA-1003™)



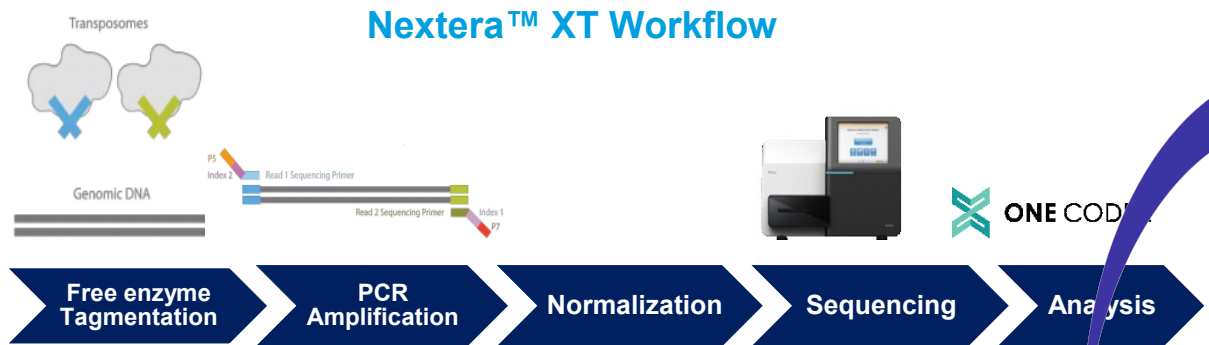
MSA-1004™/MSA-2004™ Oral Microbiome

Name	Gram Stain	% GC	Genome Size (Mb)	16S rRNA Copies
<i>Actinomyces odontolyticus</i>	POS	65.5	2.39396	2
<i>Fusobacterium nucleatum</i>	NEG	27.2	2.1745	5
<i>Haemophilus parainfluenzae</i>	NEG	39.3	2.12476	9
<i>Prevotella melaninogenica</i>	NEG	35.1	3.16828	4
<i>Streptococcus mitis</i>	POS	40.5	1.83108	4
<i>Veillonella parvula</i>	NEG	38.6	2.16347	4

Comparing Library Preparation Kits

Nextera Flex enables uniform coverage of genomes of low GC content

Oral Microbiome Genomic DNA (ATCC® MSA-1004™)



Sample Composition

Name	Estimated Abundance
Actinomyces odontolyticus	46.07%
Prevotella melaninogenica	16.09%
Streptococcus mitis	16.04%
Veillonella parvula	12.65%
Haemophilus parainfluenzae	7.80%
Fusobacterium nucleatum	1.34%



Sample Composition

Name	Estimated Abundance
Streptococcus mitis	18.83%
Veillonella parvula	18.42%
Fusobacterium nucleatum	17.13%
Haemophilus parainfluenzae	15.68%
Actinomyces odontolyticus	15.54%
Prevotella melaninogenica	14.40%



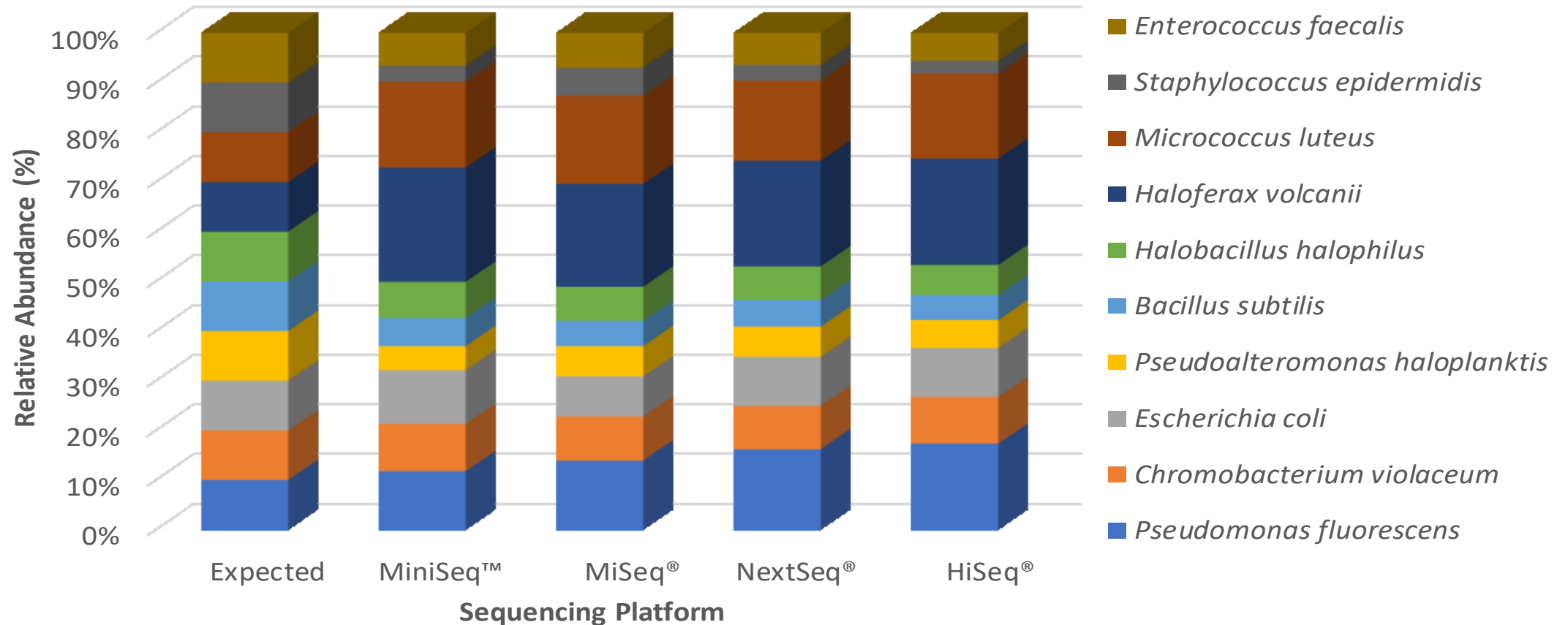


Evaluating NGS Platforms

Short-read Sequencing Platform: Illumina®

Assay reproducibility through different Illumina sequencing platforms

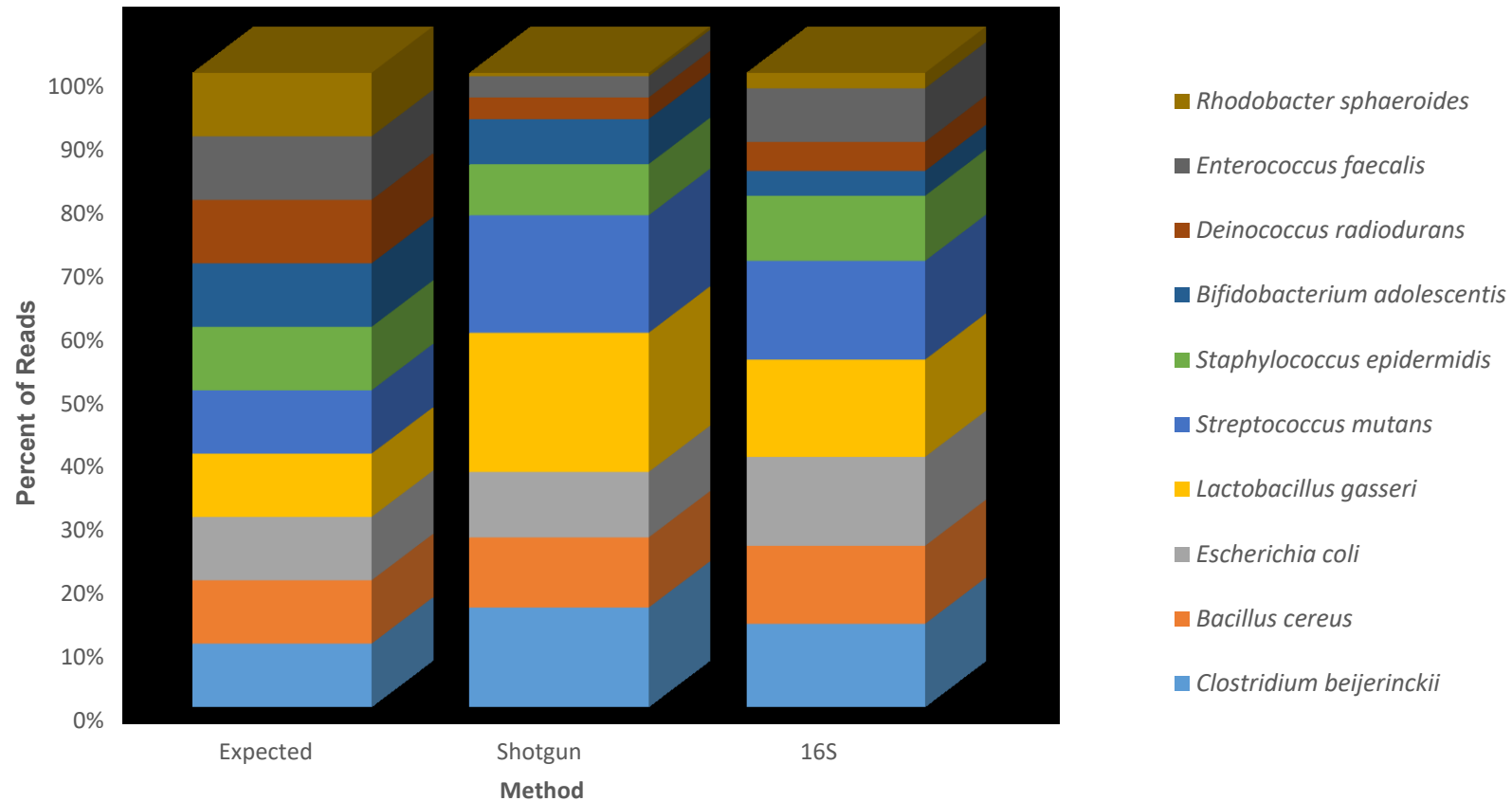
Shotgun Metagenomic Data (ATCC® MSA-3001™)



Short-read Sequencing Platform: Ion Torrent™

16S rRNA and shotgun data on the Ion GPM Platform (ATCC® MSA-1003™)

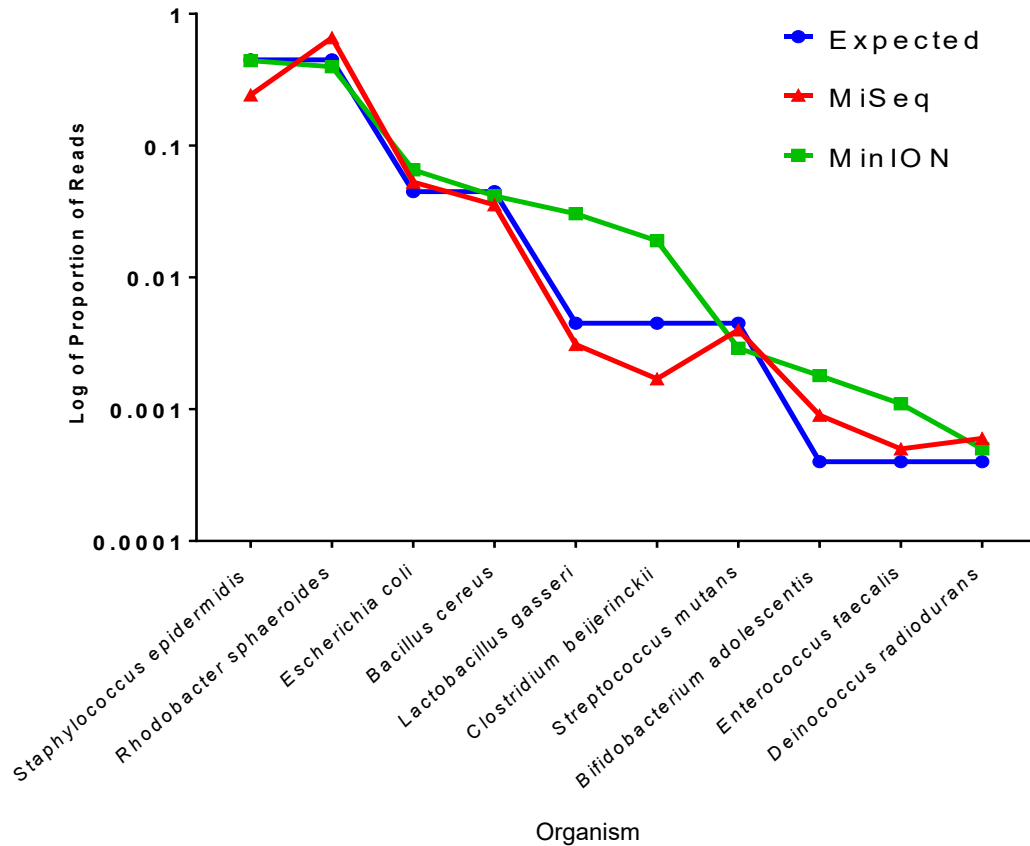
Shotgun vs 16S rRNA assay (V1/V2) (ATCC® MSA-1000™)



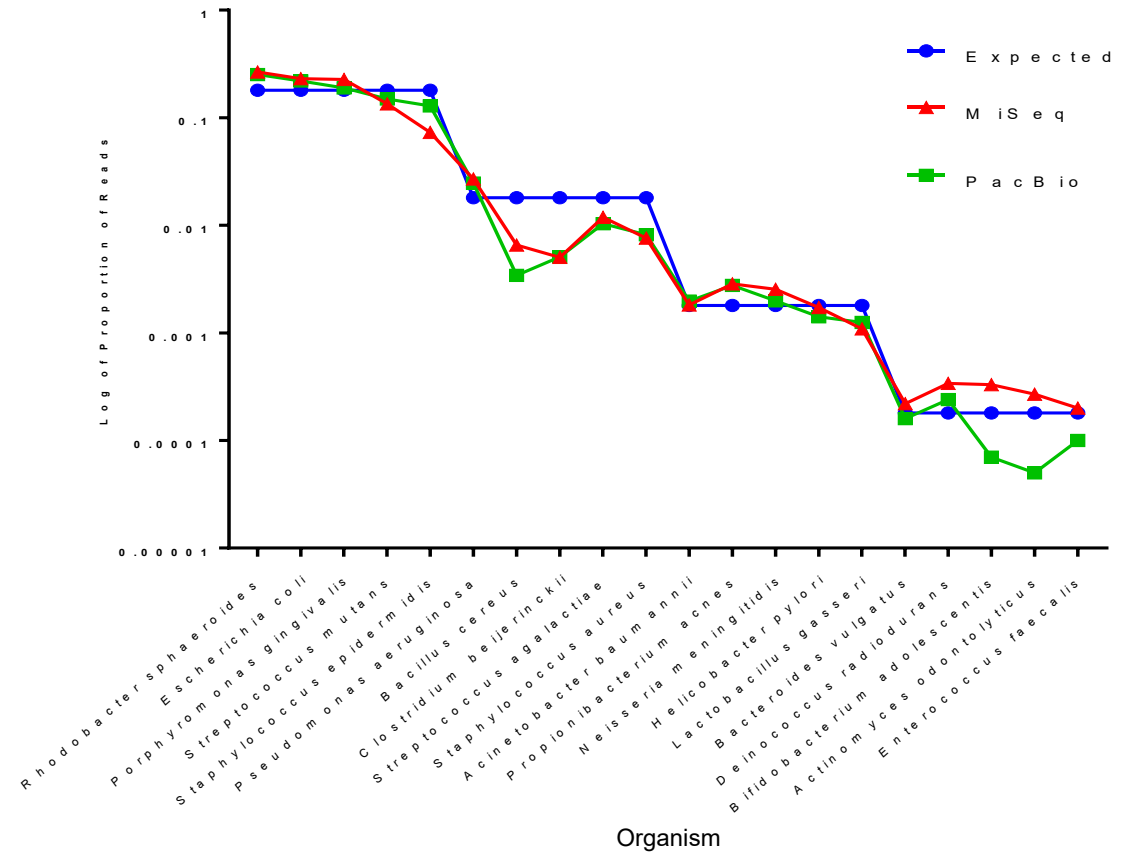
Shotgun Metagenomic Analysis: Short vs Long Reads

ATCC Microbiome Standards are technology agnostic

MSA-1001™ (Staggered 10 Strains)



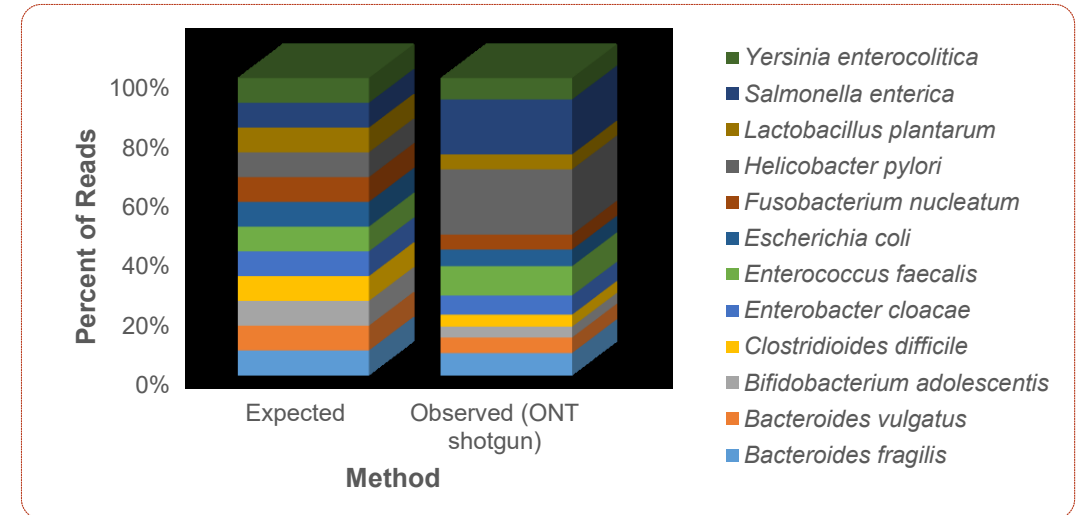
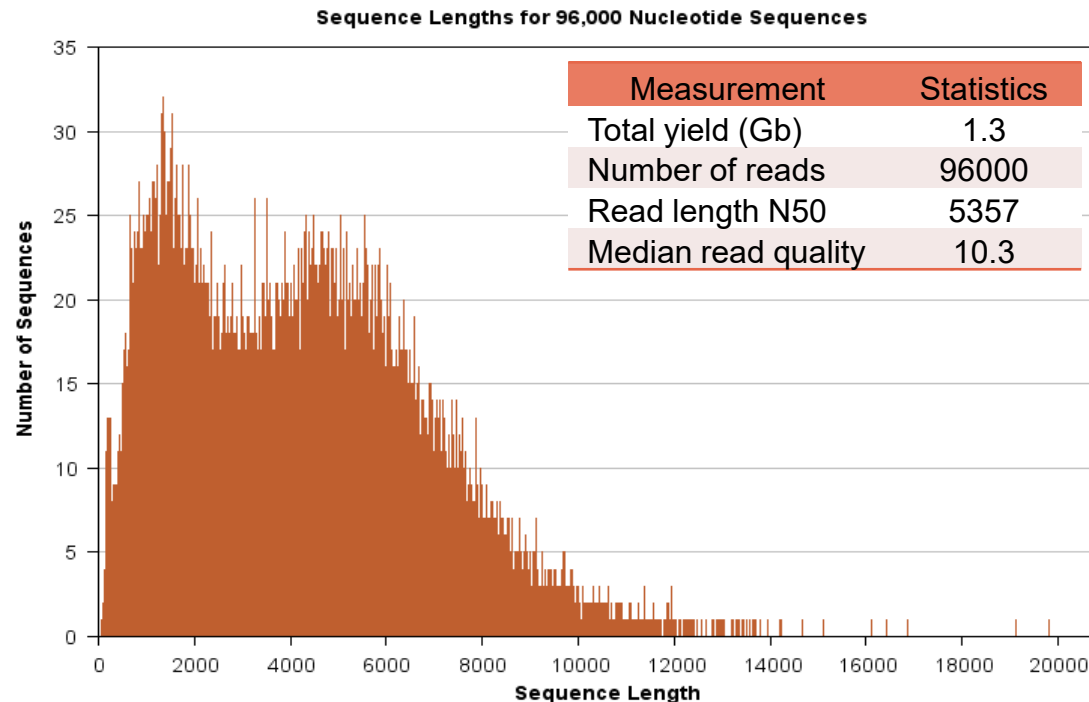
MSA-1003™ (Staggered 20 Strains)



Long-read Sequencing Platform: Nanopore®

One hour sequencing coverage was enough to identify all organisms in the mix with sufficient genome coverage

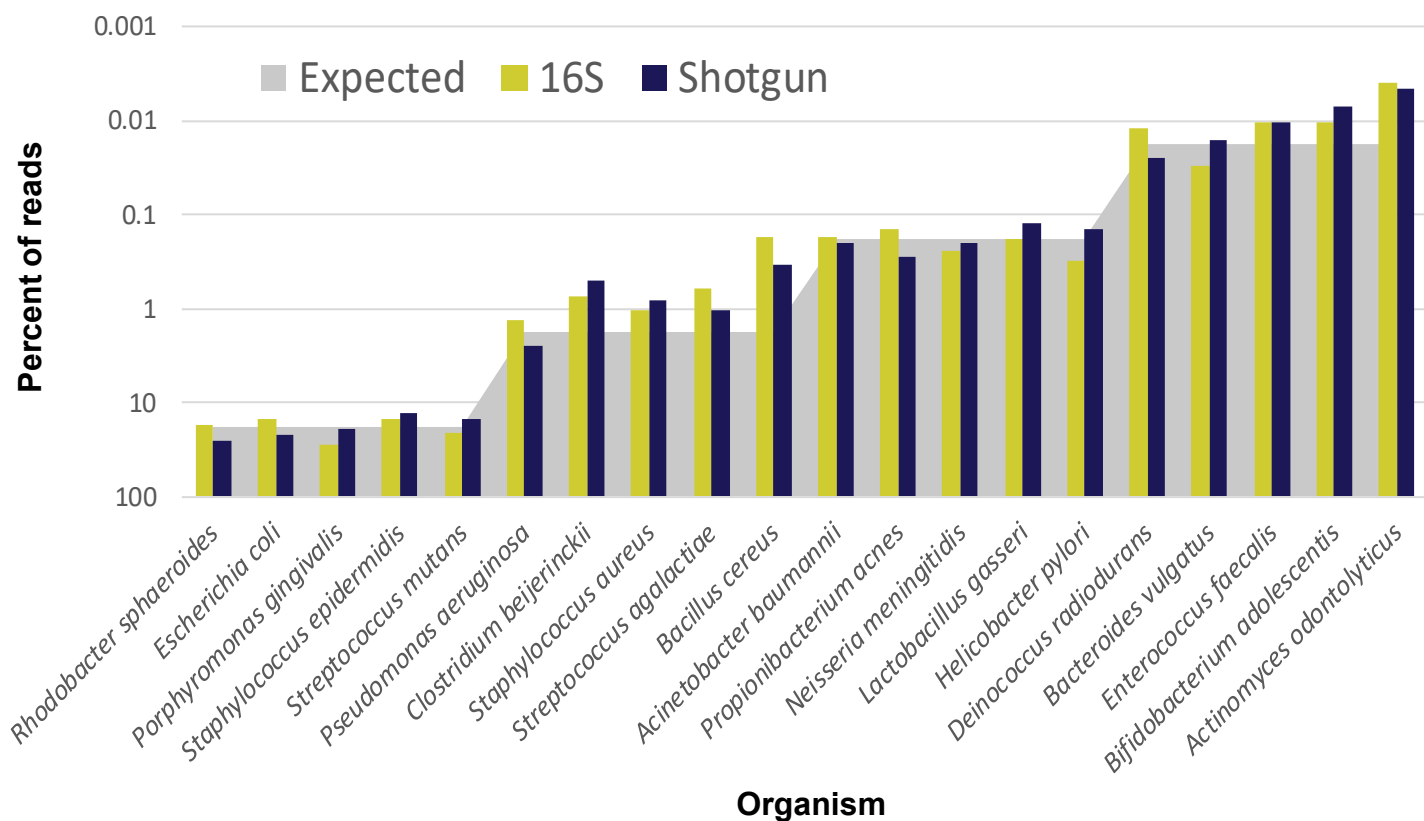
The Gut Microbiome Whole Cell Standard (ATCC® MSA-1006™) was analyzed via shotgun sequencing on the MinION platform



Organism	Genome Coverage (x)
<i>Enterobacter cloacae</i>	9.1
<i>Enterococcus faecalis</i>	14.1
<i>Bacteroides fragilis</i>	8.1
<i>Bacteroides vulgatus</i>	6.6
<i>Bifidobacterium adolescentis</i>	1.8
<i>Clostridioides difficile</i>	7.4
<i>Escherichia coli</i>	6.5
<i>Fusobacterium nucleatum</i>	4.6
<i>Helicobacter pylori</i>	16.6
<i>Lactobacillus plantarum</i>	6.0
<i>Salmonella enterica</i>	11.1
<i>Yersinia enterocolitica</i>	11.3

Long-read Sequencing Platform: PacBio®

16S rRNA (full-length) and shotgun data on the PacBio Sequel Platform (ATCC® MSA-1003™)



ATCC quality control score (One Codex)

One Codex Analysis	16S rRNA run 1	16S rRNA run 2	Shotgun run 1	Shotgun run 2
True positives	100%	100%	100%	100%
Relative abundance	95%	95%	97%	97%
False positives	100%	100%	88%	84%
Overall score	98%	98%	95%	95%



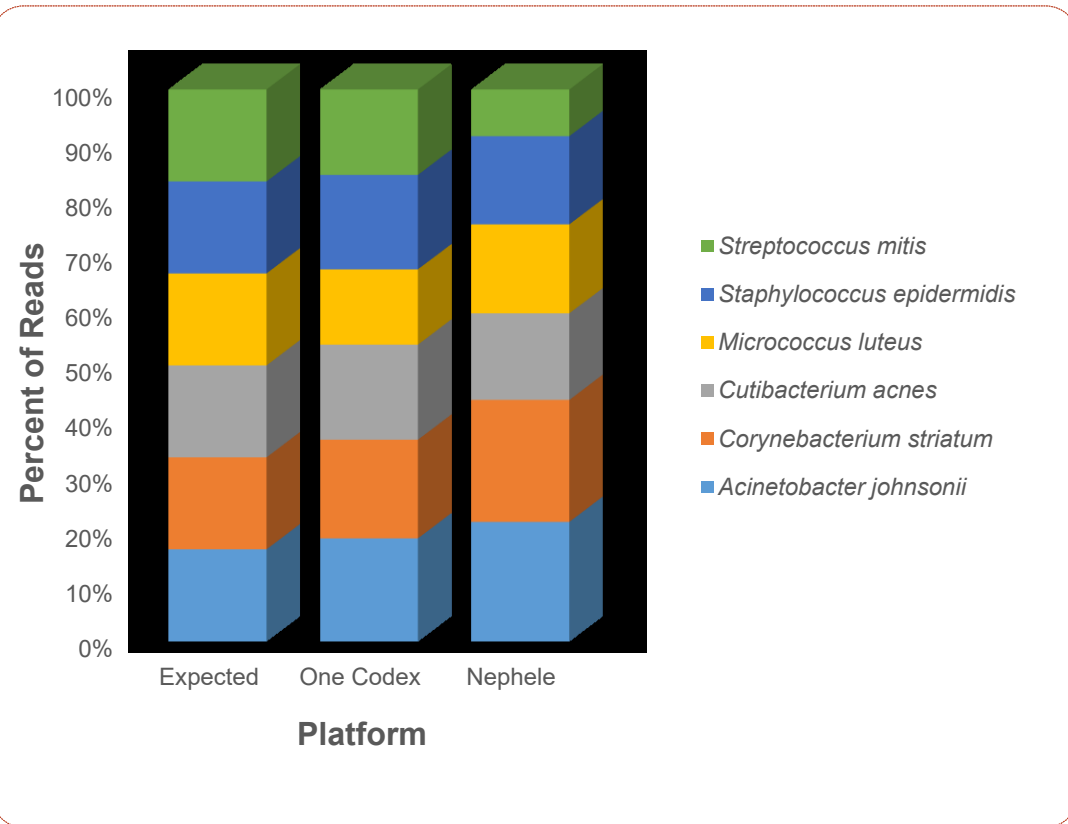
Comparing Bioinformatics and Databases

Data Analysis Using Different Databases

Evaluation of NGS data from microbiome standards in multiple analysis platforms and databases

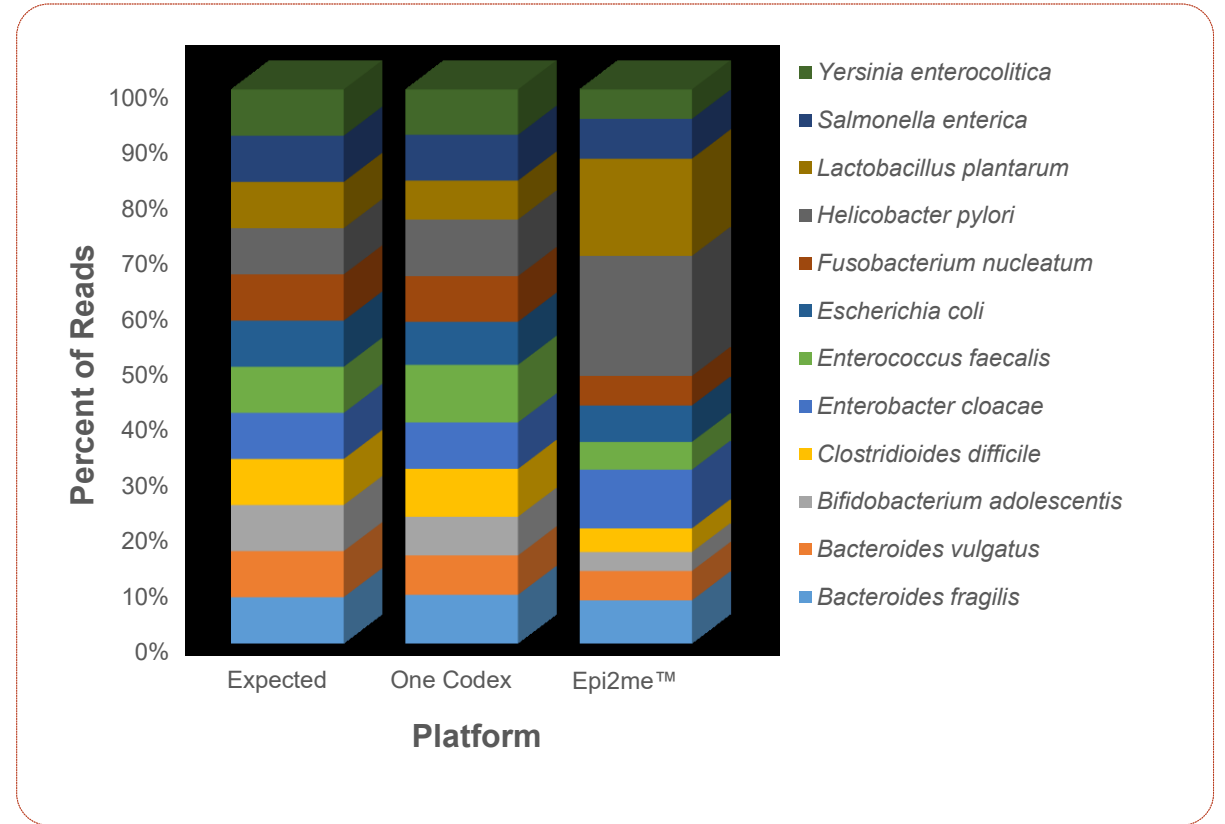
Nephele vs One Codex

Short-read sequencing data from the Skin Genomic DNA Mix (ATCC® MSA-1005™)



Epi2Me vs One Codex

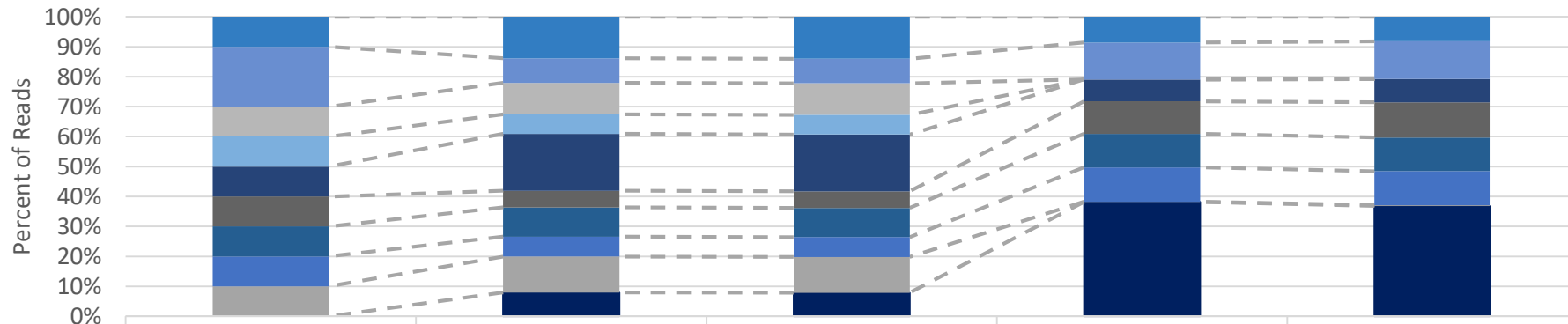
Long-read sequencing data from the Gut Genomic DNA Mix (ATCC® MSA-1006™)



Mycobiome Standards

Data analysis platform impacts strain identification and taxonomic resolution

Shotgun Analysis of Genomic DNA standards (ATCC® MSA-1010™)



	Expected	Databases A1	Databases A2	Databases B1	Databases B2
Aspergillus	10%	13.88%	13.99%	8.60	8.16
Candida	20%	8.17%	8.22%	12.31	12.61
Cryptococcus	10%	10.52%	10.51%	0.00	0.00
Cutaneotrichosporon	10%	6.53%	6.61%	0.00	0.00
Fusarium	10%	19.00%	18.94%	7.32	7.78
Malassezia	10%	5.53%	5.55%	10.86	11.74
Penicillium	10%	9.81%	9.76%	11.17	11.25
Saccharomyces	10%	6.63%	6.62%	11.48	11.35
Trichophyton	10%	12.01%	11.98%	0.05	0.29
Unclassified	0%	7.93%	7.82%	38.21	36.82

ATCC Data Analysis Solution



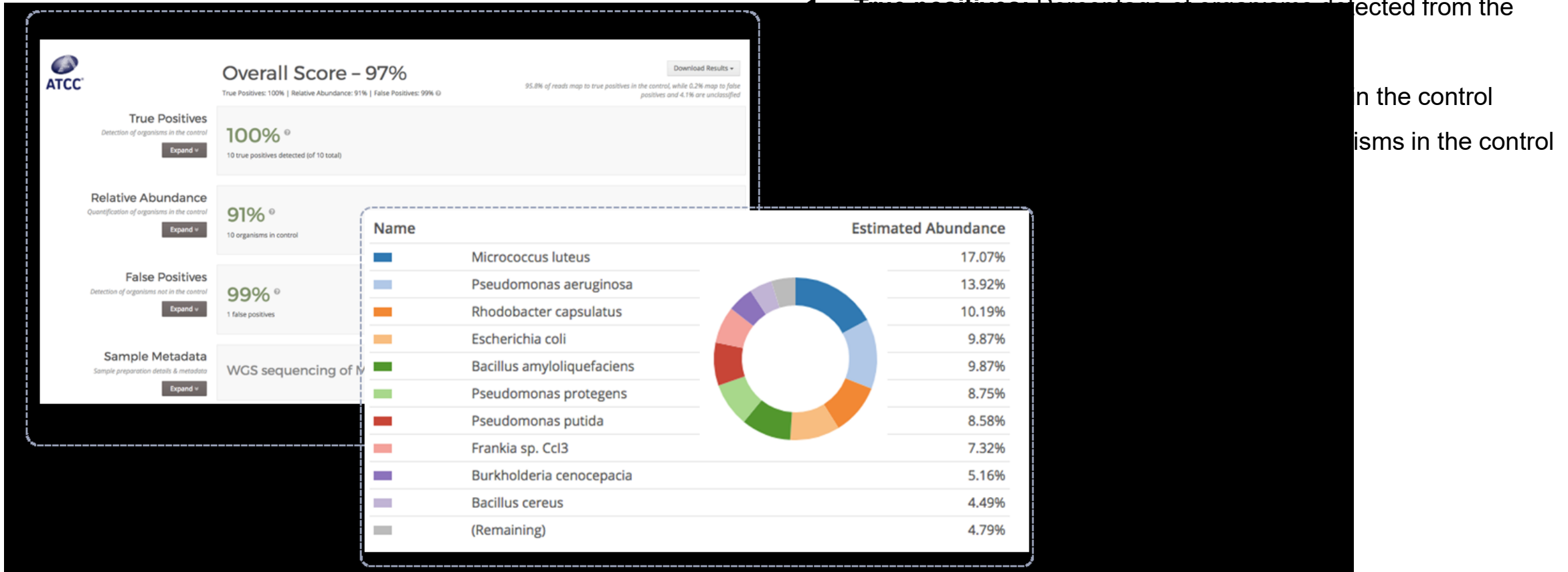
A screenshot of the ATCC ONE CODEX web interface. The left sidebar shows navigation options like 'Samples', 'Upload / Import', 'Search', 'Compare Analyses', 'Cluster View', 'Notebooks', 'Run Analyses', 'ATCC Standards', 'Help & Documentation', and 'ATCC'. The main content area is titled 'Microbiome Reference Standards' and lists six ATCC MSA products with their descriptions and 'Order now' links. The products are: ATCC MSA-2002™ (20 Strain Even Mix Whole Cell Material), ATCC MSA-2003™ (10 Strain Even Mix Whole Cell Material), ATCC MSA-2004™ (Oral Microbiome Whole Cell Material), ATCC MSA-2005™ (Skin Microbiome Whole Cell Material), ATCC MSA-2006™ (Gut Microbiome Whole Cell Material), and ATCC MSA-2007™ (Vaginal Microbiome Whole Cell Material). The interface also includes a 'Choose your ATCC product' section with 'Whole Cell' and 'Genomic DNA' tabs, a 'Sequencing' section with 'Shotgun' and '16S' tabs, and an upload area for 'FASTQ files'.

WORKFLOW:

1. Drag and drop Fastq files or export via cloud
2. Choose your ATCC product and analysis (16s and shotgun)
3. Download your reports

ATCC Data Analysis Solution

RESULTS ARE PROVIDED ON A SCORECARD REPORTS:



ATCC Data Analysis Solution



Mock Microbial Communities

- Genomic DNA and whole cell standards
- Even and staggered mixtures comprising 10 or 20 strains
- Environmental and pathogen mixtures



Site-specific Standards

- Genomic DNA and whole cell standards
- Even mixtures of 6-12 strains
- Bacterial strains prevalent in the oral, skin, gut, and vaginal microbiome



Spike-In Standards

- Recombinant strains with a unique DNA tag stably integrated into the chromosome
- Recombinant standards include the Gram-negative and Gram-positive bacteria



New Products

- Genomic DNA and whole cell mock communities representing:
 - Virome
 - Mycobiome

Bundled with data analysis on the One Codex platform

www.atcc.org/Microbiome

Conclusion

- ✓ The use of standards in all areas of research is absolutely essential.
- ✓ Microbiome research is challenging and flush with biases
- ✓ The ATCC® Microbiome Standards portfolio and upcoming new products
- ✓ Applications of standards in microbiome research
 - ✓ Extraction method, assay variability, NGS library preparation, and bioinformatics analysis

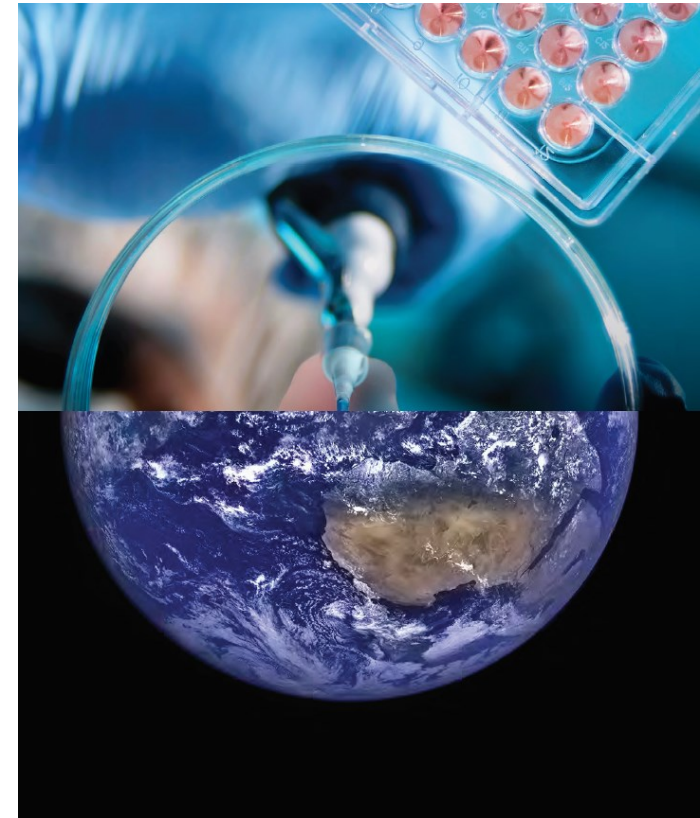
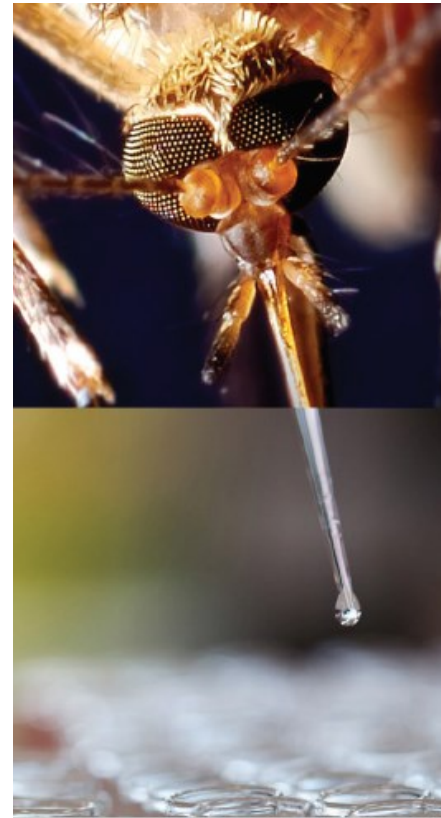
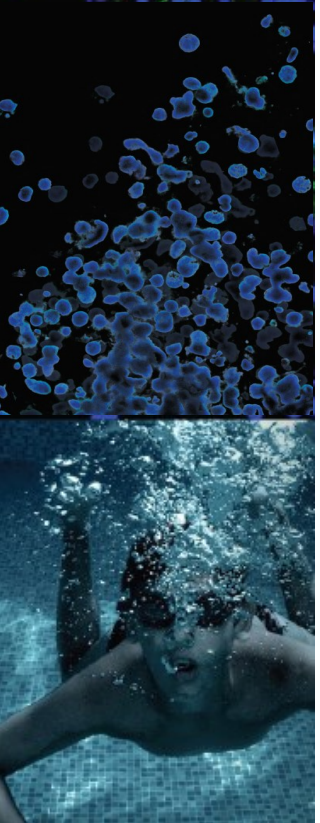
Acknowledgements

- Monique Hunter, MS
- Anna McCluskey, BS
- Stephen King, MS
- Juan Lopera, PhD
- Cara Wilder, PhD
- Dev Mittar, PhD
- Scott Tighe, PhD, UVM
- Denise O'Sullivan, PhD, LGC
- Nick Greenfield, MA, One Codex
- Pat Gillevet, PhD, Microbiome Analysis Center, GMU
- Rohan Patil, Microbiome Analysis Center, GMU
- Stefan Green, PhD, UIC (ABRF-MGRG)
- Joan Wong, PhD, PACBIO®
- Tony Lialin, Loop Genomics



Thank you!
Questions?

Credible Leads to Incredible™



Cultivating collaboration to support global health

Upcoming webinars

- STR Profiling for Mouse Cell Lines: Another Tool to Combat Cell Line Misidentification | September 12, 12:00 ET

www.atcc.org/webinars

