Metagenomics and Metatranscriptomics Analysis Using ATCC Whole Cell Microbiome Standards

Poster: 1093
Date: June 9, 2018

Phone: 800.638.6597
Email: Tech@atcc.org

Ian Miller, Ph.D.[1], Juan Lopera, Ph.D.[2], Jason Kwan, Ph.D.[1], and Dev Mittar, Ph.D.[2]
[1]Pharmaceutical Science Division, University of Wisconsin-Madison, WI 53706; [2]ATCC, Manassas, VA 20110

Abstract

Metagenomic analyses have provided insight into the abundance and taxonomic profiles of microbiomes. As the clinical and biotechnological applications of microbiome research continue to expand, researchers are now leveraging metatranscriptomics to explore organism-level function in microbiome samples via RNA-Seq technology. To facilitate this research, we have developed whole cell mock community standards representing complex mixtures of diverse bacterial species. In the following study, we used these mock community standards to create metagenomics and metatranscriptomic profiles to validate the bioinformatics analysis of whole genome shotgun sequencing and RNA-Seq data. DNA and RNA were extracted from 100 µL aliquots of the whole cell microbiome standards (ATCC® MSA-2002™ and ATCC® MSA-2003™), and shotgun sequencing of DNA and RNA was performed using the Illumina® platform. Sequencing analysis was performed by the Kwan lab (University of Wisconsin-Madison) using the Autometa bioinformatics pipeline, which automated the taxonomic separation of genomic DNA contigs and profiled their RNA expression from metagenomes[1]. Results from *de novo* assembly and contig separation ("binning") via the developed bioinformatics pipeline suggested taxonomic separation. There were high fractions of recovery and coverage for individual genomes assembled *de novo* from whole cell microbiome standards. The mapping of RNA reads to genome bins showed detectable expression levels in all of the assembled genomes with detectable RNA read coverage for all strains. Functional annotation of biosynthetic gene clusters identified multiple pathways in different genomes. RNA expression profiles of microbiome standards showed different levels of gene expression and individual genome resolution. This proof-of-concept study, using integrated genome-resolved metagenomics and metatranscriptomics, demonstrates the utility, flexibility, and power of whole cell mock community standards to benchmark both the characterization and functional profiling of microbiomes.

Whole Cell Microbiome Standards

Table 1. Individual bacterial strains within the ATCC® Microbiome Standards (ATCC® MSA-2002™ and ATCC® MSA-2003™)

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

Multi-omics Applications

Metagenomics
- Compare various DNA extraction kits
- Full functional controls for metagenomics applications
- Pipeline validation

Metatranscriptomics
- Compare RNA extraction protocols
- RNA-Seq applications
- Metatranscriptomics analysis

Metabolomics
- Mass spectrometry
- Metabolites
- Proteins

Metagenomics Analysis

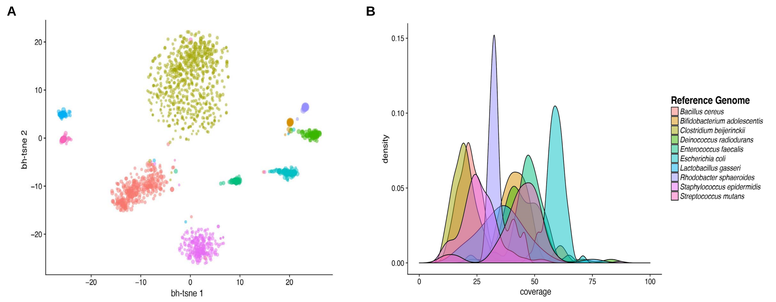

Figure 1. Metagenomics analysis of the whole cell microbiome standard (ATCC® MSA-2003™). (A) Metagenomic binning plot. Bh-tsne represents two-dimensional clustering of 5-*mer* frequencies of individual contigs (colored spots) via the Barnes-Hut Stochastic Neighbor Embedding method (Bh-tsne)[2]. (B) Sequencing coverage of individual organisms present in the whole cell microbiome standard. Density plot of % of coverage in the respective genome assembly bins.

Table 2. Metagenomic summary of individual assembled genomes composing the whole cell microbiome standard (ATCC® MSA-2003™)

Ref. genome 20150	Length (Mbp)	GC (%)	Coverage (x)	No. of contigs	Completeness	Purity
Staphylococcus epidermidis	2.32	31.7	33.1	186	100.0	99.3
Enterococcus faecalis	2.68	37.4	48.0	37	100.0	98.6
Escherichia coli	4.54	49.8	58.5	64	100.0	98.6
Rhodobacter sphaeroides	4.36	69.0	33.5	19	99.3	98.6
Streptococcus mutans	1.96	36.1	45.4	30	100.0	98.6
Lactobacillus gasseri	1.79	34.5	37.3	34	97.8	98.6
Bacillus cereus	5.16	34.6	22.3	260	100.0	92.1
Clostridium beijerinckii	3.92	30.2	19.8	623	83.5†	99.3
Bifidobacterium adolescentis	2.03	59.1	41.8	11	98.6	100.0
Dinococcus radiodurans	3.27	65.3	64.4	68	100.0	100.0

†*C. beijerinckii* had the lowest coverage and did not assemble well

Metranscriptomics Analysis

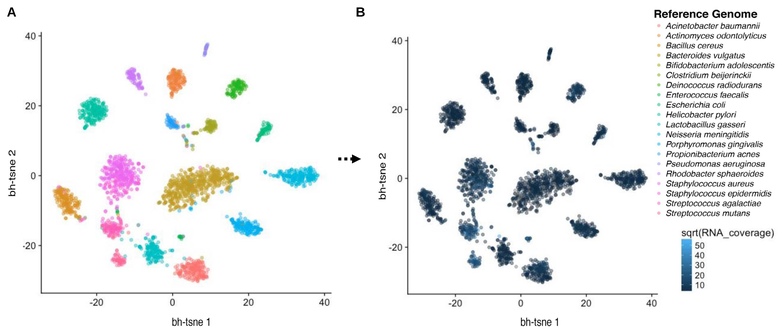

Figure 2. Metatranscriptomics analysis of the whole cell microbiome standard (ATCC® MSA-2003™). (A) Metagenomic binning. (B) Mapping RNA to the metagenome of the whole cell microbiome standard. Sqrt(RNA_coverage) represents the RNA coverage (number of RNA reads per kilobase of ORF/cluster length per million of total RNA reads that map to individual contigs) of a given contig by measuring the square root normalized by the length of the contig.

Annotation of Biosynthetic Gene Clusters

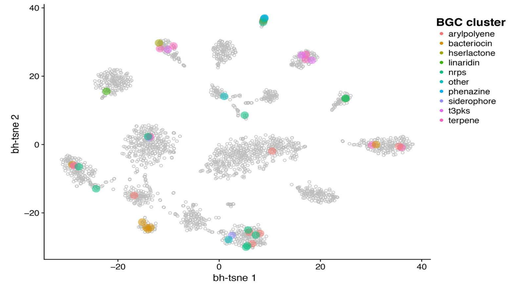

Figure 3. Functional annotation of biosynthetic gene clusters (BGC) in the whole cell microbiome standard (ATCC® MSA-2002™). Whole-genome resolution enables the assignment of functions to individual genomes by searching for ORFs in assembly contigs that have homology to secondary metabolytepathways. Expression of BGCs were identified via the antiSMASH database.

Expression Profiling of Individual Genome

Table 3. *A. baumannii* assembly and functional annotation summary

Feature	Value
Fraction assembled (%)	73.4
Length (Mbp)	3.75
GC (%)	39.1
No. contigs	184
N_{50} (bp)	30,768
No. BGCs	9†
Completeness (%)	97.8
Purity (%)	95.0

†:Arylpolyene (3), NRPS (4), siderophore (1), other (1)

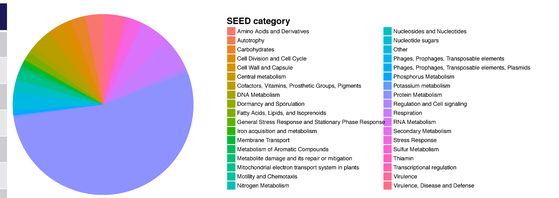

Figure 4. Expression profile of *A. baumannii* in the whole cell microbiome standard (ATCC® MSA-2002™). Genes belonging to the SEED category (Protein Metabolism) had the highest RNA coverage for the *A. Baumannii* genome. Functional analysis using the SEED classification was obtained using MEGAN5 software[3].

Conclusions

ATCC whole cell microbiome standards, which are comprised of 10 or 20 live lyophilized bacterial species, are ideal for use as controls in metagenomics, metatranscriptomics, and metabolomics applications.

- The Autometa bioinformatics pipeline was able to separate the individual genomes composing the ATCC microbiome standard.
- Metatrascriptomics enables the functional analysis of communities via whole genome RNA alignments.
- Whole-genome resolution enables the assignment of an RNA expression profile to individual genomes.
- Transcriptional activity of microbiome standards can be further classified into functional categories.

Acknowledgements/References

We would like to thank Dr. Stefan Green from the University of Illinois at Chicago for generating the RNA-Seq data.

1. Miller IJ, et al. Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. bioRxiv preprint April 30, 2018.
2. van der Maaten. Accelerating t-SNE using tree-based algorithms. JMLR 15: 3221-3245, 2014.
3. Hudson DH, et al. MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 12(6): e1004957, 2016.