June 21, 2019
FRIDAY – CPHM-940

Utility of Recombinant Bacteria as Spike-in Controls for Microbiome Studies

ATCC®
Credible leads to Incredible™

Ray-Yuan Chuang, PhD, Monique Hunter, MS, Britany Tang, BS, Briana Benton, BS, Anna McCluskey, BS, Juan Lopera, PhD
ATCC, Manassas, VA 20110

Background

To harness the quantitative potential of next-generation sequencing for data normalization, spike-in controls are essential. We have engineered three bacterial genomes (*Escherichia coli, Staphylococcus aureus,* and *Clostridium perfringens*) to contain a unique synthetic DNA tag that can be detected via 16S rRNA profiling and whole genome sequencing assays. To demonstrate the utility of the spike-in control in microbiome studies, we mixed precise quantities of genomic DNA from the recombinant bacterial strains to create a genomic DNA spike-in standard. This quantified standard was spiked into a known mock community (ATCC® MSA-1000™) containing genomic DNA prepared from 10 different bacterial strains. The resulting data showed that the unique tag of all three bacteria was identifiable and quantifiable by shotgun and 16S rRNA amplicon sequencing using V1/V2, V3/V4, and V4 primers. Spiking these recombinant bacterial genomic DNA at an optimal concentration did not affect microbiome abundance. Further, we demonstrated that the spike-in standard was applicable as an internal control for absolute quantitation. These proof-of-concept experiments support the utility of using a spike-in control with a unique 16S rRNA tag to monitor the full process of a microbiome workflow for both 16S rRNA and shotgun metagenomics assays.

Spike-in Standards for Metagenomics

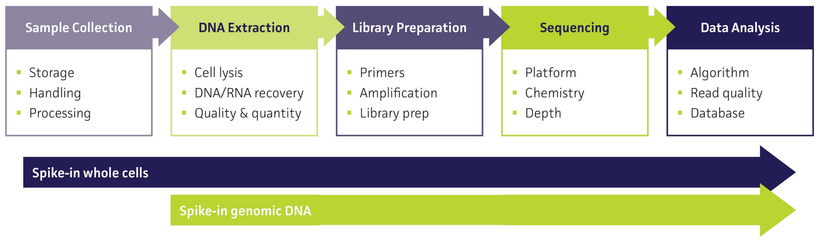


Figure 1. Spike-in standards workflow and applications. Whole cell and genomic DNA spike-in standards can be used in defined quantities at the appropriate stage to monitor biases introduced throughout the 16S rRNA gene sequencing and shotgun metagenomics analysis workflow.

Design of Three Recombinant Bacterial Strains

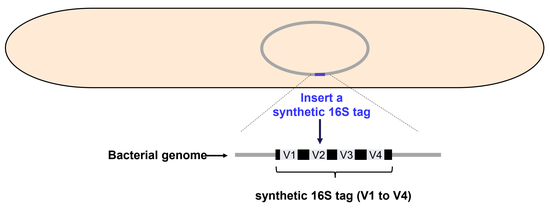


Figure 2. Production of tagged strains. ATCC created three unique synthetic 16S tag sequences mimicking the native 16S rRNA gene from 3 bacterial strains (*E. coli, C. perfringens, and S. aureus*). Each tag consists of 4 artificial variable regions (corresponding to V1 through V4 of the 16S rRNA gene) flanked by conserved regions for PCR amplification. Each tag sequence was integrated into the genome of their cognate strains to create three tagged strains.

Table 1. Description of the spike-in standards

Product Description	Gram stain	Genome size (Mb)	Tag size (bp)	% G/C	16S copies	Tag copies
Escherichia coli Tag1	Negative	4.59	829	50.8	7	1
Clostridium perfringens Tag2	Positive	3.25	799	29.0	10	1
Staphylococcus aureus Tag3	Positive	2.70	833	32.8	6	1

Development of Spike-in Standards and Quality Control

Table 2. Specifications of genomic DNA and whole cell spike-in standards

ATCC® Catalog No.	Preparation	Specification	Importance
MSA-1014™	Genomic DNA	2 x 10^7 genomes/organism per vial	➢ Microbiome measurements and data normalization
MSA-2014™	Whole cells	2 x 10^7 cells/organism per vial	➢ 16S rRNA and shotgun assay verification, validation, and quality control

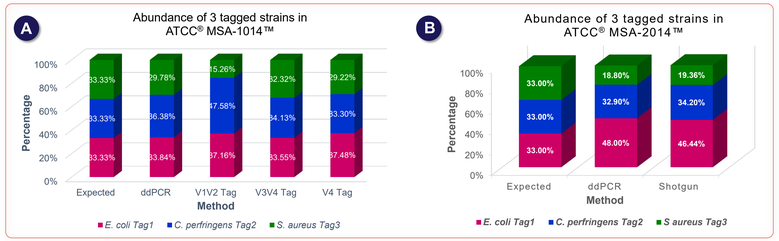


Figure 3. Relative abundance of the three tagged genomes in the genomic DNA and whole cell spike-in standards. (A) The genomic DNA spike-in standard (ATCC® MSA-1014™) was analyzed by 16S rRNA gene sequencing (V1V2, V3V4, and V4 regions) on the Illumina® MiSeq® platform, and the relative abundance of the three tags were compared to Droplet Digital™ (ddPCR) results. (B) Total DNA was extracted from the whole cell spike-in standard (ATCC® MSA-2014™) and shotgun metagenomic sequencing was performed on the Illumina® platform (Nextera XT DNA Library Preparation Kit and HiSeq 2x150). Read mapping to the tag sequences was performed via Bowtie2 tool in Geneious 11.1.4 Software and percentage abundance was normalized based on the size of each genome. ddPCR was performed using unique primers/probe sets specifically target three synthetic tags. The total genome copy number per organism per vial was calculated based on the average of triplicate reactions per specific ddPCR assay.

Evaluation of Spike-in control in a known mock community by 16S rRNA

Table 3. Experimental design of the spike-in control

Preparations and Controls	Genome copy/µL*		Percentage (%)
	MSA-1014™	MSA-1000™	Spike-in
Spike-in control	2.56 x 10^5	0	NA
Mix 1	2.33 x 10^4	2.36 x 10^5	8.99
Mix 2	2.33 x 10^3	2.36 x 10^5	0.98
Unspiked control	0	2.60 x 10^5	0

*Two different concentrations of genomic DNA spike-in standards were mixed with a constant concentration of 10 even genomic DNA (ATCC® MSA-1000™). All expedient were run in triplicated. The spike-in standard and the un-spiked genomic DNA mix were used as experimental controls.

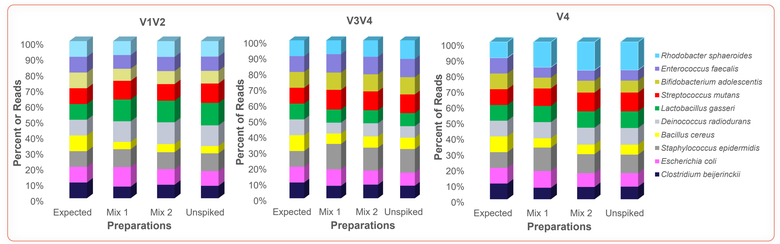
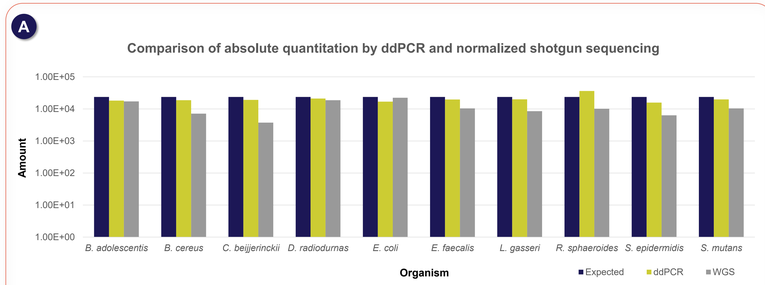


Figure 4. Relative abundance of a mock community after different concentrations of spike-in. Samples were analyzed by 16S rRNA gene sequencing (V1V2, V3V4, and V4) on the Illumina MiSeq platform and data were analyzed in the One Codex data analysis platform. The relative abundance was calculated based on the average of triplicate assay.

Table 4. Number of recovery reads of the synthetic 16S tags (V1V2, V3V4, and V4) in different spike-in concentrations

Samples	Mix 1			Mix 2		
	V1V2 Tag	V3V4 Tag	V4 Tag	V1V2 Tag	V3V4 Tag	V4 Tag
Escherichia coli Tag1	3103	660	1795	149	267	164
Clostridium perfringens Tag2	3748	534	1909	218	218	135
Staphylococcus aureus Tag3	1021	591	1453	39	245	116

Absolute quantitation of microbiome samples



Species	Expected	Genome copies (ddPCR)	Genome copies (shotgun normalization)
B. adolescentis	2.36 x 10^4	1.82 x 10^4	1.71 x 10^4
B. cereus	2.36 x 10^4	1.87 x 10^4	7.13 x 10^3
C. beijerinckii	2.36 x 10^4	1.91 x 10^4	3.74 x 10^3
D. radiodurnas	2.36 x 10^4	2.11 x 10^4	1.87 x 10^4
E. coli	2.36 x 10^4	1.69 x 10^4	2.24 x 10^4
E. faecalis	2.36 x 10^4	1.96 x 10^4	1.04 x 10^4
L. gasseri	2.36 x 10^4	2.00 x 10^4	8.50 x 10^3
R. sphaeroides	2.36 x 10^4	3.62 x 10^4	1.01 x 10^4
S. epidermidis	2.36 x 10^4	1.58 x 10^4	6.33 x 10^3
S. mutans	2.36 x 10^4	1.98 x 10^4	1.04 x 10^4

B

$$N = \frac{GR \times TB}{TR \times GB} \times SP$$

GR: # read of an organism genome
TB: base pair (bp) of average of 3 tags (821 bp)
TR: # of total tag sequence read
GB: bp of an organism genome
SP: genome copies number of spike-in

Figure 5 Genomic DNA mock microbial communities analyzed via ddPCR and shotgun metagenomics sequencing methods. (A) The DNA sample Mix 1 (see Table 3) was sequenced via a shotgun method on the Illumina platform (Nextera XT DNA Library Preparation Kit and MiniSeq 2x150). Read mapping to the tag sequences and to the 10 individual genome sequences of MSA-1000 was performed via Bowtie2 tool in Geneious 11.1.4 software and then the absolute quantitation (N) of the mock community was normalized based on the formula shown in (B).

Table 5. Number of recovered reads of the synthetic tags by shotgun sequencing

Tagged strain	Number of reads	
	Mix 1	Mix 2
Escherichia coli Tag1	396	25
Clostridium perfringens Tag2	416	24
Staphylococcus aureus Tag3	435	26

Summary

This proof-of-concept study demonstrates the utility of spike-in standards as an internal control in microbiome applications.

- The three synthetic tags are amplifiable via common 16S rRNA primers (V1-V4).
- The relative abundance of the mock community was not impacted in the presence of the proper concentration of spike-in standard.
- The reads of the tag sequence were recovered in a dose-dependent manner.
- The spike-in standards were able to serve as an internal control for absolute quantitation.

Acknowledgements: We would like to thank Nick Greenfield, One Codex, for data analysis; and Dr. Stefan Joshua Green and the Sequencing Core, University of Illinois at Chicago University, for their scientific input and sequencing expertise.

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 703.263.8277 **Email:** SalesRep@atcc.org **Web:** www.atcc.org