

Technical Data Sheet:

3 Strain Tagged Genomic DNA Even Mix

ATCC® Number	MSA-1014™
Components	33.3% <i>Escherichia coli</i> with Tag 1 33.3% <i>Clostridium perfringens</i> with Tag 2 33.3% <i>Staphylococcus aureus</i> with Tag 3
Product Description	<p>We have engineered three bacterial genomes (<i>Escherichia coli</i>, <i>Staphylococcus aureus</i>, and <i>Clostridium perfringens</i>) to contain a unique synthetic DNA tag that can be detected via 16S rRNA profiling and whole genome sequencing assays. The unique tag comprises four artificial variable regions (corresponding to the V1 through V4 regions in the 16S rRNA gene) flanked by conserved regions for PCR amplification, thereby enabling the identification of spike-in reads during the analysis of next-generation sequencing data. Using these recombinant bacterial strains, we developed an even mixture of the genomic DNA prepared from the three tagged strains for use as a spike-in control in microbiome research.</p>

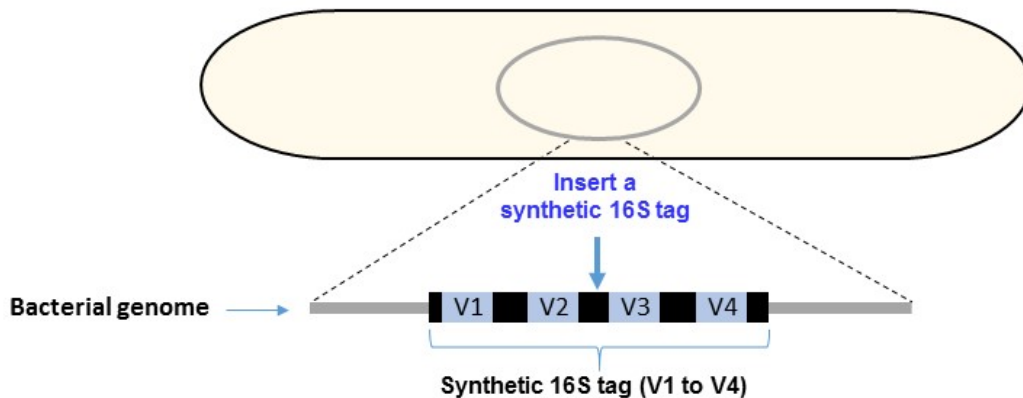


Figure 1. Production of tagged strains. ATCC created 3 unique synthetic 16S tag sequences mimicking the native 16S rRNA gene from 3 bacterial strains (*E. coli*, *C. perfringens*, and *S. aureus*). Each tag consists of 4 artificial variable regions (corresponding to V1 through V4 of the 16S rRNA gene) flanked by conserved regions for PCR amplification. Each tag sequence was integrated into the genome of their cognate strains to create three tagged strains. The tag sequences are provided below.

***Escherichia coli* 16S Tag 1**

AAATTGAAGAGTTTGGATCATGGCTCAGATTACATGCGGATTCGAGGGCCACAGGAGGCATCACTGACATGCCCTATCGTGATAGGG
GCTAGCTACAGCAGAGTGGCGGACGGGTGAGTAATGTGAGTAGCGAAAAGTCATGGCTAAAGTACTGTTTCGTCATCCGATAAGA
TTGACGGAAAATTGATTCTCACACGTCCCGATGTGGGAGCCGCGACCGTCACAGGTGAAGAATCTCTCTCAAAGATTTATGGCCATA
GTAGATTTCACTCACAATCCAGACACACGGGTAGTTCGCTGCGACTCGATTTCTAATATCTTATGGATCCTAATCTAGACTCCTACGG
GAGGCAGCAGTGGTCTACTGCATGATCAACCAAAGGTGTTCCGGCTTACGTTCAATTTGAGAACGGCGGTCTGGAGCATGAAAGGA
CGAGACGGCATTAGGACTTGCCAGGCGATGTATGCTGATCGGGAAGTAGGGAAACATGAGAGGCCGCTCTAAATCCTCTTCCGTGC
CAGCAGCCGCGTAATACGACGTGTGATCATGGTAGACGTCCACTTTTACCGTGTGTGGGCGATGAGGGATGCAAGAGGATCATTG
GTTAGCGTATTTGTACTCTGAAACTAAGGCAGGAGACCAGGTTAGGTAACGGTGCCTAGATAATCCGCAGACCGCCGGTCCGCT
AGCTTAAGGAAAGGTATGCCCGACGTGGCTTGAACAGATACCTAAACCAGGTGACGCCTATAGAGAACGACGAAGCTAATCTAT
CCGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATGTCGAC

***Clostridium perfringens* 16S Tag 2**

AAATTGAGAGTTTGGATCCTGGCTCAGGATGTACGACGAGGATTTAGGTGGGGAGGGACTGGCACGAGTAGTATACGGTTTTAAAA
GTATTGGAGCGGCGGACGGGTGAGTAACACCTAAGGCGGGTTGGGGCGTCCGAAACATACGATCCCCTGGCAAAGGTGCCAGT
GGCAGACCTGGCGGGGAGTACCGGAGCATAAAGGATTCGCAAGCACGTTACGCGGTTAGGGAGCCTGGGCTGCAAGCGCGAAGGC
CAGCGCTTTACCGTGCATGGTTAGCAAATGAGTCCCTGACCGACCACACATAATCGTACGTCCTGATCCTCTCTACAGACTCCTACG
GGAGGCAGCAGTGGTAGGCTCCATAATGCTAGTCGACCTCGTCTGGTGCCTGCTTCAACCGTTCACAAGAACTCTCCTGCCAACGT
TAATCGGCGTAGCGCGTAATCGATCACCAGTGTAGTACGTATCTATCCCTATACGTGCCAGCAGCCGCGTAATACGTGAAGGTTT
GAACTGAAATCAAGAAAGTTAATCAAGGGTTGCGTGC CGGAATCGGCGTAAAAACAAATTGAGCGGGTGGGAACAAAACGAAGA
TGGTAGTTCTATAGTTGCAGATAACTCCCCTAAGTGTGCGTACGGGATAGCGATATGTACTGGCC
TAAGAGCTCCGATTTCTAAGCTGCTGGGTGCAATGTAAGCACGACACATCTAATCCGCAAACAGGATTAGATACCCTGGTAGTCC
ACGCCGTAACGATGAATAC

***Staphylococcus aureus* 16S Tag 3**

TTTATGGAGAGTTTGGATCCTGGCTCAGGATCATTACCTCGATTGAGATAAGCAAACAAGTCTCGCTAGTGAAGGCACGTCTGATCGT
CACACCTGGAAGCGGCGGACGGGTGAGTAACACCTTACAGTATGGTATCCTGGATATTCACCTAGTATTACTGGTCTCCATAGTAA
CGGCGTGTCTGATAAGCGGTCCCAGGATTAGATAAATCGACTGGCCTATGAATGCCTAGGCACACGAGTGGGATACGATCAAGAGT
GACACAGAGCGAACCCCGTCAAATGTCCGGAATATGGGTGTATATGACGTACCCAGAGTGGTATAGAAAATCCTTGTGAAGACTCC
TACGGGAGGCAGCAGTAGAGGTGGTGCAGGCACAGTTAGGTTTCATGATCATGGCGACAGAGAAGCTGGGCACAGCGAAGGTCAA
ACCAACAACAGTAGCAGTCACTACTAAGCTGGGAGGTGATGGTCTGCGCATAGGGTATTAACACACTGGTAATTGATCCTGATAC
GTGCCAGCAGCCGCGTAATACCGAAGTCTATTATCTCGGCATGCTCGTGGAGCTCAGACCGCTGAGGTGAAGTATAAAGTGTTCCG
AGGATCGAGATATAACGGCTCATATATGTGATGGGACAGTTTAAAATACGCGGATATGCAGTGCACGGACCAGGAGGGACGGAG
AGGGACCTCTACTTGCAATCGTTCAATGGAGGTGAGTACCGCAGAGAGTAGGTAATACTGTGAGACGAAGAGAAAGAGATTTGTG
AATCCTCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATGAGTGC

Table 1. Composition of the 3 Strain Tagged Genomic DNA Even Mix (ATCC® MSA-1014™)

Species	Gram Stain	Genome size (Mb)	Tag size (bp)	G/C Content (%)	Number of 16S Copies	Number of Tag Copies	Genome copies 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

*Specification Range: Total of 6.0 x 10⁷ genome copies/vial ± 1 log. Genome copy number indicates ATCC® manufacturing specifications and are provided as a reference only.

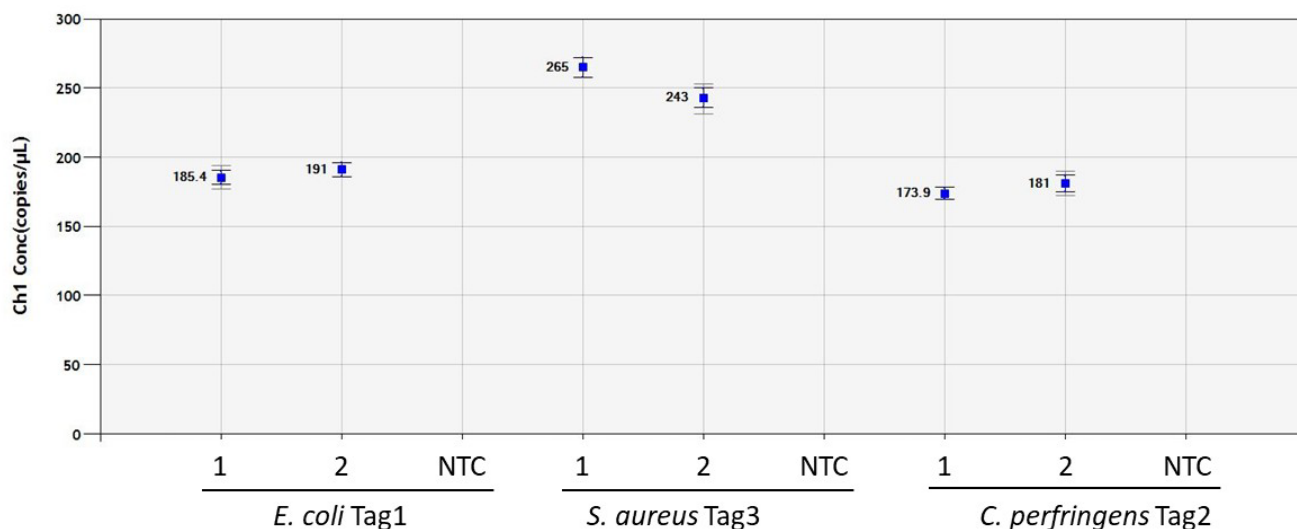


Figure 2. Calculating genome copy number of individual preparations in ATCC® MSA-1014™ by Droplet Digital™ PCR (ddPCR™; Bio- Rad) assay. Two vials of ATCC® MSA-1014™ were diluted 500-fold and assayed in triplicate (1, 2) along with a no template control (NTC). The primer and probe set used for the three tag amplicons were as follows: *E. coli* Tag1 (Forward: 5'- GAACGGGTGAGTAATGTGAGTAG -3'; Reverse: 5'-CGGGACGTGTGAGAATCAAT -3'; Probe: 5'-TCGGATGACGAAACAGTACCTTTAGCC -3'), *C. perfringens* Tag2: (Forward: 5'- TAACCATGCACGGTAAAGCG -3'; Reverse: 5'-AAGGATTCGCAAGCACGTTT -3'; Probe: 5'-TTCGCGCTTGACGCCAGGCTCCCT -3'), and *S. aureus* Tag3: (Forward: 5'-AATACGCGGATATGCAGTGC 3'; Reverse: 5'-CTGACCTCCATTGAACGATTGC -3'; Probe: 5'-AGAGTCCCTCTCCGTCCTCCTGGTCCGT -3'). The concentration (genome copies/μL) represents the average of three wells in triplicate assays. The total genome copy number per organism per vial was calculated based on the average of triplicate reactions per specific ddPCR™ assay (Table 2).

Table 2. Genome copy number of individual bacteria in the 3 Strain Tagged Genomic DNA Even Mix

Species	Genome copy number per vial	
	Vial 1	Vial 2
<i>Escherichia coli</i> Tag1	$1.86 \pm 0.06 \times 10^7$	$1.90 \pm 0.02 \times 10^7$
<i>Clostridium perfringens</i> Tag2	$1.75 \pm 0.04 \times 10^7$	$1.78 \pm 0.05 \times 10^7$
<i>Staphylococcus aureus</i> Tag3	$2.65 \pm 0.05 \times 10^7$	$2.42 \pm 0.08 \times 10^7$

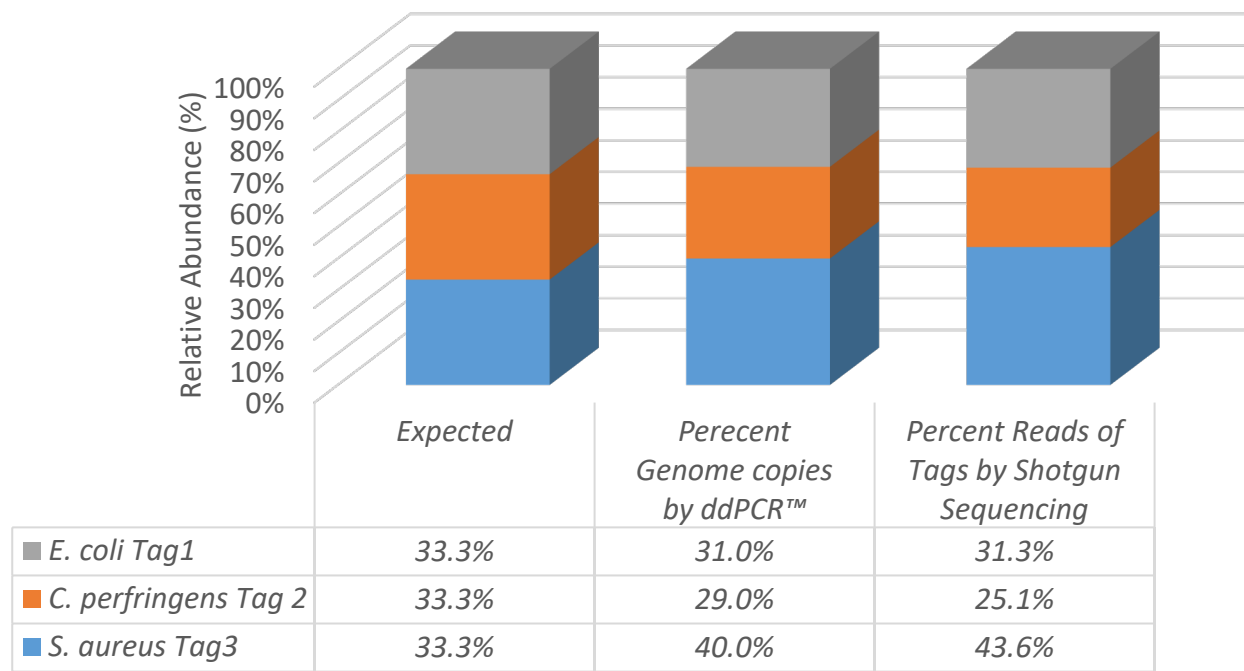


Figure 3. Relative abundance of three genomes extracted from ATCC® MSA-1014™. Shotgun metagenomic sequencing was performed on the Illumina® platform (Nextera XT DNA Library Preparation Kit and MiSeq 2x250). Total shotgun reads (29.6 million) were mapped to unique tag sequences (total = 6020 reads), and the percent of reads mapped to individual tags were calculated.

© 2019 American Type Culture Collection. The ATCC trademark and trade name, and any other trademarks listed in this publication are owned by the American Type Culture Collection unless indicated otherwise. ddPCR and Bio-Rad are trademarks of Bio-Rad Laboratories, Inc. Illumina, Nextera, and MiSeq are registered trademarks of Illumina, Inc.

These products are for laboratory use only. Not for human or diagnostic use. ATCC products may not be resold, modified for resale, used to provide commercial services, or to manufacture commercial products without prior ATCC written approval. The quantified data provided in this document is a reference point. It is user's responsibility to determine the number of genome copies for spike-in studies on the basis of the DNA extraction method and other methodology used.